Université Paris-Est

Lab. Informatique Gaspard Monge

UMR CNRS 8049

Lebanese University

Doctoral school

of sciences and technology

THESIS

presented by

Mireille EL GHECHE

27 Mai 2014, Paris

Proximal methods for convex minimization of φ-divergences. Application to computer vision.

Examiner	Michèle Basseville	CNRS research director
Reporter	Amel Benazza	Prof. Université de Carthage
PhD co-supervisor	Caroline Chaux	CNRS researcher, LATP
Examiner	Bilal Chebaro	Prof. Lebanese University
PhD supervisor	Joumana farah	Prof. Holy-Spirit University of Kaslik
Examiner	Alfred Hero	Prof. University of Michigan
Reporter	François Malgouyres	Prof. Université Paul Sabatier
PhD supervisor	Jean-Christophe Pesquet	Prof. Université Paris-Est Marne-la-Vallée

إلى أمّي الحبيبة وأبي العزيز،
وأحبائي:
ميرنا، هيثم، مروان ومي.

To my lovely family:
Mom, Dad,
Mirna, Haytham,
Marwan and May.

كذا تتغير الأشياء أمام أعيننا بتغير عواطفنا،
وهكذا نتوهم الأشياء متشحة بالسحر والجمال،
عندما لا يكون السحر والجمال إلا في نفوسنا.

جبران خليل جبران

The appearance of things changes according to the emotions,
and thus we see magic and beauty in them,
while the magic and beauty are really in ourselves.

Gibran Khalil Gibran

Acknowledgments

I would like to acknowledge to my supervisors Jean-Christophe Pesquet and Joumana Farah for their support along the three-year period of my PhD and the invaluable guidance on my research topic. It was a real pleasure to work with them, to share passion for optimization, to be enrolled into scientific discussions and to learn how to proceed from an idea to an article.

I would especially like to thank Caroline Chaux for her generous support during the first year of my PhD. She shared her expertise with me very generously and I have learned a lot from her. My thanks and appreciations also go to Béatrice Pesquet-Popescu for her assistance and expert guidance.

My deepest gratitude is due to the members of my jury, Michèle Basseville, Amel Benazza, Bilal Chebaro, Alfred Hero and François Malgouyres, for their time. It was a great honor for me to present my work to them.

I would like to acknowledge to the LIGM and Telecom ParisTech groups for the great atmosphere in these laboratories. I was lucky to meet there many friends and great young scientists. I would like to thank to Anna Jezierska, Audrey Repetti, Mai Quyen Pham, Feriel abboud, Yosra Marnisi, Emilie Chouzenoux, Aurélie Pirayre, Mounir Kaaniche, Florian Dupuy, Olena Tankyevych, Camille Couprie, Ravi Kiran, Anisia Florescu and Giovanni Petrazzuoli for their kind assistance. Without forgetting Patrice Herault who has always been there. Special thanks also to all my friends, especially Aida Arida, Pascale Asmar, Mohammad El Achkar, Mashhour El Chakouch, Ahmad Diab, Karam Fayad, Rim Kaddah, Ahlem Khlass, Karim Youssef, Pierre Youssef and Kifah Sarraf.

Special thanks go to Giovanni Chierchia for his insights into the problems I studied, his expertise in convex optimization and his lovely support.
At last but not least, I would like to thank my family: without them, I would not have been strong enough to complete this thesis, go through all the difficulties, and overcome them.

Abstract

Convex optimization aims at searching for the minimum of a convex function over a convex set. While the theory of convex optimization has been largely explored for about a century, several related developments have stimulated a new interest in the topic. The first one is the emergence of efficient optimization algorithms, such as proximal methods, which allow one to easily solve large-size nonsmooth convex problems in a parallel manner. The second development is the discovery of the fact that convex optimization problems are more ubiquitous in practice than was thought previously.

In this thesis, we address two different problems within the framework of convex optimization. The first one is an application to computer stereo vision, where the goal is to recover the depth information of a scene from a pair of images taken from the left and right positions. The second one is the proposition of new mathematical tools to deal with convex optimization problems involving information measures, where the objective is to minimize the divergence between two statistical objects such as random variables or probability distributions.

We propose a convex approach to address the problem of dense disparity estimation under varying illumination conditions. A convex energy function is derived for jointly estimating the disparity and the illumination variation. The resulting problem is tackled in a set theoretic framework and solved using proximal tools. It is worth emphasizing the ability of this method to process multicomponent images under illumination variation. The conducted experiments indicate that this approach can effectively deal with the local illumination changes and yields better results compared with existing methods.

We then extend the previous approach to the problem of multi-view disparity estimation. Rather than estimating a single depth map, we estimate a sequence of disparity maps, one for each input image. We address this problem by adopting a discrete reformulation that can be efficiently solved through a convex relaxation. This approach offers the advantage of handling both convex and nonconvex similarity measures within the same framework. We have shown that the additional complexity required by the application of

our method to the multi-view case is small with respect to the stereo case.

Finally, we have proposed a novel approach to handle a broad class of statistical distances, called φ-divergences, within the framework of proximal algorithms. In particular, we have developed the expression of the proximity operators of several φ-divergences, such as Kulback-Leibler, Jeffrey-Kulback, Hellinger, Chi-Square, I_α, and Renyi divergences. This allows proximal algorithms to deal with problems involving such divergences, thus overcoming the limitations of current state-of-the-art approaches for similar problems. The proposed approach is validated in two different contexts. The first is an application to image restoration that illustrates how to employ divergences as a regularization term, while the second is an application to image registration that employs divergences as a data fidelity term.

Resumé

Cette thèse s'inscrit dans le contexte de l'optimisation convexe. Elle apporte à ce domaine deux contributions principales. La première porte sur les méthodes d'optimisation convexe non lisse appliquées à la vision par ordinateur. Quant à la seconde, elle fournit de nouveaux résultats théoriques concernant la manipulation de mesures de divergences, telles que celles utilisées en théorie de l'information et dans divers problèmes d'optimisation.

Le principe de la stéréovision consiste à exploiter deux images d'une même scène prises sous deux points de vue, afin de retrouver les pixels homologues et de se ramener ainsi à un problème d'estimation d'un champ de disparité. Dans ce travail, le problème de l'estimation de la disparité est considéré en présence de variations d'illumination. Ceci se traduit par l'ajout, dans la fonction objective globale à minimiser, d'un facteur multiplicatif variant spatialement, estimé conjointement avec la disparité. Nous avons mis l'accent sur l'avantage de considérer plusieurs critères convexes et non-nécessairement différentiables, et d'exploiter des images multicomposantes (par exemple, des images couleurs) pour améliorer les performances de notre méthode. Le problème d'estimation posé est résolu en utilisant un algorithme parallèle proximal basé sur des développements récents en analyse convexe.

Dans une seconde partie, nous avons étendu notre approche au cas multi-vues qui est un sujet de recherche relativement nouveau. Cette extension s'avère particulièrement utile dans le cadre d'applications où les zones d'occultation sont très larges et posent de nombreuses difficultés. Pour résoudre le problème d'optimisation associé, nous avons utilisé des algorithmes proximaux en suivant des approches multi-étiquettes relaxés de manière convexe. Les algorithmes employés présentent l'avantage de pouvoir gérer simultanément un grand nombre d'images et de contraintes, ainsi que des critères convexes et non convexes. Des résultats sur des images synthétiques ont permis de valider l'efficacité de ces méthodes, pour différentes mesures d'erreur.

La dernière partie de cette thèse porte sur les problèmes d'optimisation convexe impliquant des mesures d'information (φ-divergences), qui sont largement utilisés dans le codage source et le codage canal. Ces mesures peuvent être également employées avec succès dans des problèmes inverses rencontrés dans le traitement du signal et de l'image. Les problèmes d'optimisation associés sont souvent difficiles à résoudre en raison de leur grande taille.

Dans ce travail, nous avons établi les expressions des opérateurs proximaux de ces divergences. En s'appuyant sur ces résultats, nous avons développé une approche proximale reposant sur l'usage de méthodes primales-duales. Ceci nous a permis de répondre à une large gamme de problèmes d'optimisation convexe dont la fonction objective comprend un terme qui s'exprime sous la forme de l'une de ces divergences.

Vue d'ensemble du travail effectué

Optimisation convexe

Le chapitre 2 a pour but d'introduire le lecteur aux problèmes inverses dans le cadre convexe. Notre exposé se limite à la description des outils mathématiques nécessaires à notre étude. Pour ce faire, nous avons fourni, dans un premier temps, un bref aperçu des problèmes inverses en nous basant sur des approches variationnelles, pour ensuite présenter les algorithmes d'optimisation convexe. Une attention particulière a été accordée à la projection sur des ensembles convexes, à la fin de ce chapitre.

Estimation de la disparité

La stéréoscopie est l'un des problèmes fondamentaux de la vision par ordinateur. Le procédé de la stéréoscopie est calqué sur la perception humaine du relief grâce aux deux images planes que perçoit chaque œil. Ce processus constitue un enjeu majeur en vision par ordinateur. Une étape cruciale dans la résolution de ce problème est la mise en correspondance stéréoscopique.

Le problème de la mise en correspondance est un procédé de reconstruction de la profondeur à partir de deux images de la même scène acquises avec des angles différents. En établissant que deux points correspondants, dans deux images acquises, représentent la projection d'un même point de la scène, on peut calculer par triangulation la position tridimensionnelle exacte de ce point dans l'espace.

L'un des problèmes les plus délicats en vision par ordinateur, notamment dans le cadre de la mise en correspondance stéréoscopique, est

celui de changements d'illumination qui peuvent affecter deux vues d'une même scène. Un autre problème critique provient de la présence d'objets dont la surface est uniforme ou de la présence de textures répétitives. En effet, les pixels dans ces régions ne représentent pas de primitives discriminantes, c'est à dire que plusieurs correspondants potentiels ayant la même valeur de luminance existent dans l'autre image. Les régions homogènes sont donc difficiles à apparier. Par ailleurs, une autre difficulté rencontrée par les méthodes de mise en correspondance concerne les occultations. Des points de la scène peuvent être visibles dans une image de la paire stéréoscopique mais pas dans l'autre.

Pour pallier ces difficultés, les méthodes de mise en correspondance sont amenées à exploiter toutes les informations disponibles afin de faciliter la recherche et la détermination des correspondants. Les techniques les plus simples et les plus courantes se basent sur une recherche locale de similarité entre de petites régions, typiquement par optimisation d'une fonction de corrélation. On peut aussi segmenter les images et mettre en correspondance les régions obtenues. L'avantage de ces méthodes est que la mise en œuvre est simple et efficace pour un temps de calcul relativement faible. Cependant, les techniques locales sont sensibles aux occultations, aux régions faiblement texturées et aux motifs répétitifs. Les approches globales tentent de pallier ces inconvénients en minimisant une fonction de coût globale qui fait intervenir l'ensemble des pixels des images, mais au prix d'une complexité de calcul plus élevée. Plusieurs méthodes d'optimisation ont été proposées dans la littérature. Nous pouvons distinguer les méthodes à variables discrètes, qui effectuent un appariement de couples de pixels, comme la programmation dynamique ou les coupures de graphes, des méthodes à variables continues, où la disparité estimée prend des valeurs réelles "quelconques", comme dans les approches variationnelles.

Dans notre travail, nous nous plaçons dans ce dernier cadre en employant des méthodes parallèles reposant sur des opérateurs proximaux. Dans le Chapitre 3, le problème de mise en correspondance est alors formulé comme un problème d'optimisation convexe sous contraintes. Une fonction objective convexe est minimisée sur l'intersection d'ensembles convexes. Ces ensembles sont associés à différentes contraintes modélisant des informations *a priori* sur les propriétés des champs à estimer. L'originalité de cette approche se traduit par plusieurs aspects:

- L'estimation de la carte de disparité est effectuée ici en présence de variations d'illumination dans la scène observée. Ceci se traduit par l'ajout, dans la fonction objective globale à minimiser, d'un

facteur multiplicatif variant spatialement, qui est estimé conjointement avec la disparité.

- Des images multi-composantes sont prises en considération. Cette technique est élaborée en prenant en compte les trois canaux couleurs (RGB, YUV, ...). La formulation globale n'étant pas convexe, on propose une méthodologie d'approximation convexe basée sur un développement de Taylor au premier ordre.
- Le critère convexe (distance) considéré peut être choisi de manière très générale et peut être non différentiable (une norme ℓ_1 par exemple).
- L'algorithme considéré offre une grande flexibilité quant au choix du critère à minimiser et à l'incorporation de contraintes multiples. Par ailleurs, il faut souligner que la méthode peut être implantée sur une architecture parallèle (multi-coeurs, multi-processeurs, GPU, ...).

Dans les résultats de simulation fournis, nous démontrons l'efficacité de notre approche.

Estimation de la disparité en multi-vues

L'imagerie multi-vues a suscité un vif intérêt dans différentes applications telles que le cinéma numérique, la médecine, le trafic aérien, les technologies militaires, les jeux vidéos, etc. Le développement de la télévision 3D et des supports auto-stéréoscopiques a suscité de nombreuses études portant sur le traitement de séquences d'images. Il s'agit en particulier, d'extraire de l'information 3D à partir des vues disponibles, sous forme de cartes de disparités.

Dans le Chapitre 4, nous étendons l'approche présentée dans le Chapitre 3 au cas multi-vues. On dispose de caméras calibrées, dont les paramètres sont connus et les images acquises sont rectifiées. Nous proposons d'estimer les cartes de disparité de chaque image présente dans la séquence en exploitant toutes les données dans les vues adjacentes. La stratégie adoptée permet d'exploiter les zones occultées qui se trouvent dans certaines vues, ainsi que d'améliorer la précision des estimations (par rapport au cas stéréo).

Pour résoudre les problèmes d'optimisation associés, nous avons utilisé des algorithmes proximaux en suivant deux approches de relaxation convexe différentes:

- Premièrement, nous avons considéré une approche multi-étiquettes fondée sur des développements récents de l'analyse convexe. La technique employée présente l'avantage de pouvoir manipuler un grand nombre d'images avec une complexité faible par rapport

au cas stéréo. Cet algorithme offre une grande flexibilité quant à l'incorporation de critères convexes mais aussi non convexes (ℓ_1-tronquée, $\ell_{\frac{1}{2}}$-tronquée, ...)

- Deuxièmement, nous appliquons une linéarisation basée sur un développement de Taylor amélioré.

Des résultats sur des images ont permis, pour différentes formes de critère de mesure d'erreur, de valider l'efficacité de la première méthode par rapport à la seconde et d'autres approches existantes.

Opérateurs proximaux des φ-divergences

La théorie de l'information est une branche émergente des mathématiques appliquées et de l'informatique, issue de la volonté de quantifier les échanges d'information. Dans les années 40, Claude Shannon a développé la théorie de l'information qui décrit les aspects les plus fondamentaux des systèmes de transmission. Cette théorie s'intéresse à la construction et à l'étude de modèles mathématiques, en se basant essentiellement sur la théorie des probabilités. Elle s'est développée au fur et à mesure et est devenue aujourd'hui incontournable dans la conception des systèmes de communication. Elle a été étendue à d'autres champs d'applications, ou le codage, comme pour le codage source/canal.

Une fonction fondamentale de la mesure de l'information est la divergence de Kullback -Leibler, qui mesure la distance entre deux objets statistiques telles que des variables aléatoires par le biais de leurs distributions de probabilité. Une telle divergence est un cas particulier d'une classe plus large de distances appelées φ-divergences, incluant celles de Jeffreys-Kullback, Hellinger, Chi-Square, I_α, et de Renyi. Nous présentons dans ce qui suit un bref résumé sur l'état de l'art concernant ces divergences:

- La divergence de Kullback-Leibler(KL) a un rôle de premier plan dans le domaine de codage source/canal pour le calcul de la capacité de transmission et de la fonction de débit-distorsion. Ces problèmes sont souvent résolus par des algorithmes basés sur la minimisation alternée proposés par Blahut et Arimoto. La symétrisation classique de Kullback- Leibler est la divergence de Jeffreys-Kullback, qui est connue dans plusieurs applications surtout dans les méthodes basées sur les algorithmes des k-moyennes.
- La divergence d'Hellinger a été initialement introduite par Beran et elle est connue sous différentes dénominations (telles que la

divergence de Jeffreys-Matusita). Dans le domaine de la théorie de l'information, la divergence d'Hellinger est couramment utilisée dans l'analyse des données, des statistiques et de l'apprentissage automatique.

- La divergence Chi-carré a été introduite par Pearson qui l'a utilisée pour évaluer quantitativement si un phénomène observé tend à confirmer une hypothèse donnée. Cette divergence a été appliquée avec succès dans différents contextes, tels que la théorie de l'information et le traitement du signal, comme une mesure de dissimilarité entre deux distributions de probabilités.
- La I_α-divergence a été initialement proposée pour évaluer statistiquement l'efficacité d'un test d'hypothèse. Elle a été étendue par de nombreux chercheurs et est surtout considérée dans le contexte de la factorisation des matrices non-négatives.
- La divergence de Rényi a été introduite par Alfréd Rényi comme une mesure de l'information liée à l'entropie Rényi. Elle a été étudiée et appliquée dans de nombreux domaines, y compris les problèmes de recalage d'image.

Etant des fonctions de dissimilarité, les φ-divergences peuvent servir comme critères de minimisation dans les problèmes d'optimisation. Cependant, ce sont des fonctions non lisses, ce qui exclut les techniques d'optimisation lisses. De plus, les approches existantes impliquant ce type de mesures se limitent souvent à des problèmes où la minimisation est effectuée par rapport à l'un des arguments de la divergence, ou en utilisant des approches alternées qui nécessitent des hypothèses spécifiques pour être valides.

Les méthodes proximales constituent un cadre approprié pour les φ-divergences, puisque ce sont des méthodes permettant de traiter des fonctions convexes non lisses. La difficulté provient du fait que le calcul de l'opérateur proximal des ces divergences n'est pas trivial. En effet, une φ-divergence est une somme de fonctions à deux variables non séparables. Par conséquent, le défi consiste à calculer des expressions explicites de ces opérateurs. Dans le Chapitre 5, nous avons établi les expressions des opérateurs proximaux de ces divergences. Ce calcul nous a permis de développer de nouvelles méthodes proximales qui permettent d'aborder des formes plus générales de problèmes d'optimisation. Les expressions des opérateurs proximaux présentées, directement liée à des fonctions convexes de deux variables réelles, ont enrichi la liste des fonctions pour lesquelles un tel opérateur proximal peut être facilement calculé.

Projections épigraphiques

Les projections épigraphiques revêtent un grand intérêt lorsqu'il s'agit de traiter des contraintes pour lesquelles une forme explicite de la projection n'existe pas. Dans ce contexte, nous avons fourni dans le Chapitre 5 des résultats théoriques pour montrer comment les formes des opérateurs proximaux des divergences peuvent être utilisées pour calculer l'expression de la projection sur l'épigraphe d'un certain nombre de fonctions convexes.

Applications des opérateurs proximaux des φ-divergences

Les divergences ont été largement utilisées dans la littérature dans différents contextes. Des mesures d'information ont été exploitées pour comparer deux fonctions de densité de probabilité, et certaines servent à mesurer la différence entre deux vecteurs aléatoires de même distribution. Dans le Chapitre 6, nous utilisons ces mesures d'information dans deux applications différentes. Tout d'abord, nous étendons le cadre de la variation totale non locale en utilisant des divergences pour construire une nouvelle régularisation (en débruitage et en déconvolution). Le recalage d'images constitue un autre exemple illustrant l'utilité des opérateurs proximaux proposés.

Conclusion

Nous résumons nos contributions et dégageons quelques pistes pour de futurs travaux dans le Chapitre 7.

Contents

- Chapter 1 -

Introduction

"What is written without effort is in general read without pleasure."
`Samuel Johnson`

Optimization is a rich and thriving mathematical discipline. Historically, the first methodology for solving nonlinear optimization problems is due to Fermat (1601–1665). Because of this work, Lagrange (1736–1813) stated clearly that he considered Fermat to be the inventor of calculus (as opposed to Newton (1643–1727) and Leibnitz (1646–1716) who were later locked in a bitter struggle for this honor). At the beginning of the nineteenth century, the least-squares method was firmly established as a scientific approach by Gauss, Legendre, and Laplace within the space of a decade. Since then, optimization theory has grown ever more sophisticated. The powerful and elegant language of convex analysis unifies much of this theory, to which a strong contribution was brought, among others, by Moreau (1923–2014). A large number of important analytic problems have illuminating optimization formulations and hence can be approached through convex analysis tools. In this thesis, we present two main contributions within the framework of convex optimization. The first one concerns an application of recent nonsmooth optimization methods to computer stereo vision, while the second one provides novel theoretical tools to deal with divergence measures in optimization problems.

Disparity estimation One of the main challenges in computer vision consists of reproducing human abilities by means of mathematical models involving geometry, statistics, and learning theory. A popular example is *computer stereo vision*, inspired from a biological process called *stereopsis*. Stereopsis is the impression of depth that is perceived when a scene is observed with two eyes. Indeed, binocular vision creates two slightly different images of the scene in the two eyes, due to the different positions of the eyes on the head. These differences, referred to as *binocular disparity*, provide information that the brain can use to calculate depth in the visual scene, providing a major means of *depth perception*.

In stereo vision, binocular disparity refers to the difference in location of an object seen from two different viewpoints. Such a difference results from the horizontal separation of cameras (see Figure 1.1). In the same way as the brain computes the depth of the visual scene, the disparity serves to extract 3D information from 2D digital images, since the disparity is inversely proportional to the depth. Disparity map can be recovered by estimating the relative position of features in the stereo pair. This process is called *stereo matching*. Looking for correspondences between stereo images is a difficult task, because of the presence of hidden areas (i.e. *occlusions*) and because of the fact that the light is reflected differently depending on the viewing angle. A huge amount of solutions have been proposed to address stereo matching. In this thesis, we present original approaches based on convex optimization.

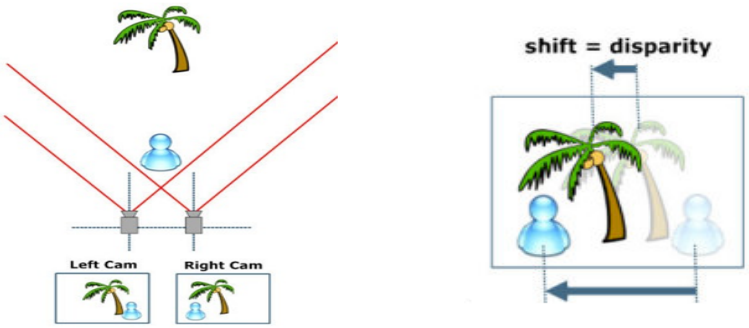


Figure 1.1

Divergence measures Information theory is a branch of applied mathematics and computer science involving the quantization of information. It was developed by Shannon to find fundamental limits on communication and signal processing operations. Since its inception, it has broadened to find applications in many other areas, including statistical inference, natural language processing, cryptography, neurobiology, quantum computing, and other forms of data analysis. Important sub-fields of information theory are source/channel coding, algorithmic information theory, complexity theory, and measures of information (see Figure 1.2). A key measure of information is the Kullback-Leibler divergence, which quantifies the distance between two statistical objects such as random variables or probability distributions. Such a divergence is a special case of a broader class of statistical distances called φ-divergences. Thanks to their interesting properties, φ-divergences have been widely used in optimization problems as similarity criteria between the sought solution and a given observation. However, existing approaches involving these kinds of measures are often restricted to problems where the minimization is performed w.r.t. one of the arguments of the divergence, or by using alternating minimization approaches which require specific assumptions to be valid. In this thesis, we present an original convex optimization approach that circumvents the above limitations.

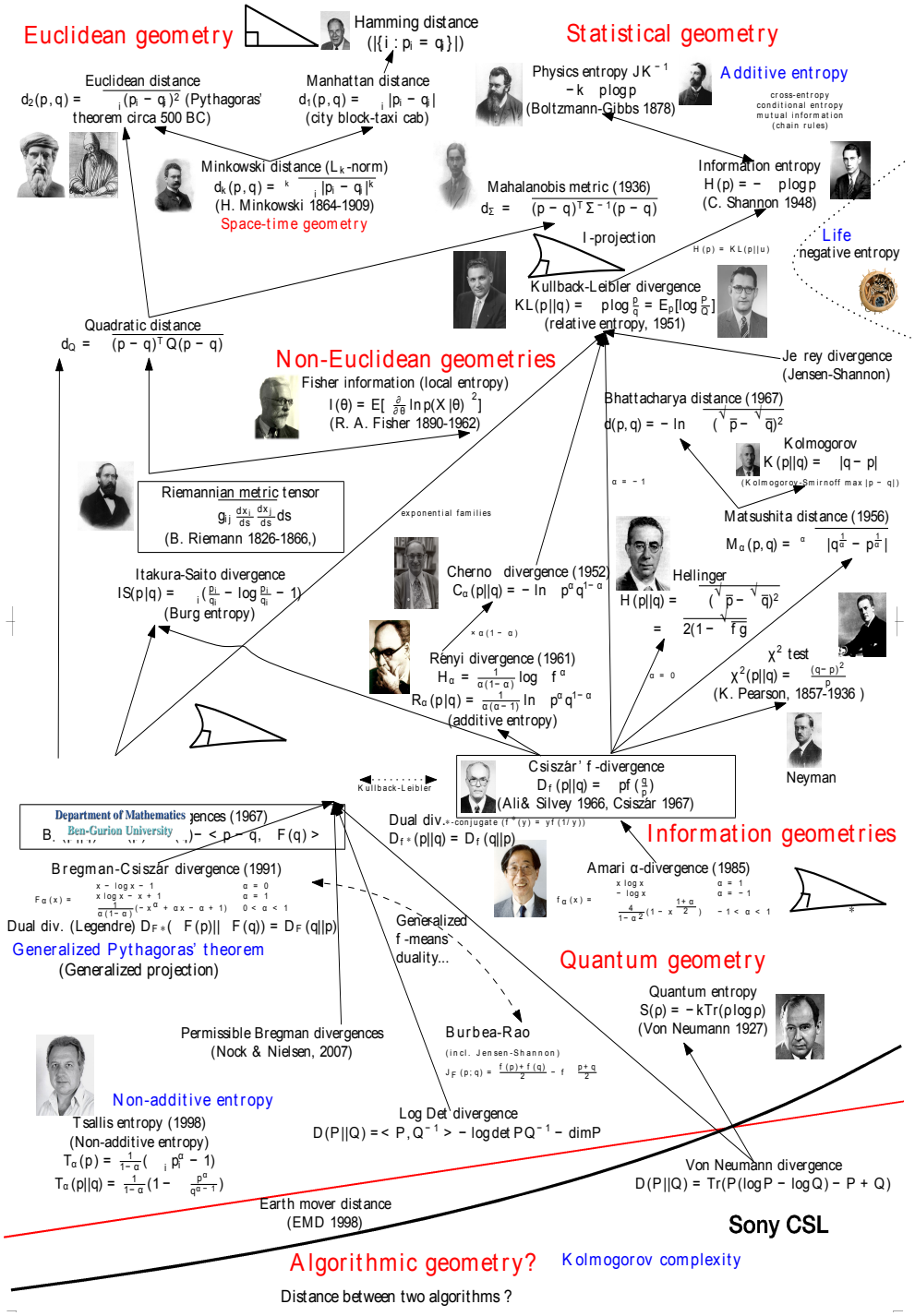


Figure 1.2: *Taxonomy of principal distances.* ©*Frank Nielsen, 2007. Version 0.2 Last updated, May 2008*

Challenges

Estimating depth from stereo images is a central challenge in computer vision and pattern analysis. Recovering 3D information from a pair of stereo cameras has been a popular topic because the additional 3D information provided by this technology contains significantly more information than 2D clues produced by traditional cameras. However, this 2D to 3D evolution has always been facing many challenges, which can be grouped into two main categories: accuracy and efficiency. Accuracy becomes an important concern in applications such as precise 3D surface modeling, especially when dealing with object surfaces with reflectance behavior, rich geometric structure, significant amount of occlusion and poor texture. Efficiency is one of the main issues when the stereo system is employed in real-time applications such as robot navigation, video surveillance, and interactive user interfaces. Unfortunately, these challenges often conflict with each other: in order to improve the quality of stereo matching, people usually cast the problem as a global optimization problem, which results in a high computational cost and a poor efficiency. On the other hand, most efficient stereo matching algorithms are based on only local information, which leads to poor accuracy in some difficult situations. Another challenge in depth estimation from stereo images is the presence of hidden surfaces (occlusions), moving objects or surfaces that reflect the light differently depending on the viewing angle, therefore complicating the matching task.

The rapid development of computer technology in the last decades has changed the methodologies underlying the 3D reconstruction. In this context, a possible approach consists of first extracting the depth from multiple images, and then reconstructing 3D scene using the depth information. Therefore, multi-view depth estimation has become a crucial step in computer vision. However, this kind of methods have to face the challenge of efficiently handling a massive amount of multi-view data (w.r.t. the stereo case).

φ-divergences are often used as discrete measures in signal processing problems. They serve as dissimilarity functions in many information theoretic models, data recovery tasks, and machine learning. Their popularity stems from their intimate connections with information theory concepts such as the entropy and the mutual information. Being dissimilarity functions, they may serve as criteria to be minimized in optimization problems. However, φ-divergences are nonsmooth functions, which rules out smooth optimization techniques. Proximal methods represent an ideal framework for φ-divergences, as these methods are able to deal with nonsmooth convex functions. Here, the difficulty stems from the fact that computing the proximity operator of φ-divergences is not trivial since they are not separable functions of scalar variables. Consequently, the challenge is to derive closed form expressions of these operators.

Contributions

Applications to computer stereo vision:

- We propose a convex optimization framework to address the problem of dense disparity estimation under varying illumination conditions. A convex energy function that takes into account the illumination variation model is derived by resorting to a relaxation based on a first-order Taylor approximation around an initial estimate. This energy is then minimized while taking into consideration various convex constraints arising from prior knowledge and observed data (Chapter 3). The proposed method allows us to incorporate various convex distances and it relies on the extension of the Parallel ProXimal Algorithm (PPXA).

- We show, for the same optimization problem, the ability to consider multicomponent images with illumination variation, and also the flexibility in minimizing various (possibly nonsmooth) convex similarity measures and combining them with various convex constraints.

- In the context of multi-view images, we address the problem of disparity estimation from rectified images. We estimate a sequence of disparity maps by exploiting, for each of these maps, information from all the available views.

- We investigate the possibility of handling nonconvex similarity measures by adopting a discrete formulation of the disparity estimation problem which is exactly solved by a tight convex relaxation (Chapter 4). We show that the additional complexity incurred by our method in the multi-view context is small with respect to the stereo case.

Theoretical tools for dealing with divergence measures in optimization:

- φ-divergences are often restricted to problems where the minimization is performed w.r.t. one of the arguments of the divergence, or they are based on an alternating minimization process which requires specific assumptions to converge. In this work, we develop a novel proximal method that allows us to address more general forms of optimization problems, by computing the expressions of the proximity operators of two-variable convex functions. The expressions we have derived for these divergences enrich the list of functions for which such proximity operators can be easily calculated (Chapter 5).

- We have provided two different applications for divergence proximity operators in signal processing:
  ◇ We extended the Non Local Total Variation (NLTV) framework

by using information divergences to build new sparsity regularization measures for signal recovery.
◇ We have involved the divergence in the expression of the data fidelity term for image registration, more precisely for disparity estimation under illumination variation.

Publications

Journal paper:

1. C. Chaux, M. El Gheche, J. Farah, J.-C. Pesquet, and B. Pesquet-Popescu, A parallel proximal splitting method for disparity estimation from multicomponent images under illumination variation, in Journal of Mathematical Imaging and Vision, vol. 47, no. 3, pages 167-178, 2013.

Conference papers:

2. M. El Gheche, J.-C. Pesquet, J. Farah, M. Kaaniche and B. Pesquet-Popescu, Proximal splitting methods for depth estimation, in ICASSP 2011, Pages 853-856, Prague, Czech Republic, 22-27 May 2011.

3. M. El Gheche, J.-C. Pesquet, C. Chaux, J. Farah et B. Pesquet-Popescu, Méthodes proximales pour l'estimation du champ de disparité à partir d'une paire d'images stéréoscopiques en présence de variations d'illumination, GRETSI 2011, Bordeaux, France, 5-8 Sep. 2011.

4. M. El Gheche, C. Chaux, J.-C. Pesquet, J. Farah and B. Pesquet-Popescu, Disparity map estimation under convex constraints using proximal algorithms, in SIPS 2011, Pages 293-298, Beirut, Lebanon, 4-7 Oct. 2011.

5. M. El Gheche, J.-C. Pesquet and J. Farah, A proximal approach for optimization problems involving Kullback divergences, in ICASSP 2013, Pages 5984-5988, Vancouver, Canada, 26-31 May 2013.

6. M. El Gheche, A. Jezierska, J.-C. Pesquet and J. Farah, A proximal approach for signal recovery based on information measures, in EUSIPCO 2013, Marrakech, Moracco, 9-13 Sep. 2013.

Oral presentations:

M. El Gheche, C. Chaux and J.-C. Pesquet, Proximal methods in stereo vision, seminar in Ayazaga Campus of Istanbul Technical University, Turkey, 15 Oct. 2012.

M. El Gheche, C. Chaux, J.-C. Pesquet, J. Farah and B. Pesquet-Popescu, A parallel proximal splitting method for disparity estimation under illumination variation, Journée des doctorants, Champs sur Marne, France, 12 Jun. 2012.

C. Chaux, M. El-Gheche, J.-C. Pesquet and B. Pesquet-Popescu, Proximal methods for disparity map estimation, Symposium on Variation Image Analysis, Heidelberg, Allemagne, 11 Jul. 2011.

M. El Gheche, C. Chaux, J.-C. Pesquet, B. Pesquet-Popescu and J. Farah, Disparity estimation under convex constraints, EDST, Centre AZM pour la recherche en biotechnologies et applications, Lebanon, 26 Jul. 2011.

M. El Gheche, C. Chaux, J.-C. Pesquet, B. Pesquet-Popescu and J. Farah, Proximal splitting methods for depth estimation, Semaine conjointe des GDR MOA et MSPC, La Londe les Maures, France, 6-10 Jun. 2011.

Organization and Summary

This thesis is organized as follows. Chapter 2 introduces the reader to the topic of convex optimization and the associated algorithmic solutions.

Then, Chapter 3 addresses the dense disparity estimation problem from multicomponent images, formulated as an inverse problem by minimizing a convex objective function under multiple convex constraints. These constraints arise from prior knowledge and rely on various properties of the field to be estimated. We address the problem of disparity estimation under varying illumination conditions, and develop a spatially varying multiplicative model that accounts for photometric changes between both images in the stereo pair. Results on synthetic and real stereo pairs demonstrate the efficiency of the proposed method in recovering illumination changes and disparity map jointly, making disparity estimation very robust w.r.t. such changes.

For large scale or complex scenes, we address, in Chapter 4, the problem of multi-view disparity estimation from rectified images. A sequence of disparity maps is estimated by using two different approaches, which differ in the type of convex relaxation employed. For each of these maps, information from all the available views are exploited without adding much computational cost with respect to the two-view case.

In Chapter 5, we focus on convex optimization problems involving information measures (φ-divergence), which have been extensively investigated in source and channel coding. These measures can also be successfully used in inverse problems encountered in signal and image processing. The related

optimization problems are often challenging due to their large size. We derive closed-form expressions of the proximity operators of these divergences. Building upon these results, we develop an efficient proximal approach. This allows us to address, in Chapter 6, two applications of convex optimization problems involving one of these divergences.

Finally, we draw some conclusions and perspectives in Chapter 7.

- Chapter 2 -

Convex optimization background

"Everything is ~~possible~~ convex."
P. L. Combettes and J.-C. Pesquet

Optimization constitutes a wide and rich mathematical field. Many applications in computer design or analysis, in particular for real-time reactive or automatic systems, rely on a wealth of these thriving tools. In recent years, the theory underlying the computational techniques in convex and nonconvex optimization have been studied by many authors. However, the elegant and dynamical language of convex analysis enlights much more the potential of optimization theory. Convex analysis, its extensions and its applications have known many developments in the recent years.

This chapter mainly focuses on the convex formulation of inverse problems. In this part, we aim at describing the tools relevant to our studies. Firstly, we begin with a brief overview of inverse problems addressed by variational approaches. Then, we provide a brief presentation of convex optimization algorithms. Finally, we discuss different methods for computing the projection onto a convex set.

§ 2.1 INVERSE PROBLEMS

In many optimization problems, one is interested in measuring a physical signal which is not directly available. In this context, we often classify the problems along two axes:

- Direct problem: having a complete description of a physical system, we can estimate the outcome of some measurements. One can say it is a problem of inferring the consequences from given causes.
- Inverse problem: finding the unknown causes of known consequences (it is associated with the reversal of the cause-effect relationship).

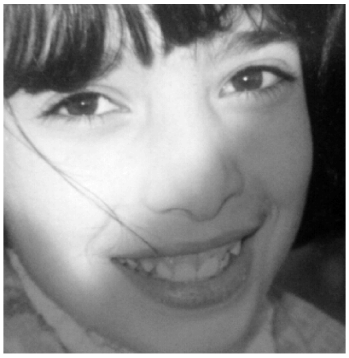
Using the actual result of some measurements, we are able to infer the values of the parameters that characterize the system Bertero and Boccacci [1998]. The inherent inverse problem consists in recovering the target signal by making explicit any available a priori information on the model parameters. There exist many applications to inverse problems which can be found in many fields such as satellite image restoration Facciolo et al. [2009]; Hajlaoui et al. [2010], microscopy image deconvolution Dupé et al. [2009]; Jezierska et al. [2012], computed tomography Vandeghinste et al. [2011], machine learning Bach et al. [2012]; Theodoridis et al. [2011], stereo vision Deriche et al. [1996]; Miled et al. [2009a]; Pock et al. [2010], audio processing Kowalski et al. [2010]; Akyildiz and Bayram [2012]...

A popular inverse problem consists in recovering a signal (e.g. an image) which is as close as possible to a reference one. Because of sensor imperfections and acquisition errors, the measured data are often noisy and degraded by a linear operator. Let us consider the basic approach in restoration where the direct problem can be formulated as follows:

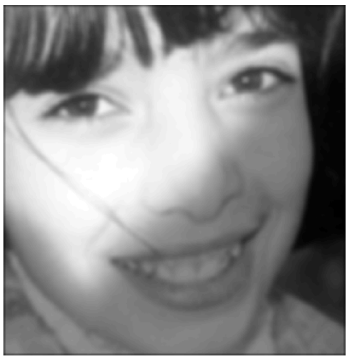
$$z = T\overline{x} + w, \tag{2.1}$$

where $z = (z_j)_{1\leq j\leq K} \in \mathbb{R}^K$ is the degraded observation and $\overline{x} = (\overline{x}_i)_{1\leq i\leq N} \in \mathbb{R}^N$ is the image to be recovered. The linear operator $T = (T_{j,i})_{1\leq j\leq K, 1\leq i\leq N} \in \mathbb{R}^{K\times N}$ describes the physical laws linking $\overline{x}$ to the measurements z (e.g. motion blur, defocussing blur, ...) and the components of $w = (w_j)_{1\leq j\leq K} \in \mathbb{R}^K$ are realizations of a mutually independent zero-mean (for example Gaussian) noise process. Solving the inverse problem associated with (2.1) aims at recovering $\overline{x}$ from z. One of the most well-known approaches in this context is the least squares method Gauss [1820]; Kay [1998]. It consists in minimizing the quadratic distance between the observation and the image to be recovered:

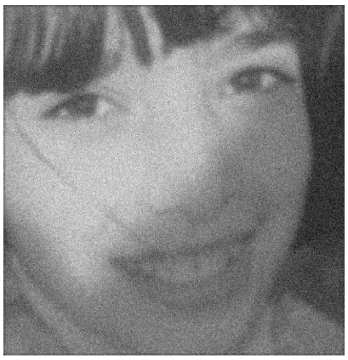
$$\underset{x\in\mathbb{R}^N}{\operatorname{argmin}} \ \|Tx - z\|^2. \tag{2.2}$$



(a) *Original image* $\overline{x}$



(b) *Blury image* $T\overline{x}$*:* T *of size* 7×7



(c) *Noisy image* z*: with zero-mean Gaussian noise and variance=400*

Figure 2.1: *Example of a degraded image resulting from the direct problem:* $z = T\overline{x} + w$.

The solution to Problem (2.2) may be highly unstable and sensitive to small changes in the data. In such a case, we can certainly see why an inverse problem would be "difficult". A more precise way to look at this might be defined by how "solvable" a given problem is, which leads to the notion of well-posed and ill-posed problems proposed by Hadamard [1902].

Well-posed problems Hadamard proposed three properties that a problem must possess in order to be classified as a well-posed one: existence, uniqueness and stability. For our purposes, considering only finite dimensional linear operators, we can think of data and solutions as the input and output vectors of some linear transformations. The first two properties seem rather obvious: a solution $\hat{x}$ exists for each observation z and this solution is unique. The third property is a question of stability, requiring that small changes to the input do not produce arbitrarily large changes to the output. A problem that lacks any one of these properties is by definition *ill-posed*. Thus, to solve the problem in a reliable manner, it is necessary to incorporate

additional information Tikhonov [1963]. One of the first approaches to get a stable and satisfactory solution consists in adding a quadratic penalization term in order to achieve a small energy:

$$\underset{x\in\mathbb{R}^N}{\operatorname{argmin}} \ \|Tx - z\|^2 + \lambda\|x\|^2, \tag{2.3}$$

where $\lambda > 0$ is the regularization parameter which balances the solution between the data fidelity term $\|Tx - z\|^2$ and the penalization $\|x\|^2$. A more general formulation of the regularized problem will be presented afterwards.

2.1.1 Data fidelity and prior information

Since many inverse problems are inherently ill-posed, a Bayesian approach may help in finding a stable and satisfactory solution. Let x be a realization of a random variable X with prior probability density function $\mathcal{P}_X$ and z a realization of a random variable Z with conditional likelihood $\mathcal{P}_{Z|X=x}$. One aims at finding the signal x having the maximum *a posterior* probability density function $\mathcal{P}_{X|Z=z}(x)$, given the observation z, i.e.

$$x^* = \arg\max_x \ \mathcal{P}_{X|Z=z}(x). \tag{2.4}$$

Using Bayes rule, the posterior probability density function can be decomposed into likelihood and prior probability measures :

$$x^* = \arg\max_x \ \mathcal{P}_{Z|X=x}(z)\,\mathcal{P}_X(x). \tag{2.5}$$

Taking the minus-logarithm of the above expression, eventually leads to the following minimization problem:

$$\underset{x}{\text{minimize}} \ \underbrace{-\log \mathcal{P}_{Z|X=x}(z)}_{g(x)} \ + \ \underbrace{(-\log \mathcal{P}_X(x))}_{\Psi(x)}, \tag{2.6}$$

where the function g is the *data fidelity* term and the function Ψ is the *prior* term. Note that the fidelity term typically presents a separable form, as it is common to assume the statistical conditional independence of the components of vector Z:

$$g(x) = \sum_{i=1}^{N} g^{(i)}(x, z_i). \tag{2.7}$$

The terms $g^{(i)}(x, z_i)$ may be explicitly defined by a particular likelihood distribution $\mathcal{P}_{Z_i|X=x}(z_i)$ (as in denoising problems), or they may be implicitly derived from local matching scores (as in disparity estimation problems).

From a practical standpoint, it is important to balance the importance of the data fidelity term g against the regularization Ψ, as it has an impact over

the signal to be recovered. A possible solution is to introduce a regularization constant $\lambda \in]0, +\infty[$ into the minimization problem, leading to the following *unconstrained* formulation:

$$\underset{x}{\text{minimize}} \quad g(x) + \lambda\Psi(x). \tag{2.8}$$

The alternative solution is to enforce an upper bound on the function g or Ψ, yielding the following alternative *constrained* formulations:

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad g(x) \quad \text{s. t.} \quad \Psi(x) \leq \tau \\ (\text{resp. } &\underset{x}{\text{minimize}} \quad \Psi(x) \quad \text{s. t.} \quad g(x) \leq \eta), \end{aligned} \tag{2.9}$$

where τ and η are some positive constants which are chosen depending on some prior information. Indeed, under technical assumptions, both formulations are equivalent for some specific values of λ and τ (resp. η). However, the constrained formulation may be considered more practical, as the solution is less sensitive to the choice of regularization parameter λ, and even the choice of τ (resp. η) may be easier than that of λ Combettes and Trussell [1991]; Chierchia et al. [2013]; Afonso et al. [2011]. However, it is worth noting that there exists well-known relationships between the constrained formulation and the regularized one Ciak et al. [2012]; Teuber et al. [2013]. The dual problem associated with (2.8) often has an interesting interpretation in terms of the original problem, and sometimes leads to efficient or distributed methods for solving it.

Convex data fidelity term In general, the function g may be either convex or nonconvex. Unfortunately, finding a global minimum to Problems (2.8) and (2.9), when g is nonconvex, turns out to be a very hard problem. In some cases, good results may be obtained by local minimization, starting from a good initialization, using for example PDEs (Partial Differential Equations) or other variational approaches. However, such methods cannot guarantee any form of quality of the result, as they may get stuck in local minima.

On the other hand, when g is convex, there exist efficient numerical methods to solve the above problems with guaranteed convergence (under weak conditions). Consequently, a variety of convex relaxation techniques have been developed to make the function g convex: first order Taylor approximation Miled et al. [2009b], convex-hull approximation Jiang et al. [2007], exact relaxation for binary problems Chan and Nikolova [2006], exact relaxation for multi-label problems Pock et al. [2010], and semidefinite programming Lieven and Stephen [1995].

Convex regularization: The aim of regularization is to convey some prior knowledge about the signal to be recovered, in order to make the original problem well-posed. A priori knowledge should not reflect any specific measurements of the physical signal which is currently available and analyzed. In this regard, the more carefully the regularity is modeled (in terms of energy to be minimized), the better the quality of the estimated signal. Since the signals of interest are often piecewise smooth, popular regularization models tend to enforce a sparsity condition by penalizing an analysis representation of the target signal. In this context, Total Variation (TV) Rudin et al. [1992]; Malgouyres [2002]; Peyré and Fadili [2011]; Bayram and Kamasak [2012] has emerged as a simple, yet successful, regularization, consisting in penalizing the gradient coefficients. However, TV fails to preserve textures, details, and fine structures, because they are hardly distinguishable from noise. To improve this behaviour, the TV model has been extended by using a non-locality principle Gilboa and Osher [2009]; Werlberger et al. [2010]; Peyré [2011]; Chierchia et al. [2013]; Couprie et al. [2013]. Another approach to overcome these limitations is to replace the gradient operator with a frame-based transformation which may yield a more suitable sparse representation of the signals of interest Mallat [1997]. The connections between these two different approaches have been studied in Cai et al. [2012].

- Introducing a Total Variation regularization has an effect on smoothing homogeneous regions in the image while preserving edges. Initially introduced in image processing by Rudin et al. [1992], the TV regularity measure has been proven to be very useful in image recovery and denoising problems Combettes and Pesquet [2004]; Aujol [2009] and variational stereo methods Miled et al. [2009a].
  Let $\widehat{\nabla^{(1)}}$ and $\widehat{\nabla^{(2)}}$ denote discrete horizontal and vertical gradients (for example, obtained by cyclic convolutions Combettes and Pesquet [2008]; Pustelnik et al. [2011]). Then, a discrete version of the Total Variation is the following one:

$$(\forall x \in \mathbb{R}^N) \quad \mathrm{TV}(x) = \sum_{s \in \mathcal{D}} \sqrt{|\widehat{\nabla^{(1)}} x(\mathbf{s})|^2 + |\widehat{\nabla^{(2)}} x(\mathbf{s})|^2} \tag{2.10}$$

  where $\mathcal{D} \subset \mathbb{Z}^2$ is the image domain and $\mathbf{s} \in \mathbb{R}^2$ is the spatial position in the image x, here column-wise reshaped as an N-dimensional vector. The associated constraint set is

$$C = \{x \in \mathbb{R}^N | \, \mathrm{TV}(x) \leq \tau\} \tag{2.11}$$

  where τ can be computed based on some prior information on the target solution (or approximated from an initial estimate). The projection onto the convex set has been solved with various methods Van Den Berg and Friedlander [2008]; Weiss et al. [2009]; Fadili and Peyré [2011] and in particular with epigraphical ones Chierchia et al. [2013].

- While TV is a simple convex optimization tool, it fails to preserve textures. To improve this behaviour, the TV model has been extended in Gilboa and Osher [2009] by relying on the non-locality principle introduced in Buades et al. [2006].

$$\mathrm{NLTV}(x) = \sum_{s\in\mathcal{D}} \sqrt{\sum_{n\in\mathcal{W}_s} \omega_{s,n}|x(\mathbf{s}) - x(n)|^2}, \tag{2.12}$$

where $\mathcal{W}_s$ is the *neighbourhood support* of $\mathbf{s}$, that is a subset of positions in $\mathcal{D}\backslash\{\mathbf{s}\}$ located in a relatively large $Q\times Q$ window centered at $\mathbf{s}$, where $Q \in \mathbb{N}$ is odd and $\omega_{s,n}$ are some weights in $[0, +\infty[$. The considered NLTV constraint is expressed as

$$C' = \{x \in \mathbb{R}^N | \, \mathrm{NLTV}(x) \leq \tau'\} \tag{2.13}$$

For computing the projection onto the convex set C', specific numerical methods Chierchia et al. [2013] have been recently developed.

§ 2.2 Algorithms in convex optimization

Notation: In the following, we denote by $\|\cdot\|$ the standard Euclidean norm, and by Id the identity matrix. The domain of a function $f : \mathbb{R}^N \to]-\infty, +\infty]$ is $\mathrm{dom}\, f = \{x \in \mathbb{R}^N | f(x) < +\infty\}$. $\Gamma_0(\mathbb{R}^N)$ is the class of lower semi-continuous convex functions from $\mathbb{R}^N$ to $]-\infty, +\infty]$ such that $\mathrm{dom}\, f \neq \emptyset$. Let $f \in \Gamma_0(\mathbb{R}^N)$. The conjugate of f is the function $f^* \in \Gamma_0(\mathbb{R}^N)$ defined by

$$f^* : \mathbb{R}^N \to]-\infty, +\infty] : u \mapsto \sup_{x\in\mathbb{R}^N} x^\top u - f(x) \tag{2.14}$$

and the subdifferential of f is the set-valued operator

$$\partial f : \mathbb{R}^N \to 2^{\mathbb{R}^N} : x \mapsto \{u \in \mathbb{R}^N | (\forall y \in \mathbb{R}^N)(y-x)^\top u + f(x) \leq f(y)\}. \tag{2.15}$$

When f is Gâteaux-differentiable at $y \in \mathbb{R}^N$, $\partial f(y) = \{\nabla f(y)\}$ where $\nabla f(y)$ is the gradient of f at y. A differentiable convex function has β-Lipschitz continuous gradient ∇f if

$$(\forall (x, y) \in \mathbb{R}^N \times \mathbb{R}^N) \quad \|\nabla f(x) - \nabla f(y)\| \leq \beta\|x - y\|, \tag{2.16}$$

where $\beta \in]0, +\infty[$.
Let C be a nonempty subset of $\mathbb{R}^N$. The indicator function of C is

$$\iota_C : \mathbb{R}^N \to]-\infty, +\infty] : x \mapsto \begin{cases} 0, & \text{if } x \in C \\ +\infty, & \text{if } x \notin C. \end{cases} \tag{2.17}$$

Let S be a convex subset of $\mathbb{R}^N$. The *relative interior* of S, i.e. the set of points $x \in S$ such that the cone generated by $-x+S$ is a vector subspace of $\mathbb{R}^N$, is denoted by $\operatorname{ri} S$.
The epigraph of $f \in \Gamma_0(\mathbb{R}^N)$ is the nonempty closed convex subset of $\mathbb{R}^N \times \mathbb{R}$ defined as

$$\operatorname{epi} f = \{(x,\zeta) \in \mathbb{R}^N \times \mathbb{R} \mid f(x) \leq \zeta\}, \tag{2.18}$$

and the lower level set of f at height $\zeta \in \mathbb{R}$ is the nonempty closed convex subset of $\mathbb{R}^N$ defined as

$$\operatorname{lev}_{\leq \zeta} f = \{x \in \mathbb{R}^N \mid f(x) \leq \zeta\}. \tag{2.19}$$

Let $f \in \Gamma_0(\mathbb{R}^N)$ and $g \in \Gamma_0(\mathbb{R}^N)$. We denote by $f \,\square\, g$ the infimal convolution of f and g

$$f \,\square\, g : \mathbb{R}^N \to [-\infty,+\infty] : x \mapsto \inf_{y \in \mathbb{R}^N} f(y) + g(x-y). \tag{2.20}$$

Problem formulation and algorithms

By generalizing (2.6), a large range of inverse problems can be formulated under the following form:

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \quad h(x) + \sum_{r=1}^{R} g_r(T_r x) \quad \text{s.t.} \quad \begin{cases} H_1 x \in C_1, \\ \vdots \\ H_S x \in C_S, \end{cases} \tag{2.21}$$

where, for every $r \in \{1,\dots,R\}$, $T_r \in \mathbb{R}^{K_r \times N}$ and $g_r \in \Gamma_0(\mathbb{R}^{K_r})$, $h\colon \mathbb{R}^N \mapsto (-\infty,+\infty]$ is a convex differentiable function with a μ-Lipschitzian gradient for some $\mu \in (0,+\infty)$ and, for every $s \in \{1,\dots,S\}$, $H_s \in \mathbb{R}^{M_s \times N}$ and $C_s \subset \mathbb{R}^{M_s}$ is a nonempty closed convex subset.

Gradient descent The first methods for finding a solution to an inverse problem were restricted to the use of a differentiable cost function Tikhonov [1963], i.e. Problem (2.21) where $R = 0$ and $S = 0$. In this context, gradient-based algorithms, e.g. nonlinear conjugate gradient or quasi-Newton methods, are popular (see Chouzenoux et al. [2011] and references therein). However, in order to model additional properties, sparsity promoting penalizations ($R \geq 1$) or hard constraints ($S \geq 1$) may be introduced, and the differentiability property is not satisfied anymore. One way to circumvent this difficulty is to resort to smart approximations in order to smooth the involved non-differentiable functions Aubert and Tahraoui [1980]; Ben-Tal and Teboulle [1989]; Hiriart-Urruty and Lemaréchal [1996]. If one wants to address the original nonsmooth problem without introducing approximation errors, one

may apply some specific algorithms, e.g. Gauss-Seidel or Uzawa methods, the convergence of which is guaranteed under restrictive assumptions Tseng [2001]. Interior point methods Nesterov and Nemirovskii [1994], Wright [1997] can also be employed for small to medium size optimization problems.

Projection methods In order to circumvent the limitations of the aforementioned optimization techniques, some iterative methods were developed to solve Problem (2.21) with $R = 0$ and $h = 0$, i.e. in finding a vector belonging to the intersection of convex sets. The *projection onto convex sets* algorithm (POCS) is one of the most popular approaches to solve convex feasibility problems for data recovery Bregman [1965]; Gurin et al. [1967]; Youla and Webb [1982]; Combettes [1993]. A drawback of POCS is that it is not well-suited for parallel implementations. The Parallel Projection Method (PPM) and Method of Parallel Projections (MOPP) are variants of POCS making use of parallel projections. Moreover, these algorithms were designed to efficiently solve inconsistent feasibility problems (e.g. when the intersection of the convex sets is empty). Thorough comparisons between projection methods have been performed in Combettes [1997]; Censor et al. [2012].

Proximal methods Recently, a new class of iterative algorithms has emerged in order to efficiently solve Problem (2.21) in its general form. The key tool in these approaches is the *proximity operator* Moreau [1965] of a function $\varphi \in \Gamma_0(\mathbb{R}^N)$, defined as

$$(\forall y \in \mathbb{R}^N) \qquad \operatorname{prox}_{\varphi}(y) = \underset{u \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|u - y\|^2 + \varphi(u). \tag{2.22}$$

The proximity operator can be interpreted as a sort of subgradient step for the function φ, as $p = \operatorname{prox}_{\varphi}(y)$ is uniquely defined through the inclusion $y - p \in \partial\varphi(p)$ Combettes and Pesquet [2011]. This explains why proximity operators, yielding first order methods, are useful in optimization. As a matter of fact, the proximity operator is the cornerstone of so-called *proximal methods*.

Proximity operators enjoy many interesting properties Chaux et al. [2007]; Combettes and Pesquet [2011]. In particular, they generalize the notion of projection onto a nonempty closed convex subset C of $\mathbb{R}^N$, in the sense that

$$(\forall y \in \mathbb{R}^N) \qquad \operatorname{prox}_{\iota_C} y = P_C(y) = \underset{u \in C}{\operatorname{argmin}} \ \|u - y\|. \tag{2.23}$$

Hence, proximal methods provide a unifying framework that allows one to address nonsmooth penalizations ($R \geq 1$) and hard constraints ($S \geq 1$).

The key feature of proximal methods is that they proceed by *splitting*. The operators $(\operatorname{prox}_{g_r})_{1 \leq r \leq R}$, $(T_r)_{1 \leq r \leq R}$, $(P_{C_s})_{1 \leq s \leq S}$ and $(H_s)_{1 \leq s \leq S}$ are used

individually to solve Problem (2.21). This yields easily implementable algorithms, which circumvent the problem of computing the proximity operator of a sum of functions or a function composed with a linear operator, for which closed-form expressions are rarely available.

Proximal algorithms can be classified in primal methods Daubechies et al. [2004]; Chaux et al. [2007]; Combettes and Pesquet [2007b]; Figueiredo et al. [2007]; Beck and Teboulle [2009]; Fornasier and Schönlieb [2009]; Steidl and Teuber [2010]; Combettes and Pesquet [2011]; Pesquet and Pustelnik [2012] and primal-dual methods Chen and Teboulle [1994]; Esser et al. [2010]; Chambolle and Pock [2011]; Briceño-Arias and Combettes [2011]; Combettes and Pesquet [2012]; Vũ [2013]; Condat [2013]. The main difference is that primal algorithms generally require the computation of the inverse of the operator $\sum_{r=1}^{R} T_r^\top T_r + \sum_{s=1}^{S} H_s^\top H_s$ (see Algorithm 1), while primal-dual ones only require the calculation of $(T_r)_{1\leq r\leq R}, (H_s)_{1\leq s\leq S}$ and their adjoints (see Algorithm 2). Consequently, primal-dual methods are often easier to implement than primal ones, but their convergence may be slower Pustelnik et al. [2012]; Couprie et al. [2013].

2.2.0.1 Primal algorithms

Primal methods aim at directly solving Problem (2.21), without resorting to any duality arguments. In this context, forward-backward splitting and Douglas-Rachford splitting are popular, although they are limited to solving particular instances of Problem (2.21). Nonetheless, they are the building blocks of more sophisticated algorithms (such as PPXA+ and M+LFBF) that allow one to solve Problem (2.21) in its general form.

Forward-Backward splitting Let $R = 1$, $T_1 = \mathrm{Id}$ and $S = 0$ in (2.21). Then the problem becomes

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; h(x) + g(x). \tag{2.24}$$

Forward-backward splitting aims at finding a point $\bar{x} \in \mathbb{R}^N$ satisfying the condition.[1]

$$0 \in \partial g(\bar{x}) + \nabla h(\bar{x}) \quad \Leftrightarrow \quad \bar{x} = \mathrm{prox}_g\big(\bar{x} - \nabla h(\bar{x})\big). \tag{2.25}$$

Indeed, under the assumption that $\lim_{\|x\|\to+\infty} g(x) + h(x) = +\infty$, it can be shown Combettes and Wajs [2005] that the sequence $(x^{[n]})_{n\in\mathbb{N}}$ defined as

$$x^{[n+1]} = x^{[n]} + \lambda_n\Big(\mathrm{prox}_{\gamma_n g}\big(x^{[n]} - \gamma_n \nabla h(x^{[n]})\big) - x^{[n]}\Big) \tag{2.26}$$

converges to a solution to Problem (2.24) for $\gamma_n \in \,]0, 2/\mu[$ and $\lambda_n \in \,]0, 1]$.

[1] $0 \in \partial g(\bar{x}) + \nabla h(\bar{x}) \;\Leftrightarrow\; \bar{x} - \nabla h(\bar{x}) - \bar{x} \in \partial g(\bar{x}) \;\Leftrightarrow\; \bar{x} = \mathrm{prox}_g\big(\bar{x} - \nabla h(\bar{x})\big)$

Douglas-Rachford splitting Let $h = 0$, $R = 2$, $T_1 = T_2 = \mathrm{Id}$ and $S = 0$ in (2.21). Then the problem reads

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; g_1(x) + g_2(x). \tag{2.27}$$

Douglas-Rachford splitting aims at finding some points $\bar{x}, \bar{u} \in \mathbb{R}^N$ that satisfy the condition.[2]

$$0 \in \partial g_1(\bar{x}) + \partial g_2(\bar{x}) \;\Leftrightarrow\; \begin{cases} \bar{u} - \bar{x} \in \partial g_2(\bar{x}) \\ \bar{x} - \bar{u} \in \partial g_1(\bar{x}) \end{cases} \;\Leftrightarrow\; \begin{cases} \bar{x} = \mathrm{prox}_{g_2}(\bar{u}) \\ 0 = \mathrm{prox}_{g_1}(2\bar{x} - \bar{u}) - \bar{x}. \end{cases} \tag{2.28}$$

As a matter of fact, under the assumption $(\mathrm{ri}\,\mathrm{dom}\, g_1) \cap (\mathrm{ri}\,\mathrm{dom}\, g_2) \neq \emptyset$, it can be shown Combettes and Pesquet [2007b] that the sequence $(x^{[n]})_{n\in\mathbb{N}}$ defined as

$$\begin{cases} x^{[n]} &= \mathrm{prox}_{\gamma g_2}(u^{[n]}), \\ u^{[n+1]} &= u^{[n]} + \lambda_n \Big(\mathrm{prox}_{\gamma g_1}(2\,x^{[n]} - u^{[n]}) - x^{[n]} \Big), \end{cases} \tag{2.29}$$

converges to a solution to Problem (2.27) for $\lambda_n \in \left]0, 2\right[$ and $\gamma > 0$.

PPXA+ Let $h = 0$ in Problem (2.21). Then, we have to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; \sum_{r=1}^{R} g_r(T_r x) + \sum_{s=1}^{S} \iota_{C_s}(H_s x). \tag{2.30}$$

Let $L = [T_1^\top \ldots T_R^\top \;\; H_1^\top \ldots H_S^\top]^\top \in \mathbb{R}^{Q\times N}$ with $Q = \sum_{r=1}^{R} K_r + \sum_{s=1}^{S} M_s$,

$$\big(\forall v = [v_1^\top \;\ldots\; v_{R+S}^\top]^\top \in \mathbb{R}^Q\big) \qquad g(v) = \sum_{r=1}^{R} g_r(v_r) + \sum_{s=1}^{S} \iota_{C_s}(v_{R+s}), \tag{2.31}$$

and $D = \big\{ v \in \mathbb{R}^Q \;\big|\; v = Lx,\; x \in \mathbb{R}^N \big\}$. Then, Problem (2.30) can be cast over the Douglas-Rachford framework by setting $g_1 = \iota_D$ and $g_2 = g$,

$$\underset{v\in\mathbb{R}^Q}{\text{minimize}} \;\; \iota_D(v) + g(v). \tag{2.32}$$

Moreover, it can be shown Combettes and Pesquet [2008] that,[3]

$$(2.29) \quad\Leftrightarrow\quad \begin{cases} v^{[n]} &= \mathrm{prox}_{\gamma g}(u^{[n]}) \\ \widehat{v}^{[n]} &= P_D(v^{[n]}) \\ u^{[n+1]} &= u^{[n]} + \lambda_n \left(2\,\widehat{v}^{[n]} - \widehat{u}^{[n]} - v^{[n]}\right) \\ \widehat{u}^{[n+1]} &= \widehat{u}^{[n]} + \lambda_n \left(\widehat{v}^{[n]} - \widehat{u}^{[n]}\right), \end{cases} \tag{2.33}$$

[2] $x - u \in \partial g_1(x) \;\Leftrightarrow\; (2x - u) - x \in \partial g_1(x) \;\Leftrightarrow\; x = \mathrm{prox}_{g_1}(2x - u)$

[3] In (2.29) $P_D(2\,v^{[n]} - u^{[n]}) = 2\,\widehat{v}^{[n]} - \widehat{u}^{[n]}$, with $\widehat{v}^{[n]} = P_D(v^{[n]})$ and $\widehat{u}^{[n]} = P_D(u^{[n]}) = P_D\big(u^{[n-1]} + \lambda_n\,(2\,\widehat{v}^{[n-1]} - \widehat{u}^{[n-1]} - v^{[n-1]})\big) = \widehat{u}^{[n-1]} + \lambda_n(\widehat{v}^{[n-1]} - \widehat{u}^{[n-1]})$.

and the sequence $(\widehat{u}^{[n]})_{n\in\mathbb{N}}$ converges to a solution of (2.32). Furthermore, the constraint D enforces $(\widehat{u}^{[n]})_{n\in\mathbb{N}} = (L\,x^{[n]})_{n\in\mathbb{N}}$, for some sequences $(x^{[n]})_{n\in\mathbb{N}}$. When $L^\top L$ is invertible, it can be proved Pesquet and Pustelnik [2012] that $(x^{[n]})_{n\in\mathbb{N}}$ converges to a solution of (2.30) and that $P_D(v^{[n]}) = L\,\widehat{x}^{[n]}$ with $\widehat{x}^{[n]} = (L^\top L)^{-1}L^\top v^{[n]}$. As a result, if there exists an $\bar{x} \in \mathbb{R}^N$ such that $L\bar{x} \in \operatorname{ri}\operatorname{dom} g$, the sequence $(x^{[n]})_{n\in\mathbb{N}}$ generated by

$$\begin{cases} v^{[n]} &= \operatorname{prox}_{\gamma g}(u^{[n]}) \\ \widehat{x}^{[n]} &= (L^\top L)^{-1}L^\top v^{[n]} \\ u^{[n+1]} &= u^{[n]} + \lambda_n \left(L(2\,\widehat{x}^{[n]} - x^{[n]}) - v^{[n]}\right) \\ x^{[n+1]} &= x^{[n]} + \lambda_n\,(\widehat{x}^{[n]} - x^{[n]}), \end{cases} \tag{2.34}$$

converges to a solution to Problem (2.30). The detailed iterations are illustrated in Algorithm 1. Such an algorithm can be considered as a generalization of numerous methods Fortin and Glowinski [1983]; Eckstein [1994]; Goldstein and Osher [2009]; Attouch and Soueycatt [2009]; Setzer et al. [2010], among which *augmented Lagrangian* techniques.

Algorithm 1 PPXA+ Pesquet and Pustelnik [2012] for Problem (2.21).

$\gamma \in]0,+\infty[, \quad (\lambda_n)_{n\in\mathbb{N}} \subset]0,2[$
$(\forall r \in \{1,\dots,R\}) \quad u_r^{[0]} \in \mathbb{R}^{K_r}$
$(\forall s \in \{1,\dots,S\}) \quad u_{R+s}^{[0]} \in \mathbb{R}^{M_s}$
$x^{[0]} \in \mathbb{R}^N$
for $n = 0, 1, \dots$
$\quad (\forall r \in \{1,\dots,R\}) \quad v_r^{[n]} = \operatorname{prox}_{\gamma g_r}(u_r^{[n]})$
$\quad (\forall s \in \{1,\dots,S\}) \quad v_{R+s}^{[n]} = P_{C_s}(u_{R+s}^{[n]})$
$\quad \widehat{x}^{[n]} = \Big(\sum_{r=1}^{R} T_r^\top T_r + \sum_{s=1}^{S} H_s^\top H_s\Big)^{-1}\Big(\sum_{r=1}^{R} T_r^\top v_r^{[n]} + \sum_{s=1}^{S} H_s^\top v_{R+s}^{[n]}\Big)$
$\quad (\forall r \in \{1,\dots,R\}) \quad u_r^{[n+1]} = u_r^{[n]} + \lambda_n \left(T_r(2\,\widehat{x}^{[n]} - x^{[n]}) - v_r^{[n]}\right)$
$\quad (\forall s \in \{1,\dots,S\}) \quad u_{R+s}^{[n+1]} = u_{R+s}^{[n]} + \lambda_n \left(H_s(2\,\widehat{x}^{[n]} - x^{[n]}) - v_{R+s}^{[n]}\right)$
$\quad x^{[n+1]} = x^{[n]} + \lambda_n\,(\widehat{x}^{[n]} - x^{[n]})$

2.2.0.2 Primal-dual algorithms

The scope of primal methods is somewhat limited, as they require to determine the inverse of some linear operators (see Algorithm 1), which turns out to be burdensome or even impossible to compute in many practical cases. Primal-dual algorithms overcome the inversion issue by resorting to the Fenchel-Rockafellar duality framework. In Problem (2.21), let us add a

function $f \in \Gamma_0(\mathbb{R}^N)$, let $L = [T_1^\top \dots T_R^\top \;\; H_1^\top \dots H_S^\top]^\top \in \mathbb{R}^{Q\times N}$ and define $g \in \Gamma_0(\mathbb{R}^Q)$ as in (2.31). Instead of seeking a solution to the primal problem

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; f(x) + g(Lx) + h(x), \tag{2.35}$$

or a solution to the dual problem

$$\underset{v\in\mathbb{R}^Q}{\text{minimize}} \qquad g^*(v) + (f^* \,\square\, h^*)(-L^\top v), \tag{2.36}$$

one aims at finding a saddle-point solution to the primal-dual problem

$$\min_{x\in\mathbb{R}^N} \max_{v\in\mathbb{R}^Q} \quad \underbrace{f(x) - g^*(v)}_{F_1(x,v)} + \underbrace{v^\top Lx + h(x)}_{F_2(x,v)}. \tag{2.37}$$

The classical Karush-Kuhn-Tucker theory asserts that a solution to the primal-dual problem is a pair $(\bar{x}, \bar{v})$ satisfying the inclusion,[4]

$$\begin{cases} 0 \in \partial f(\bar{x}) + L^\top \bar{v} + \nabla h(\bar{x}) \\ 0 \in \partial g^*(\bar{v}) - L\bar{x} \end{cases} \quad \Leftrightarrow \quad \begin{cases} \bar{x} = \operatorname{prox}_f\big(\bar{x} - L^\top \bar{v} - \nabla h(\bar{x})\big) \\ \bar{v} = \operatorname{prox}_{g^*}(\bar{v} + L\bar{x}), \end{cases} \tag{2.38}$$

where $\bar{x}$ and $\bar{v}$ are solutions, respectively, to the primal and dual problems.

The primal-dual inclusion in (2.38) falls in the forward-backward framework. Hence, it can be solved by resorting to the classical forward-backward splitting method or to the forward-backward-forward splitting method Tseng [1998]. The detailed iterations associated with a primal-dual algorithm based on Tseng's splitting are illustrated in Algorithm 2.

§ 2.3 Projections

As mentioned before, computing the projection P_C onto a nonempty closed convex subset C of a real Hilbert space $\mathbb{R}^N$ requires to solve a constrained quadratic minimization problem:

$$(\forall y \in \mathbb{R}^N) \qquad P_C(y) = \underset{u\in C}{\operatorname{argmin}} \;\; \|u - y\|. \tag{2.39}$$

The distance to C of every point $y \in \mathbb{R}^N$ is then given by $d_C(y) = \|y - P_C(y)\|$. However, it turns out that a closed form expression of the solution to (2.39) is available in a limited number of instances.

Such well-known examples are the projections onto a closed half-space and a hyperslab:

[4] $0 \in \partial g^*(\bar{v}) - L\bar{x} \;\Leftrightarrow\; (\bar{v} + L\bar{x}) - \bar{v} \in \partial g^*(\bar{v}) \;\Leftrightarrow\; \bar{v} = \operatorname{prox}_{g^*}(\bar{v} + L\bar{x})$.

Algorithm 2 M+LFBF Combettes and Pesquet [2012] for Problem (2.21).

$\beta = \mu + \left(\sum_{r=1}^{R}\|T_r\|^2 + \sum_{s=1}^{S}\|H_s\|^2\right)^{1/2}$
$(\gamma_n)_{n\in\mathbb{N}} \subset]0, \beta^{-1}[$
$(\forall r \in \{1,\dots,R\})\quad u_r^{[0]} \in \mathbb{R}^{K_r}$
$(\forall s \in \{1,\dots,S\})\quad v_s^{[0]} \in \mathbb{R}^{M_s}$
$x^{[0]} \in \mathbb{R}^N$
`for` $n = 0, 1, \dots$
$p^{[n]} = \nabla h(x^{[n]}) + \sum_{r=1}^{R} T_r^\top u_r^{[n]} + \sum_{s=1}^{S} H_s^\top v_s^{[n]}$
$\widehat{x}^{[n]} = \operatorname{prox}_{\gamma_n f}\left(x^{[n]} - \gamma_n p^{[n]}\right)$
$(\forall r \in \{1,\dots,R\})\quad \widehat{u}_r^{[n]} = \operatorname{prox}_{\gamma_n g_r^*}\left(u_r^{[n]} + \gamma_n T_r x^{[n]}\right)$
$(\forall r \in \{1,\dots,R\})\quad u_r^{[n+1]} = \widehat{u}_r^{[n]} + \gamma_n T_r(\widehat{x}^{[n]} - x^{[n]})$
$(\forall s \in \{1,\dots,S\})\quad \widehat{v}_s^{[n]} = \operatorname{prox}_{\gamma_n \iota_{C_s}^*}\left(v_s^{[n]} + \gamma_n H_s x^{[n]}\right)$
$(\forall s \in \{1,\dots,S\})\quad v_s^{[n+1]} = \widehat{v}_s^{[n]} + \gamma_n H_s(\widehat{x}^{[n]} - x^{[n]})$
$\widehat{p}^{[n]} = \nabla h(\widehat{x}^{[n]}) + \sum_{r=1}^{R} T_r^\top \widehat{u}_r^{[n]} + \sum_{s=1}^{S} H_s^\top \widehat{v}_s^{[n]}$
$x^{[n+1]} = \widehat{x}^{[n]} - \gamma_n\left(\widehat{p}^{[n]} - p^{[n]}\right)$

Example 2.3.1 *Let $u \in \mathbb{R}^N$, let $\eta \in \mathbb{R}$, and let*

$$C = \{x \in \mathbb{R}^N \,|\, \langle x \,|\, u\rangle \leq \eta\}. \tag{2.40}$$

Then

- *If $u = 0$ and $\eta \geq 0$, in which case $C = \mathbb{R}^N$ and $P_C = \mathrm{Id}$.*
- *If $u = 0$ and $\eta < 0$, in which case $C = \emptyset$.*
- *If $u \neq 0$, in which case $C \neq \emptyset$ and*

$$(\forall x \in \mathbb{R}^N) \qquad P_C(x) = \begin{cases} x, & \text{if } \langle x \,|\, u\rangle \leq \eta, \\ x + \dfrac{\eta - \langle x \,|\, u\rangle}{\|u\|^2}u, & \text{otherwise.} \end{cases} \tag{2.41}$$

Example 2.3.2 *Let $u \in \mathbb{R}^N$, let $(\eta_1, \eta_2) \in \mathbb{R}^2$, and let*

$$C = \{x \in \mathbb{R}^N \,|\, \eta_1 \leq \langle x \,|\, u\rangle \leq \eta_2\}. \tag{2.42}$$

Then exactly one of the following holds:

- *$u = 0$ and $\eta_1 \leq 0 \leq \eta_2$, in which case $C = \mathbb{R}^N$ and $P_C = \mathrm{Id}$.*
- *$u = 0$ and $\eta_1 > 0$ (resp. $\eta_2 < 0$), in which case $C = \emptyset$.*

- $u \neq 0$ *and* $\eta_1 > \eta_2$, *in which case* $C = \emptyset$.
- $u \neq 0$ *and* $\eta_1 \leq \eta_2$, *in which case* $C \neq \emptyset$ *and*

$$(\forall x \in \mathbb{R}^N) \qquad P_C(x) = \begin{cases} x + \dfrac{\eta_1 - \langle x \,|\, u \rangle}{\|u\|^2} u, & \text{if } \langle x \,|\, u \rangle < \eta_1, \\ x, & \text{if } \eta_1 \leq \langle x \,|\, u \rangle \leq \eta_2, \\ x + \dfrac{\eta_2 - \langle x \,|\, u \rangle}{\|u\|^2} u, & \text{otherwise.} \end{cases} \tag{2.43}$$

The projection onto the ℓ_2-norm ball has also a closed-form expression:

Example 2.3.3 *Let* $\tau \in \mathbb{R}$, *and set* $C = \{x \in \mathbb{R}^N \mid \ \|x\|^2 \leq \tau\}$. *Then*

$$(\forall x \in \mathbb{R}^N), \qquad P_C(x) = \begin{cases} x, & \text{if } \|x\|^2 \leq \tau, \\ x \dfrac{\sqrt{\tau}}{\|x\|} & \text{otherwise.} \end{cases} \tag{2.44}$$

Subgradient projection When an expression of the direct projection is not available, a possible solution is to approximate the convex set C by a half-space, which leads to the concept of subgradient projection. Let us assume that $C = \big\{w \in \mathbb{R}^N \;\big|\; \varphi(w) \leq \eta\big\}$ is nonempty, where $\eta \in \mathbb{R}$ and $\varphi \in \Gamma_0(\mathbb{R}^N)$. Then, for any $y \in \mathbb{R}^N$ and $t \in \partial\varphi(y)$, the convex set C is a subset of the half-space

$$C_\eta^y = \big\{u \in \mathbb{R}^N \;\big|\; \langle u - y \,|\, t \rangle \leq \eta - \varphi(y)\big\}. \tag{2.45}$$

If $t \neq 0$, the *subgradient projection* of $y \in \mathbb{R}^N$ onto C reads

$$P_{C_\eta^y}(y) = \begin{cases} y, & \text{if } \varphi(y) \leq \eta, \\ y + \dfrac{\eta - \varphi(y)}{\|t\|^2}\, t, & \text{otherwise.} \end{cases} \tag{2.46}$$

Subgradient projection was used in several works. It is involved in Polyak's algorithm by alternating the exact projection and the subgradient projection Polyak [1969] (see also Shor [1985] for alternative projected subgradient approaches). In Combettes [2003], a parallel block iterative algorithm was proposed in order to solve Problem (2.21). However, the main limitation of this method is that the objective function to be minimized must be quadratic and strictly convex. For recent works about subgradient projection methods, the reader may refer to Slavakis et al. [2006]; Bouboulis et al. [2012].

Epigraphical projection An alternative approach to circumvent the computation of a direct projection consists in resorting to a set of epigraphical projections Chierchia et al. [2013]. This approach applies to constraints that can be expressed as the lower level set of a sum of convex functions evaluated over different blocks of the signal:

$$(\forall y = [\mathsf{y}_1^\top, \dots, \mathsf{y}_L^\top]^\top \in \mathbb{R}^M) \qquad \varphi(y) = \sum_{\ell=1}^{L} \varphi_\ell(\mathsf{y}_\ell) \leq \eta, \tag{2.47}$$

where, for every $\ell \in \{1, \dots, L\}$, $\mathsf{y}_\ell \in \mathbb{R}^{M_\ell}$ and $\varphi_\ell \in \Gamma_0(\mathbb{R}^{M_\ell})$ is such that $\mathrm{ri}(\mathrm{dom}\, \varphi_\ell) \neq \varnothing$. The *epigraphical-splitting* technique consists in introducing an auxiliary vector $\zeta = (\zeta_\ell)_{1 \leq \ell \leq L} \in \mathbb{R}^L$ into the minimization problem, so that Condition (2.47) can be equivalently rewritten as

$$\begin{cases} \displaystyle\sum_{\ell=1}^{L} \zeta_\ell \leq \eta, & (2.48) \\ (\forall \ell \in \{1, \dots, L\}) \qquad \varphi_\ell(\mathsf{y}_\ell) \leq \zeta_\ell. & (2.49) \end{cases}$$

Dealing with additional constraints in the original minimization is not a problem for proximal splitting algorithms, as far as the projections onto the associated constraint sets can be computed. In the present case, the projection onto (2.48) is given by Example 2.3.1 with $u = (1, \dots, 1) \in \mathbb{R}^L$, whereas the projection onto (2.49) is given by

$$(\forall (y, \zeta) \in \mathbb{R}^M \times \mathbb{R}^L) \qquad (p, \theta) = P_{\mathrm{epi}\, \varphi_1 \times \cdots \times \mathrm{epi}\, \varphi_L}(y, \zeta), \tag{2.50}$$

where $\theta = (\theta_\ell)_{1 \leq \ell \leq L}$, $p = [\mathsf{p}_1^\top, \dots, \mathsf{p}_L^\top]^\top \in \mathbb{R}^M$ is blockwise decomposed as y, and

$$(\forall \ell \in \{1, \dots, L\}) \qquad (\mathsf{p}_\ell, \theta_\ell) = P_{\mathrm{epi}\, \varphi_\ell}(\mathsf{y}_\ell, \zeta_\ell). \tag{2.51}$$

Therefore, the problem reduces to the lower-dimensional problem of the determination of the projection onto the convex subset $\mathrm{epi}\, \varphi_\ell$ of $\mathbb{R}^{M_\ell} \times \mathbb{R}$ for each $\ell \in \{1, \dots, L\}$. The key point of this approach is that epigraphical projections are closely related to proximity operators. Indeed, the epigraphical projector in (2.51) is given by Chierchia et al. [2013]:

$$\begin{cases} \mathsf{p}_\ell &= \mathrm{prox}_{\frac{1}{2}(\max\{\varphi_\ell - \zeta_\ell, 0\})^2}(\mathsf{y}_\ell), \\ \theta_\ell &= \max\{\varphi_\ell(\mathsf{p}_\ell), \zeta_\ell\}. \end{cases} \tag{2.52}$$

In Chierchia et al. [2013], the closed-form expression of the epigraphical projection has been given for several functions φ_ℓ, such as the absolute power raised to a power $\beta \in [1, +\infty[$, the distance to a convex set and the ℓ_p-norm with $p \in \{1, 2, +\infty\}$. Note also that, when the epigraphical projection has no explicit form, a polyhedral approximation can still be applied.

Numerical methods Beside the aforementioned approaches, specific numerical methods have been developed to directly solve the constrained quadratic minimization problem in (2.39) when

$$C = \{x \in \mathbb{R}^N \mid \varphi(x) \leq \eta\}, \tag{2.53}$$

with $\eta > 0$ and $\varphi \in \Gamma_0(\mathbb{R}^N)$. Indeed, it can be shown that [Bauschke and Combettes, 2011, Prop. 28.30]

$$P_C(x) = \begin{cases} x, & \text{if } \varphi(x) \leq \eta, \\ \text{prox}_{\bar{\lambda}\varphi}(x), & \text{otherwise}, \end{cases} \tag{2.54}$$

where $\bar{\lambda}$ is a solution in $]0, +\infty[$ to the equation:

$$\varphi(\text{prox}_{\lambda\varphi}(x)) = \eta. \tag{2.55}$$

In other words, the projection onto the lower level set of φ can be computed by using the proximity operator of φ, provided that the correct λ is found. In general, the solutions to (2.55) admit no closed-form expressions. However, one may find λ by standard numerical methods, as the left-hand side of (2.55) is decreasing in λ Parikh and Boyd [2013].

In sparse reconstruction applications, some interesting algorithms have been developed such as *spectral projected gradient techniques* Van Den Berg and Friedlander [2008],[5] *ℓ_1-regularized least squares methods* Koh et al. [2007],[6] *sparse reconstruction by separable approximation* Wright et al. [2009],[7] and many others Duchi et al. [2008], Schmidt et al. [2009], Weiss et al. [2009], Fadili and Peyré [2011]. Note that there also exist numerical methods that do not follow the aforementioned approach, such as Quattoni et al. [2009] and Kyrillidis et al. [2013].

§ 2.4 Conclusion

In this chapter, we have presented a quick overview of modern convex optimization. Firstly, we have introduced inverse problems and discussed the advantages of formulating them as convex optimization problems. Then, we have emphasized the need to look for prior information which aims at incorporating some additional knowledge in the problem formulation. To maintain some smoothness, while preserving some discontinuities of the signal, we have used a regularization term (such as TV and NLTV) which can be used either as a convex constraint set or a penalization function.

[5] http://www.cs.ubc.ca/labs/scl/SPGL11/
[6] http://www.stanford.edu/~boyd/l1_ls/.
[7] thttp://www.lx.it.pt/~mtf/SpaRSA/

Moreover, we have presented some algorithms that can efficiently deal with a large panel of convex problems. A special attention has been given to proximal algorithms, as they are suitable for large-size problems and are amenable to parallel implementations. Finally, to compute the projection onto some convex sets, we have detailed the existing approaches for the exact computation of projections and we have made an overview of the numerical methods in the case where the projection operator admits no closed-form expression.

In the next chapters, we will investigate some examples of constrained convex optimization problems in disparity estimation and multi-view reconstruction.

- Chapter 3 -

A parallel proximal splitting method for disparity estimation from multicomponent images under illumination variation

"Two Eyes = Three"
`3D Dimensions`

This chapter deals with the problem of matching a pair of multi-component images by jointly estimating the disparity and the illumination variation. Since the global formulation is nonconvex, the problem is addressed by solving a sequence of convex relaxations. Each convex relaxation is non trivial and involves many constraints aiming at imposing some regularity on the solution. Experiments demonstrate that the method is efficient and provides better results compared with other approaches.

§ 3.1 Introduction

The purpose of computer vision is to duplicate the abilities of human vision in order to perceive the observable world through digital images. Among the large panel of human abilities, depth perception is certainly one of the most appealing features to be electronically reproduced, as it may serve as a basis for more complex tasks, such as object recognition, robot navigation, reconstruction of three-dimensional shapes, and realistic scene visualization.

The depth of an object can be perceived thanks to binocular vision. As a simple experiment, hold your finger vertically in front of your eyes and close each eye alternately: you will notice that the finger jumps left and right relative to the background of the scene. Indeed, because of the horizontal separation, the eyes view an object from a slightly different angle, causing a horizontal displacement from the left eye to the right eye (Figure 3.1). The same mechanism underlies *computer stereo vision*, where the eyes are

replaced by two cameras viewing the scene from slightly different angles (Figure 3.2).

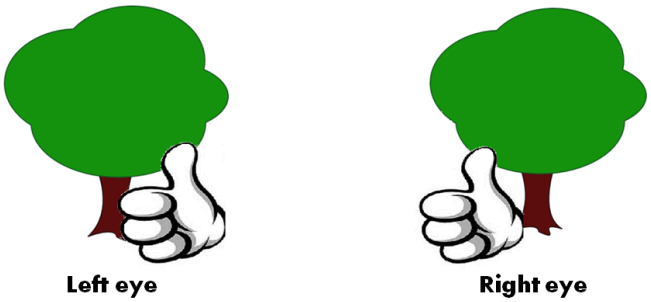


Figure 3.1: *If you focus on a tree, closing first your right eye and then your left eye, the tree will appear to move relatively to your finger.*

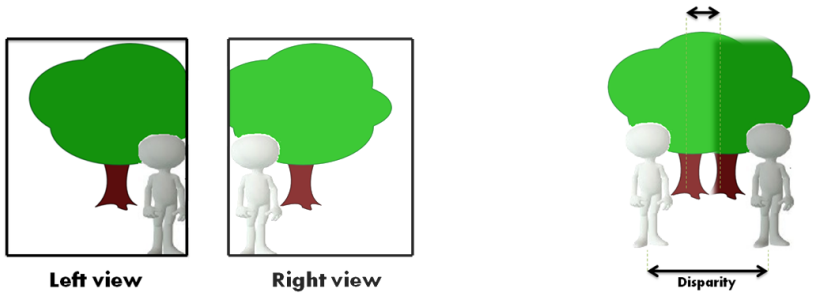


Figure 3.2: *Cameras (at two different positions) capture two separate images. The disparity is the difference between the positions of the same object in the two images. If an object is far away, the disparity will be small; otherwise, the disparity will be large if the object is close.*

Stereo matching is the process of finding the pixels in two (or more) images that correspond to the same 3D point of the observed scene, in order to compute the disparity between such pixels (i.e. the difference in coordinates) and, hence, to infer the depth of the associated 3D point.

Several preliminary steps are required before that stereo matching can take place: image acquisition, calibration, and rectification. In the present work, we focus on stereo matching, but in order to provide a comprehensive introduction to stereo vision, we briefly explain these preliminary steps in the following.

Acquisition and calibration

The pin-hole model is one of the simplest tools to describe image acquisition. In this model, the image is formed by a projection onto a 2D plane:

$$P : \quad \mathbb{R}^3 \longrightarrow \mathbb{R}^2 \tag{3.1}$$

$$(X, Y, Z) \longrightarrow (f\frac{X}{Z}, f\frac{Y}{Z}) \tag{3.2}$$

where f is the focal distance and (X, Y, Z) are the 3D coordinates. Calibration follows the image acquisition and consists of estimating the internal and the external geometry of the acquisition system. In particular, internal calibration aims at finding the focal point, the optical center of the camera and the size/angle of pixels, while external calibration aims at finding the rotations and translations of the camera with respect to an external reference. The calibration problem is usually considered as a solved problem, but current research still focuses on this subject, because the calibration is a crucial step in 3D reconstruction.

Rectification

Let us consider a 3D point P that is visible from both cameras. Given its projection P_L on the left image, its corresponding projection P_R on the right image plane has to be located on the epipolar line. The epipolar line is the intersection between the epipolar plane (defined by the optical centers of the stereo cameras C_L and C_R and the 3D point P) and the image plane 3.3.

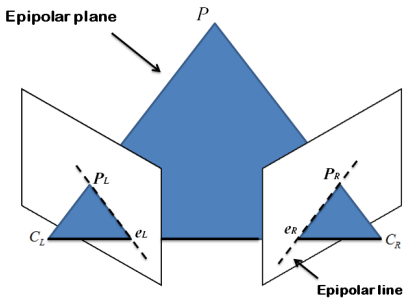


Figure 3.3: *The epipolar lines are defined by the intersection of the left and right image planes with the epipolar plane (C_L, C_R, P). The points e_L and e_R are the epipoles and are defined by the intersection of the image planes with the line connecting the two cameras: the baseline (C_L, C_R).*

Stereo matching uses triangulation based on epipolar geometry to find the correspondences between left and right images. Indeed, the epipolar line corresponding to a pixel in one image is used to constrain the search for the corresponding pixel in the other image. Rectification is a transformation process that puts the original images onto a common image plane, where the epipolar lines are parallel to the horizontal image axes. So doing, the corresponding search simplifies to a one-dimensional one. Note that the epipolar rectification is trivial when the calibration parameters are known; otherwise, it can be easily achieved by exploiting some corresponding points between the images.

Stereoscopic vision

In the following, we consider that the mapping is done from the left image (called "reference") to the right one. Each pixel in the left image is associated with a pixel in the right one which corresponds to the same point in the scene: they are homologous pixels. The difference between the pixel coordinates of the homologue points is called *disparity*.

The matching problem is then equivalent to searching for the disparity field u, which attributes a disparity for each pixel in the left image (x_L, y_L) as follows:

$$\begin{aligned} u : \quad & \mathbb{R}^2 \mapsto \mathbb{R}^2 \\ & (x_L, y_L) \mapsto (x_L - x_R, y_L - y_R), \end{aligned} \tag{3.3}$$

where (x_R, y_R) are the coordinates in the right image of the point corresponding to (x_L, y_L). If the stereo images are rectified, the vertical components of the corresponding pixels are equal ($y_L = y_R$). In this case, the disparity is a scalar and reads

$$\begin{aligned} u : \quad & \mathbb{R}^2 \mapsto \mathbb{R} \\ & (x_L, y_L) \mapsto x_L - x_R. \end{aligned} \tag{3.4}$$

The 3D position (X, Y, Z) of a point P can be reconstructed from the perspective projection of P on the image planes of the cameras (Figure 3.4), once the relative position and orientation of the two cameras is known. Let us consider the optical setting in the figure, which is also called standard model.

- C_L and C_R are two pinhole cameras with parallel optical axes. Let f be the focal length of both cameras.
- The baseline is perpendicular to the optical axes. Let B be the distance between the two lens centers.

- the optical axes lie in the XZ plane, which is parallel to the image plane of both cameras, X axis equals the baseline, and the origin of (X, Y, Z) world reference system is the lens center of the left camera.

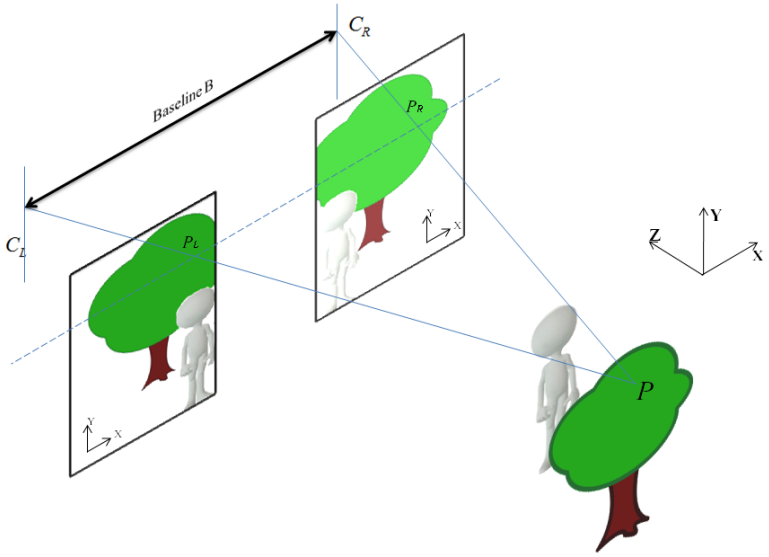


Figure 3.4: *Two cameras (acquiring images of the same scene) have two different 2D representations of a common 3D point. With proper processing, the position and depth of the 3D point can be extracted from the images.*

Figure 3.5 depicts the camera configuration. By applying Thales theorem to the triangles of hypotenuses C_LP and C_RP, we get:

$$\frac{x_L}{f} = \frac{X}{Z} \qquad \text{and} \qquad \frac{x_R}{f} = \frac{X - B}{Z}. \tag{3.5}$$

One can express the disparity $u(x_L, y_L) = x_L - x_R$ of a 3D point as a function of the depth Z between the corresponding 2D points as follows:

$$u(x_L, y_L) = x_L - x_R = \frac{fB}{Z} \tag{3.6}$$

Using the relationship (3.6), we can draw two important conclusions:

- disparity is inversely proportional to the depth of a point Z, i.e. far points have low disparity (for example the horizon has disparity of zero) and close points have a high disparity;

- the disparity is proportional to the baseline B: the smaller the baseline, the lower the disparity.

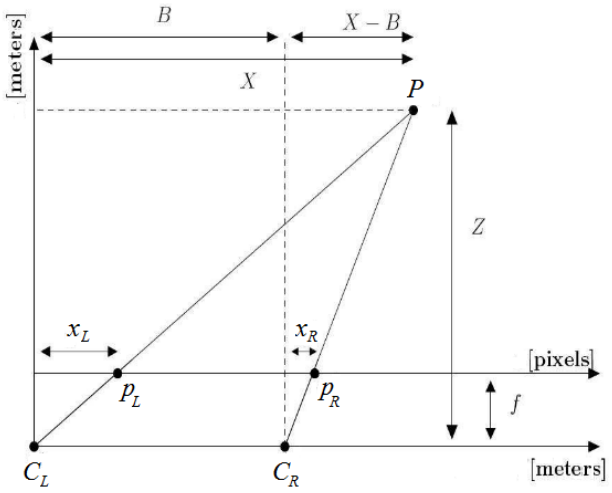


Figure 3.5: *The connection between disparity and depth.*

Stereo matching is a challenging task, because of the intrinsic difficulty in finding reliable correspondences between images in the presence of several sources of uncertainty, such as noise, illumination variations, and occlusions (which are pixels only visible from one view of the stereo images). A large number of algorithms have been proposed to perform stereo matching. Since the publication of the Middlebury database, which makes it possible to quantitatively compare the different algorithms, many solutions have been proposed to improve different aspects of stereo matching algorithms.

3.1.1 Related work

In recent years, much progress has been made in *dense* stereo matching, thanks to the development of

(i) powerful global combinatorial optimization methods, such as graph-cuts Kolmogorov and Zabih [2001]; Ishikawa [2003]; Boykov et al. [2011]; Komodakis et al. [2011] and belief propagation Klaus et al. [2006]; Yang et al. [2006]; Besse et al. [2012];

(ii) variational approaches which have proven to be also very effective for globally solving the matching problem Deriche et al. [1996]; Miled et al. [2006, 2009a]; Pock et al. [2010]; Cremers et al. [2011]; Wanner and Goldluecke [2013].

In general, dense stereo techniques can be also classified as local Hirschmuller and Scharstein [2007]; Tsai et al. [2003]; Yoon and Kweon [2006] or global Kolmogorov and Zabih [2001]; Ishikawa [2003]; Boykov et al. [2011]; Klaus et al. [2006]; Yang et al. [2006]; Besse et al. [2012]; Deriche et al. [1996]; Miled et al. [2006, 2009a], depending on whether they rely on local window-based computations or on the minimization of a global energy function. In local methods, the disparity computation at a given point depends only on intensity values within a finite window. Clearly, these techniques assume that all pixels within the window have the same disparity; therefore, they are sensitive near object boundaries. Attempts to alleviate this problem include the use of adaptive windows Kanade and Okutomi [1994] and shiftable windows Kang et al. [2001].

In a majority of dense disparity estimation techniques, the scene is assumed to be Lambertian. However, in the presence of illumination variation (that often occurs in practice), this assumption is violated. Some algorithms have been proposed to compensate illumination changes by normalizing pixel values Buchsbaum [1980]; Finlayson et al. [1998]; van de Weijer et al. [2007]. These algorithm can be applied to compensate for illumination changes as a preprocessing step before the stereo matching. The grey-world assumption algorithm Buchsbaum [1980] normalizes intensities based on the averages. The comprehensive normalization algorithm Finlayson et al. [1998] attempts to remove the dependency on scene geometry as well as the effect of global illumination iteratively. The grey-edge algorithm van de Weijer et al. [2007] employs the average edge difference for the normalization.

Other algorithms have been proposed to decrease the illumination variation effect by using invariant similarity measures Hirschmuller and Scharstein [2007]; Heo et al. [2011] in the matching cost computation. The normalized cross correlation (NCC) can be used as a similarity measure, which minimizes the effect of illumination conditions by compensating the differences of the gain and the bias between stereo images Hirschmuller and Scharstein [2007]. The adaptive normalized cross correlation (ANCC) Heo et al. [2011] employs the NCC with adaptively support weights after log-chrominance transformation and a suitable normalization that revert the illumination changes to a linear model. In addition, methods based on nonparametric local transforms followed by NCC and rank-transform Zabih and Woodfill [1994]; Jung and Kim [2012] have been used.

Recently, the light field has been established as additional information to describe the visual appearance of a scene. One of these approaches uses the light detection to employ the structure tensor of an epipolar plane image to obtain a robust local disparity estimation Wanner and Goldluecke [2013]. In Heo et al. [2013], the authors proposed a stereo color histogram equalization to produce color-consistent stereo images from which they can get more accurate disparity maps.

On the other hand, several techniques have been proposed to model the illumination changes. In this context, Cox et al. Cox et al. [1995] proposed a dynamic histogram warping that consists of directly matching histogram values and performing a global optimization via dynamic programming. However, this approach is restricted to the case of a spatially invariant relationship between the intensities of the two images. In Davis et al. [2005], a new method for stereo reconstruction called Light Transport Constancy was introduced to formulate a rank constraint and enable correspondence of non Lambertian surfaces.

The work in Gennert [1988] shows that the intensities of two corresponding points are related by a spatially varying multiplicative term. Following this model, a stereo algorithm based on a nonconvex cost function was developed. The minimization of the cost function from the associated Euler-Lagrange equations becomes a difficult task. In the same context, based on this multiplicative model, dense disparity map estimation is formulated in Miled et al. [2009a] as a constrained optimization problem in which a strictly convex quadratic objective function is minimized under various convex constraints. The resulting optimization problem is solved via a parallel block iterative algorithm involving subgradient projections.

3.1.2 Chapter outline

In this chapter, we present a convex optimization approach based on a parallel proximal algorithms Figueiredo and Nowak [2003]; Chambolle and Pock [2011]; Raguet et al. [2013]; Briceño-Arias and Combettes [2011]; Combettes and Pesquet [2012]; Pesquet and Pustelnik [2012]. Similarly to the subgradient based methods, we use the multiplicative model by Gennert, which allows us to define a global similarity measure to be minimized over an intersection of convex constraint sets. The original contributions of our work consist of

(i) the ability to consider multicomponent images in the presence of illumination variation

(ii) the flexibility in minimizing various (possibly nonsmooth) convex similarity measures and combining them with various convex constraints.

The remainder of the chapter is organized as follows. The notation, background, and considered model are presented in Section 3.2. Then, in Section 3.3, we describe the adopted parallel proximal algorithm that allows us to solve the derived convex minimization problem. Some simulation results are shown in Sections 3.4 and 3.5 on the MiddleBury data set and other real images. Finally, conclusions are drawn in Section 3.6.

§ 3.2 Stereo matching model

We consider two multicomponent images of the same scene acquired by a stereoscopic camera. The left (resp. right) view is given by the function

$$\begin{aligned} I_1 \colon \mathbb{R}^2 &\mapsto \mathbb{R}^K \\ \mathbf{s} &\mapsto I_1(\mathbf{s}) = (I_1^{(1)}(\mathbf{s}), I_1^{(2)}(\mathbf{s}), ..., I_1^{(K)}(\mathbf{s}))^\top \\ (\text{resp. } I_2 \colon \mathbb{R}^2 &\mapsto \mathbb{R}^K \\ \mathbf{s} &\mapsto I_2(\mathbf{s}) = (I_2^{(1)}(\mathbf{s}), I_2^{(2)}(\mathbf{s}), ..., I_2^{(K)}(\mathbf{s}))^\top). \end{aligned} \tag{3.7}$$

For every $k \in \{1, \dots, K\}$, $I_1^{(k)}(\mathbf{s})$ (resp. $I_2^{(k)}(\mathbf{s})$) represents the k-th component of image I_1 (resp. I_2) at position $\mathbf{s} \in \mathbb{R}^2$. Particular cases of this framework include color images by taking $K = 3$. Each component then corresponds to one of the color channels of a specific color system (RGB, YUV, ...). Throughout this chapter, it will be assumed that I_1 and I_2 are differentiable functions.

Corresponding points in the stereo images tend to have similar values. The pixel at position $(i_1, i_2) \in \mathbb{R}^2$ in the left image I_1 corresponds to a pixel at position $(i_1', i_2') \in \mathbb{R}^2$ in the right image I_2, the disparity between these pixels being equal to $(i_1 - i_1', i_2 - i_2')$. As already mentioned, when stereo images are rectified Fusiello et al. [2000], the vertical component of the disparity vector vanishes. The disparity thus reduces to $u(i_1, i_2) = i_1 - i_1'$. Finding for each pixel (i_1, i_2) in the left image the corresponding pixel $(i_1 - u(i_1, i_2), i_2)$ in the right image constitutes the goal of stereo matching. The problem is then equivalent to finding a disparity field $u \colon \mathbb{R}^2 \to [0, +\infty)$ that minimizes a similarity measure. However, under varying illumination conditions, the corresponding points in a stereo pair do not have the same component values. The illumination variation model we will employ is grounded on the work by Gennert Gennert [1988] and takes the following form:

$$(\forall (i_1, i_2) \in \mathbb{R}^2) \qquad I_2(i_1 - u(i_1, i_2), i_2) \simeq v(i_1, i_2) I_1(i_1, i_2) \tag{3.8}$$

where $v \colon \mathbb{R}^2 \to [0, +\infty)$ represents the illumination field. The problem now is not only to estimate the disparity u but also to estimate the illumination field v. The next section aims at better formulating the considered problem.

3.2.1 Problem formulation

In what follows, we propose a variational approach to jointly estimate the disparity u and the illumination variation v. Based on Model (3.8), we can

formulate the stereo matching problem as the minimization of a similarity measure $\widetilde{J}$:

$$\widetilde{J}(\mathbf{u},\mathbf{v}) \;=\; \sum_{k=1}^{K} \sum_{\mathbf{s}=(i_1,i_2)\in\mathcal{A}\setminus\mathcal{O}} \phi^{(k)}\big(v(\mathbf{s})I_1^{(k)}(\mathbf{s}) \;-\; I_2^{(k)}(i_1 \;-\; u(\mathbf{s}),i_2)\big) \quad (3.9)$$

where, for every $k \in \{1,\ldots,K\}$, $\phi^{(k)}$ is assumed to belong to $\Gamma_0(\mathbb{R})$, the class of proper lower semi-continuous convex functions from $\mathbb{R}$ to $]-\infty,+\infty]$. In addition, $\mathcal{A} \subset \mathbb{Z}^2$ is the considered finite image domain, $\mathbf{u} = (u(\mathbf{s}))_{\mathbf{s}\in\mathcal{A}}$, and $\mathbf{v} = (v(\mathbf{s}))_{\mathbf{s}\in\mathcal{A}}$. The occlusion areas denoted by $\mathcal{O}$ correspond to pixels only visible from one view of the stereo pair, which should not be taken into account in the computation of the similarity measure. Consequently, they have been discarded in the expression of the cost function.

In this chapter, we will be mainly concerned with convex optimization approaches. Unfortunately, $\widetilde{J}$ is nonconvex with respect to the variable $\mathbf{u}$. To tackle this difficulty, we perform the first-order Taylor expansion of the disparity compensated right image around an initial value of the disparity $\bar{\mathbf{u}}$. The latter can be derived from a rough estimation, e.g. by a block-based correlation method Tsai et al. [2003]. In practice, we can iteratively update our initial value in order to make our final solution weakly dependent on it. When the magnitude of the difference of the fields $\mathbf{u}$ and $\bar{\mathbf{u}} = (\bar{u}(\mathbf{s}))_{\mathbf{s}\in\mathcal{A}}$ is small enough, we obtain the following expression: for every $k \in \{1,\ldots,K\}$ and $\mathbf{s} = (i_1,i_2) \in \mathcal{A}$,

$$I_2^{(k)}(i_1-u(\mathbf{s}),i_2) \simeq I_2^{(k)}(i_1-\bar{u}(\mathbf{s}),i_2)-(u(\mathbf{s})-\bar{u}(\mathbf{s}))\nabla^{(1)}I_2^{(k)}(i_1-\bar{u}(\mathbf{s}),i_2) \quad (3.10)$$

where $\nabla^{(1)}I_2^{(k)}$ denotes the horizontal gradient of the k-th component of the right image.
As a consequence of (3.9) and (3.10), we can approximate the cost function by

$$J(\mathbf{u},\mathbf{v}) \;=\; \sum_{k=1}^{K} \sum_{s\in\mathcal{A}\setminus\mathcal{O}} \phi^{(k)}\big(T_1^{(k)}(\mathbf{s})u(\mathbf{s}) \;+\; T_2^{(k)}(\mathbf{s})v(\mathbf{s}) \;-\; r^{(k)}(\mathbf{s})\big) \quad (3.11)$$

where, for every $k \in \{1,\ldots,K\}$ and $\mathbf{s} = (i_1,i_2) \in \mathcal{A}$,

$$\begin{cases} T_1^{(k)}(\mathbf{s}) = \nabla^{(1)}I_2^{(k)}(i_1-\bar{u}(\mathbf{s}),i_2) \\ T_2^{(k)}(\mathbf{s}) = I_1^{(k)}(\mathbf{s}) \\ r^{(k)}(\mathbf{s}) = I_2^{(k)}(i_1-\bar{u}(\mathbf{s}),i_2)+\bar{u}(\mathbf{s})T_1^{(k)}(\mathbf{s}). \end{cases} \quad (3.12)$$

As previously mentioned, our objective is to jointly estimate $\mathbf{u}$ and $\mathbf{v}$. Thus, by defining $\mathbf{w} = (\mathbf{u},\mathbf{v})$,

$$(\forall \mathbf{s} \in \mathcal{A}) \qquad w(\mathbf{s}) = \begin{bmatrix} u(\mathbf{s}) \\ v(\mathbf{s}) \end{bmatrix} \quad (3.13)$$

and, for every $k \in \{1, \dots, K\}$ and $\mathbf{s} \in \mathbb{R}^2$, $\mathbf{T}^{(k)}(\mathbf{s}) = [T_1^{(k)}(\mathbf{s}), T_2^{(k)}(\mathbf{s})]$, the above expression can be reexpressed more concisely as

$$J(\mathbf{w}) = \sum_{k=1}^{K} J^{(k)}(\mathbf{w}) \tag{3.14}$$

where, for every $k \in \{1, \dots, K\}$,

$$J^{(k)}(\mathbf{w}) = \sum_{s \in \mathcal{A} \setminus \mathcal{O}} \phi^{(k)}\big(\mathbf{T}^{(k)}(\mathbf{s}) w(\mathbf{s}) - r^{(k)}(\mathbf{s})\big). \tag{3.15}$$

Despite the convexity of the function J, optimizing this criterion is an ill-posed problem, since we have two variables to estimate at each point $\mathbf{s}$ and the components of $T^{(k)}(\mathbf{s})$ may locally vanish for some k and $\mathbf{s}$. Therefore, we need to incorporate additional prior information on the desired disparity and illumination variation fields.

3.2.2 Introducing prior information

Our objective here is to introduce prior information on our target solution. This can be done either by adding some regularization terms to J or by incorporating some convex constraints to the problem. In this chapter, we will follow the second approach. Let $(S_i)_{1 \le i \le m}$ denote the m nonempty closed convex sets modelling the constraints one wants to impose. The resulting optimization problem can then be formulated as:

$$\text{Find } \mathbf{w} \in S = \bigcap_{i=1}^{m} S_i \text{ such that } J(\mathbf{w}) = \inf J(S). \tag{3.16}$$

Constraint sets defined on the Hilbert space $\mathcal{H} = \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|\mathcal{A}|}$ can be described as lower level sets:

$$(\forall i \in \{1, ..., m\}) \qquad S_i = \{\mathbf{w} \in \mathcal{H} \mid f_i(\mathbf{w}) \le \delta_i\}\ , \tag{3.17}$$

where, for every $i \in \{1, \dots, m\}$, $f_i \colon \mathcal{H} \to \mathbb{R}$ is a convex function and $\delta_i \in \mathbb{R}$. One of the potential advantages of this approach, with respect to the regularization formulation, is that it appears often easier, from a physical viewpoint, to set the constraint bounds than to determine the regularization parameters.

We will now review some of the constraints that can be applied to our problem.

3.2.2.1 Constraints on the disparity

Range values As the minimum and maximum values $u_{\min}$ and $u_{\max}$ of the disparity field are often known, we propose to introduce a constraint related to these bounds. The corresponding constraint can be expressed as

$$S_{1,1} = \{(u(\mathbf{s}))_{\mathbf{s}\in\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|} \mid (\forall \mathbf{s} \in \mathcal{A}) \quad u_{\min} \leq u(\mathbf{s}) \leq u_{\max}\}. \quad (3.18)$$

It can be noticed that the disparity is always nonnegative, so that $u_{\min} \geq 0$.

First-order smoothness constraint A piecewise constant behaviour of the disparity field is often expected. Therefore, it is meaningful to introduce a constraint enforcing smoothness in homogeneous areas while preserving edges. One can think of using the Total Variation semi-norm which has played a key role in image recovery problems Rudin et al. [1992]; Combettes and Pesquet [2004]. Let $\widehat{\nabla^{(1)}}$ and $\widehat{\nabla^{(2)}}$ denote discrete horizontal and vertical gradients (for example, obtained by cyclic convolutions Combettes and Pesquet [2008]; Pustelnik et al. [2011]). Then, a discrete version of the total variation (TV) is the following one:

$$(\forall \mathbf{u} \in \mathbb{R}^{|\mathcal{A}|}) \quad \mathrm{TV}(\mathbf{u}) = \sum_{\mathbf{s}\in\mathcal{A}} \sqrt{|\widehat{\nabla^{(1)}}u(\mathbf{s})|^2 + |\widehat{\nabla^{(2)}}u(\mathbf{s})|^2}. \quad (3.19)$$

The associated constraint set is

$$S_{1,2} = \left\{\mathbf{u} \in \mathbb{R}^{|\mathcal{A}|} \;\middle|\; \mathrm{TV}(\mathbf{u}) \leq \tau_2\right\} \quad (3.20)$$

with $\tau_2 > 0$.

Frame analysis constraint Alternatively, we can adopt a frame analysis approach to construct a smoothness constraint. Such transform can be described by an analysis frame operator $F\colon \mathbb{R}^{|\mathcal{A}|} \to \mathbb{R}^Q$ with $Q \geq |\mathcal{A}|$, which associates to $\mathbf{u} \in \mathbb{R}^{|\mathcal{A}|}$ its frame coefficients $\big((F\mathbf{u})_q\big)_{1\leq q\leq Q}$. In particular, it is well-known Meyer [1990]; Aujol et al. [2005]; Chambolle et al. [1998] that wavelet frames, through their relationships with Besov spaces, constitute appropriate tools to characterize useful classes of regular signals. F is said to be a tight frame when $F^\top F = \nu I$, where $\nu > 0$. A simple example of a tight frame is the union of ν orthonormal wavelet bases Mallat [1997]. Frame representations Han and Larson [2000] and, more precisely, tight frame representations Chaux et al. [2006] have become very popular during the last decades. The considered frame analysis constraint is expressed as

$$S'_{1,2} = \left\{\mathbf{u} \in \mathbb{R}^{|\mathcal{A}|} \;\middle|\; \sum_{q=1}^{Q} \eta_q |(F\mathbf{u})_q| \leq \tau'_2\right\} \quad (3.21)$$

where $(\eta_q)_{1\leq q\leq Q} \in [0, +\infty[^Q$ and $\tau'_2 > 0$.

Second-order constraint Some recent works Lefkimmiatis et al. [2012] have shown that considering the second-order derivatives of the target image can be helpful in data recovery problems. Let a discrete version of the Hessian of $\mathbf{u} \in \mathbb{R}^{|\mathcal{A}|}$ at $\mathbf{s} \in \mathcal{A}$ be

$$\widehat{\nabla^2}\mathbf{u}(\mathbf{s}) = \begin{bmatrix} \widehat{\nabla^{(1,1)}}u(\mathbf{s}) & \widehat{\nabla^{(1,2)}}u(\mathbf{s}) \\ \widehat{\nabla^{(1,2)}}u(\mathbf{s}) & \widehat{\nabla^{(2,2)}}u(\mathbf{s}) \end{bmatrix}. \tag{3.22}$$

Consequently, a second-order variant of the discrete total variation can be defined as

$$(\forall \mathbf{u} \in \mathbb{R}^{|\mathcal{A}|}) \quad \mathrm{TV}_2(\mathbf{u}) = \sum_{\mathbf{s}\in\mathcal{A}} \sqrt{|\widehat{\nabla^{(1,1)}}u(\mathbf{s})|^2 + |\widehat{\nabla^{(2,2)}}u(\mathbf{s})|^2 + 2|\widehat{\nabla^{(1,2)}}u(\mathbf{s})|^2} \tag{3.23}$$

and the associated constraint set is

$$S_{1,3} = \left\{\mathbf{u} \in \mathbb{R}^{|\mathcal{A}|} \;\middle|\; \mathrm{TV}_2(\mathbf{u}) \leq \tau_3\right\} \tag{3.24}$$

with $\tau_3 > 0$.

3.2.2.2 Constraints on the illumination field

Range values Similarly to the disparity field, estimations of the minimum and maximum values $v_{\min} \geq 0$ and $v_{\max}$ of the illumination field are often available. The corresponding constraint can be expressed as

$$S_{2,1} = \{\big(v(\mathbf{s})\big)_{\mathbf{s}\in\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|} \mid (\forall \mathbf{s} \in \mathcal{A}) \quad v_{\min} \leq v(\mathbf{s}) \leq v_{\max}\}. \tag{3.25}$$

First-order smoothness constraint Since the illumination field is usually smoothly varying, we can add a quadratic constraint on the discrete gradient of the illumination field. This amounts to considering the following constraint set:

$$S_{2,2} = \left\{\mathbf{v} \in \mathbb{R}^{|\mathcal{A}|} \;\middle|\; \|\widehat{\nabla}\mathbf{v}\|_{\ell_2}^2 \leq \kappa_2\right\} \tag{3.26}$$

where $\kappa_2 > 0$ and

$$(\forall \mathbf{v} \in \mathbb{R}^{|\mathcal{A}|}) \quad \|\widehat{\nabla}\mathbf{v}\|_{\ell^2} = \Big(\sum_{\mathbf{s}\in\mathcal{A}} |\widehat{\nabla^{(1)}}v(\mathbf{s})|^2 + |\widehat{\nabla^{(2)}}v(\mathbf{s})|^2\Big)^{1/2}. \tag{3.27}$$

Second-order constraint We can also introduce a constraint on the discrete Hessian of the illumination field, thus yielding

$$S_{2,3} = \left\{ \mathbf{v} \in \mathbb{R}^{|\mathcal{A}|} \;\middle|\; \|\widehat{\nabla^2}\mathbf{v}\|_{\ell_2}^2 \leq \kappa_3 \right\} \tag{3.28}$$

where $\kappa_3 > 0$ and, for every $\mathbf{v} \in \mathbb{R}^{|\mathcal{A}|}$,

$$\|\widehat{\nabla^2}\mathbf{v}\|_{\ell^2} = \Big(\sum_{\mathbf{s}\in\mathcal{A}} |\widehat{\nabla^{(1,1)}}v(\mathbf{s})|^2 + |\widehat{\nabla^{(2,2)}}v(\mathbf{s})|^2 + 2|\widehat{\nabla^{(1,2)}}v(\mathbf{s})|^2 \Big)^{1/2}. \tag{3.29}$$

3.2.2.3 Resulting constraints applied to vector w

Based on the constraints presented above, four different constraint sets are considered in this work:

(i) The first one concerns the range of values taken by the disparity and the illumination fields. It is defined as

$$S_1 = S_{1,1} \times S_{2,1}. \tag{3.30}$$

(ii) The second set is based on the first-order regularity of the illumination and disparity fields:

$$S_2 = S_{1,2} \times S_{2,2} \tag{3.31}$$

(iii) The third set involves the frame analysis formulation instead of the total variation for the disparity, and reads:

$$S_2' = S_{1,2}' \times S_{2,2}. \tag{3.32}$$

(iv) The last constraint set models the second-order smoothness constraints:

$$S_3 = S_{1,3} \times S_{2,3}. \tag{3.33}$$

§ 3.3 Proximal approaches for convex optimization

3.3.1 Optimization background

The constraints proposed in Section 3.2.2.3 are separable, since each constraint set S_i with $i \in \{1, \ldots, m\}$ can be expressed as $S_{1,i} \times S_{2,i}$.[1] In addition,

[1] Constraint S_2' will be substituted for S_2 in some of our experiments.

for every $i \in \{1, \ldots, m\}$, one can express $S_{1,i}$ (resp. $S_{2,i}$) as $L_{1,i}^{-1}(C_{1,i})$ (resp. $L_{2,i}^{-1}(C_{2,i})$) where $C_{1,i}$ (resp. $C_{2,i}$) is a nonempty closed convex subset of $\mathbb{R}^{N_{1,i}}$ (resp. $\mathbb{R}^{N_{2,i}}$) and $L_{1,i}$ (resp. $L_{2,i}$) is a matrix in $\mathbb{R}^{N_{1,i} \times |\mathcal{A}|}$ (resp. $\mathbb{R}^{N_{2,i} \times |\mathcal{A}|}$). Hence, we have chosen $L_{1,1}$ and $L_{2,1}$ equal to the $|\mathcal{A}| \times |\mathcal{A}|$ identity matrix. By setting $L_{1,2}$ and $L_{2,2}$ equal to the concatenation of the horizontal and vertical gradient operators ($N_{1,2} = N_{2,2} = 2|\mathcal{A}|$), $C_{1,2}$ and $C_{2,2}$ reduce to $\ell_{2,1}$ balls Van Den Berg and Friedlander [2008]; Eldar and Mishali [2009]. A similar choice can be made for constraint S_3 by considering linear operators related to the second-order derivatives ($N_{1,3} = N_{2,3} = 3|\mathcal{A}|$). In the considered cases, the projection onto each convex set $C_{1,i}$ or $C_{2,i}$ takes a closed-form expression or it can be computed in a finite number of operations Van Den Berg and Friedlander [2008]. The considered optimization problem then takes the following generic form:

$$\begin{array}{ll} \underset{L_{1,i}\mathbf{u} \in C_{1,i},\ L_{2,i}\mathbf{v} \in C_{2,i},\ i \in \{1,\ldots,m\}}{\text{minimize}} & J(\mathbf{w}). \end{array} \tag{3.34}$$

Numerical solutions to this problem can be provided by parallel proximal splitting methods. These methods basically consist of iterating computations of proximity operators, as we have shown in Chapter 2.

3.3.2 PPXA+ algorithm

As shown by Algorithm 3, we proceed by splitting function J into the K terms corresponding to each image component and by exploiting the separability of the closed convex constraint sets $(C_{1,i} \times C_{2,i})_{1 \leq i \leq m}$. The algorithm is initialized by setting the positive weights $(\omega_1, \omega_2, \ldots, \omega_{m+K}) \in]0, +\infty[^{m+K}$ as well as the variables $(\mathbf{z}_{1,i,0})_{1 \leq i \leq m+K}$ and $(\mathbf{z}_{2,i,0})_{1 \leq i \leq m+K}$ associated with the convex constraints $(C_{1,1} \times C_{2,1}, \ldots, C_{1,m} \times C_{2,m})$ and the split data fidelity criterion $J \mapsto \mathbf{w} \mapsto (J^{(1)}(\mathbf{w}), \ldots, J^{(K)}(\mathbf{w}))$.

The main loop first consists of computing the proximity operators of the data fidelity terms, as well as the direct projections onto the different convex sets modeling prior information. By using basic properties of proximity operators Combettes and Pesquet [2011], it can be shown that, for every $k \in \{1, \ldots, K\}$, $\omega_k > 0$, and $(\mathbf{z}_1, \mathbf{z}_2) \in \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|\mathcal{A}|}$,

$$\operatorname{prox}_{\frac{J^{(k)}}{\omega_k}}(\mathbf{z}_1, \mathbf{z}_2) = (z_1(\mathbf{s}) + \mu^{(k)}(\mathbf{s}) T_1^{(k)}(\mathbf{s}), z_2(\mathbf{s}) + \mu^{(k)}(\mathbf{s}) T_2^{(k)}(\mathbf{s}))_{\mathbf{s} \in \mathcal{A}} \tag{3.35}$$

where $\mathbf{z}_1 = (z_1(\mathbf{s}))_{\mathbf{s} \in \mathcal{A}}$, $\mathbf{z}_2 = (z_2(\mathbf{s}))_{\mathbf{s} \in \mathcal{A}}$, and

Algorithm 3 PPXA+

• *Notation*

$N_{1,m+1} = \ldots = N_{1,m+K} = |\mathcal{A}|$

$N_{2,m+1} = \ldots = N_{2,m+K} = |\mathcal{A}|$

`For` $k = 1, 2, \ldots, K$

$L_{1,m+k} = I\,,\;\; L_{2,m+k} = I$

• *Initialization*

$(\omega_1, \omega_2, \ldots, \omega_{m+K}) \in]0, +\infty[^{m+K}$

$(\mathbf{z}_{1,i}^{[0]})_{1\leq i\leq m+K} \in \mathbb{R}^{N_{1,1}} \times \mathbb{R}^{N_{1,2}} \times \cdots \times \mathbb{R}^{N_{1,m+K}}$

$(\mathbf{z}_{2,i}^{[0]})_{1\leq i\leq m+K} \in \mathbb{R}^{N_{2,1}} \times \mathbb{R}^{N_{2,2}} \times \cdots \times \mathbb{R}^{N_{2,m+K}}$

$Q_1 = \Big(\sum_{i=1}^{m+K} \omega_i (L_{1,i})^\top L_{1,i}\Big)^{-1},\;\; Q_2 = \Big(\sum_{i=1}^{m+K} \omega_i (L_{2,i})^\top L_{2,i}\Big)^{-1}$

$\mathbf{u}^{[0]} = Q_1\Big(\sum_{i=1}^{m+K} \omega_i (L_{1,i})^\top \mathbf{z}_{1,i}^{[0]}\Big)\,,\;\; \mathbf{v}^{[0]} = Q_2\Big(\sum_{i=1}^{m+K} \omega_i (L_{2,i})^\top \mathbf{z}_{2,i}^{[0]}\Big)$

• *Main loop*

`For` $n = 0, 1, \ldots$

`For` $i = 1, 2, \ldots, m$

$\mathbf{p}_{1,i}^{[n]} = P_{C_{1,i}}(\mathbf{z}_{1,i}^{[n]})$ `and` $\mathbf{p}_{2,i}^{[n]} = P_{C_{2,i}}(\mathbf{z}_{2,i}^{[n]})$

`For` $k = 1, 2, \ldots, K$

$(\mathbf{p}_{1,m+k}^{[n]}, \mathbf{p}_{2,m+k}^{[n]}) =$

$\operatorname{prox}_{\frac{J(k)}{\omega_{m+k}}}(\mathbf{z}_{1,m+k}^{[n]}, \mathbf{z}_{2,m+k}^{[n]})$

`Averaging`

$\mathbf{c}_1^{[n]} = Q_1\big(\sum_{i=1}^{m+K} \omega_i (L_{1,i})^\top \mathbf{p}_{1,i}^{[n]}\big)\,,\;\; \mathbf{c}_2^{[n]} = Q_2\big(\sum_{i=1}^{m+K} \omega_i (L_{2,i})^\top \mathbf{p}_{2,i}^{[n]}\big)$

`Updates`

`For` $i = 1, 2, \ldots, m$

$\mathbf{z}_{1,i}^{[n+1]} = \mathbf{z}_{1,i}^{[n]} + \lambda^{[n]}(L_{1,i}(2\mathbf{c}_1^{[n]} - \mathbf{u}^{[n]}) - \mathbf{p}_{1,i}^{[n]})$

$\mathbf{z}_{2,i}^{[n+1]} = \mathbf{z}_{2,i}^{[n]} + \lambda^{[n]}(L_{2,i}(2\mathbf{c}_2^{[n]} - \mathbf{v}^{[n]}) - \mathbf{p}_{2,i}^{[n]})$

`For` $k = 1, 2, \ldots, K$

$\mathbf{z}_{1,m+k}^{[n+1]} = \mathbf{z}_{1,m+k}^{[n]} + \lambda^{[n]}(2\mathbf{c}_1^{[n]} - \mathbf{u}^{[n]} - \mathbf{p}_{1,m+k}^{[n]})$

$\mathbf{z}_{2,m+k}^{[n+1]} = \mathbf{z}_{2,m+k}^{[n]} + \lambda^{[n]}(2\mathbf{c}_2^{[n]} - \mathbf{v}^{[n]} - \mathbf{p}_{2,m+k}^{[n]})$

$\mathbf{u}_{n+1} = \mathbf{u}^{[n]} + \lambda^{[n]}(\mathbf{c}_1^{[n]} - \mathbf{u}^{[n]})$

$\mathbf{v}_{n+1} = \mathbf{v}^{[n]} + \lambda^{[n]}(\mathbf{c}_2^{[n]} - \mathbf{v}^{[n]})$

$$\mu^{(k)}(\mathbf{s}) = \begin{cases} \frac{1}{\gamma^{(k)}(\mathbf{s})}\left(\text{prox}_{\frac{\gamma^{(k)}(\mathbf{s})\phi^{(k)}}{\omega_k}}(\tau^{(k)}(\mathbf{s})) - \tau^{(k)}(\mathbf{s})\right) & \text{if } \gamma^{(k)}(\mathbf{s}) \neq 0 \\ & \text{and } \mathbf{s} \notin \mathcal{O} \\ 0 & \text{otherwise,} \end{cases} \tag{3.36}$$

$$\tau^{(k)}(\mathbf{s}) = T_1^{(k)}(\mathbf{s}) z_1(\mathbf{s}) + T_2^{(k)}(\mathbf{s}) z_2(\mathbf{s}) - r^{(k)}(\mathbf{s}) \tag{3.37}$$

$$\gamma^{(k)}(\mathbf{s}) = |T_1^{(k)}(\mathbf{s})|^2 + |T_2^{(k)}(\mathbf{s})|^2. \tag{3.38}$$

Subsequently, an averaging of the variables resulting from the computation of the projections and proximity operators is performed before the update equations. It must be emphasized that the matrix inversions required in the algorithm can be efficiently performed by making use of discrete Fourier diagonalization techniques Pustelnik et al. [2012].
To gain insights in the convergence properties of this algorithm, assume that, for every $k \in \{1, \ldots, K\}$, $\phi^{(k)}$ is a finite function. In our problem, the data fidelity term is convex with respect to the variable $(\mathbf{u}, \mathbf{v})$ after Taylor expansion and each individual constraint is also convex with respect to either $\mathbf{u}$ or $\mathbf{v}$, as mentioned in Sections 3.2.1 and 3.2.2. Then, provided that the sequence of relaxation parameters $(\lambda_n)_{n\in\mathbb{N}}$ involved in Algorithm 3 is such that $(\forall n \in \mathbb{N})\ \tilde{\lambda} \leq \lambda_{n+1} \leq \lambda_n < 2$, where $\tilde{\lambda} \in]0, 2[$, [Pesquet and Pustelnik, 2012, Prop. 5.2] allows us to guarantee that the sequence $(\mathbf{u}_n, \mathbf{v}_n)_{n\in\mathbb{N}}$ generated by the algorithm converges to a solution to problem (3.16). The value of this parameter is set to 1.5 for all iterations. As the relaxation parameter constitutes an internal parameter of the algorithm, its choice does not modify the optimality of the solution. However, its value plays a role in the convergence speed as illustrated in Table 3.1. According to this table, the value of λ_n has been chosen so as to maximize the convergence speed.

Table 3.1: *Influence of λ_n on execution time (in seconds) for PPXA+ algorithm. The stopping condition described in Section 3.4.3 is used. The machine is a quad core Intel Xeon Processor X5450, 3Ghz CPU and 8Gb RAM.*

λ_n	0.5	0.7	0.9	1	1.1	1.3	**1.5**	1.7	1.9
Teddy	18.16	12.97	10.09	9.09	8.26	6.98	**6.29**	6.77	7.35
Cones	19.69	14.08	10.96	9.89	9.01	10.00	**9.05**	10.55	11.07

Finally, note that PPXA+ can be similarly applied in order to minimize J over $S_1 \cap S_2$ or $S_1 \cap S_3$.

§ 3.4 Results with local constraints

3.4.1 Error measures

First of all, we would like to define the quality measures we adopt for assessing the performance of our methods. Let $\widehat{z} \in \mathbb{R}^{|\mathcal{A}|}$ denote an estimate of $z \in \mathbb{R}^{|\mathcal{A}|}$. The measures below allow us to provide quantitative evaluations of our results:

(i) Mean Absolute Error (MAE)

$$\mathrm{MAE}(z, \widehat{z}) = \frac{1}{|\mathcal{A} \setminus \mathcal{O}|} \sum_{\mathbf{s} \in \mathcal{A} \setminus \mathcal{O}} |z(\mathbf{s}) - \hat{z}(\mathbf{s})|. \tag{3.39}$$

(ii) Average percentage of bad pixels

$$\mathrm{Err}_T(z, \widehat{z}) = \frac{1}{|\mathcal{A} \setminus \mathcal{O}|} \sum_{\mathbf{s} \in \mathcal{A} \setminus \mathcal{O}} 1_{\{|z(\mathbf{s}) - \hat{z}(\mathbf{s})| > T\}} \tag{3.40}$$

where T is some given positive threshold value chosen here equal to 2 and, for every $\mathbf{s} \in \mathcal{A}$,

$$1_{\{|z(\mathbf{s}) - \hat{z}(\mathbf{s})| > T\}} = \begin{cases} 1 & \text{if } |z(\mathbf{s}) - \hat{z}(\mathbf{s})| > T \\ 0 & \text{otherwise.} \end{cases} \tag{3.41}$$

3.4.2 Initialization

As previously mentioned, initial estimates $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ are obtained based on a block matching technique. Usually, two criteria are considered in the minimization performed in the block matching procedure: the sum of squared differences and the sum of absolute differences. In this work, we consider another popular measure, namely the Normalized Cross Correlation (NCC) similarity measure Tsai et al. [2003]. Note that more sophisticated measures could be envisaged such as the one proposed in Yoon and Kweon [2006]. NCC is defined as follows:

$$(\forall (i_1, i_2) \in \mathcal{A})(\forall u \in \mathbb{N}) \qquad \psi_{i_1, i_2}(u) = \sum_{k=1}^{K} \psi_{i_1, i_2}^{(k)}(u) \tag{3.42}$$

where

$$\psi_{i_1, i_2}^{(k)}(u) = \frac{\sum\limits_{(j_1, j_2) \in \mathcal{B}} I_1^{(k)}(i_1 + j_1, i_2 + j_2) I_2^{(k)}(x - u + j_1, i_2 + j_2)}{\sqrt{\sum\limits_{(j_1, j_2) \in \mathcal{B}} (I_1^{(k)}(i_1 + j_1, i_2 + j_2))^2} \sqrt{\sum\limits_{(j_1, j_2) \in \mathcal{B}} (I_2^{(k)}(x - u + j_1, i_2 + j_2))^2}} \tag{3.43}$$

and $\mathcal{B}$ corresponds to the matching block (here of size 11×11) centered at pixel (i_1, i_2) in the reference image.
In order to get the corresponding pixel $(i_1 - \bar{u}(i_1, i_2), i_2)$ in the right image, the matching process determines $\bar{u}(i_1, i_2)$ which maximizes the similarity measure ψ_{i_1,i_2}:

$$\bar{u}(i_1, i_2) \in \underset{u \in \mathbb{N} \cap [u_{\min}, u_{\max}]}{\operatorname{argmax}} \psi_{i_1,i_2}(u). \tag{3.44}$$

Once $\bar{u}(i_1, i_2)$ is found, $\bar{v}(i_1, i_2)$ is evaluated by the following weighted least squares estimation:

$$\bar{v}(i_1, i_2) = \frac{\sum_{k=1}^{K} \theta^{(k)} \sum\limits_{(j_1,j_2)\in\mathcal{B}} I_1^{(k)}(i_1 + j_1, i_2 + j_2) I_2^{(k)}(i_1 - \bar{u}(i_1, i_2) + j_1, i_2 + j_2)}{\sum_{k=1}^{K} \theta^{(k)} \sum\limits_{(j_1,j_2)\in\mathcal{B}} I_1^{(k)}(i_1 + j_1, i_2 + j_2)^2}, \tag{3.45}$$

where $(\theta^{(k)})_{1\leq k\leq K}$ are nonnegative real weights. In our tests, the illumination variation is the same for all the components and, for YUV images, the best results have been obtained by choosing a weight equal to 1 for the Y component, and 0 for the U and V components.
However, with the previous technique, some remaining artefacts can be observed in the disparity maps, e.g. contouring effects and oversmoothing in some areas. In order to reduce these artifacts and limit the influence of occlusion areas, a bidirectional matching process is performed. More precisely, we first apply the NCC method to find the left disparity map denoted by $\bar{u}_1(i_1, i_2)$, by proceeding as described above. Similarly, the same method is applied to find the right disparity map denoted by $\bar{u}_2(i_1, i_2)$ by taking now the right view as the reference image. Finally, we derive the initial disparity $\bar{u}(i_1, i_2)$ as follows:

$$\bar{u}(i_1, i_2) = \bar{u}_2(i_1 - \bar{u}_1(i_1, i_2), i_2). \tag{3.46}$$

The upper and lower bounds in the constrained formulation are estimated using prior knowledge when available or from the initial estimates $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$, otherwise.

3.4.3 Algorithm implementation

In this section, we summarize the algorithm steps and parameters that are required in our joint disparity and illumination variation estimation method.

Inputs:

- Occluded areas $\mathcal{O}$ estimated from the initial disparity map based on the ordering Yuille and Poggio [1984] and border constraints.

- $(T_1^{(k)}, T_2^{(k)}, r^{(k)})$ for $k \in \{1, 2, 3\}$.
- Minimum and maximum estimates of the fields to be estimated.
- Upper bounds for the constraints.

Parameter choice:

- For S_1, ω_1=100, ω_2=100.
- For S_2, S_3 or S_4: ω_3=200, ω_4=200.
- For $J^{(1)}$,$J^{(2)}$ and $J^{(3)}$: ω_5=ω_6=ω_7=10.
- $\lambda_n \equiv 1.5$.

The weights were chosen so that all the terms in the averaging step have a contribution of the same order.

Stopping criterion:

- The algorithm is stopped when the following condition is satisfied for 10 successive iterations:
  $\|\mathbf{u}_{n+1} - \mathbf{u}_n\| < \epsilon \|\mathbf{u}_n\|$ where $\epsilon = 10^{-5}$.

Outputs:

- The final values of the disparity vector $\mathbf{u}_n$ and the illumination variation vector $\mathbf{v}_n$.

3.4.4 Gray level images without illumination variation

Although the purpose of this work is to focus on situations where illumination variation arises, in this first experiment we aim at illustrating the performance of the proposed approach on grayscale images ($K = 1$) without any variation of the illumination field (for every $\mathbf{s} \in \mathcal{A}, v(\mathbf{s}) = 1$). This allows us to compare our method with some recent state-of-the-art methods referenced in the Middlebury website.[2] In Table 3.2, we display the results obtained on 3 different stereo pairs (Teddy, Venus, and Cones) for which we compare the results obtained by:

- the correlation-based approach used to initialize our approach (NCC),
- a graph-cut based approach (Gc+Occ) Kolmogorov and Zabih [2001],
- a total variation based algorithm (ConvexTV) Pock et al. [2008],

[2] http://vision.middlebury.edu/stereo/

- a method based on the knowledge of Ground Control Points (Global-GCP) Wang and Yang [2011],
- a histogram-based method (HistoAggr) Min et al. [2011],
- a curvelet-based approach (CurveletSupWgt) Mukherjee et al. [2010],
- the approach based on set theoretic estimation (subg. proj.) in Miled et al. [2009a],
- our proposed method based on PPXA+ minimizing an ℓ_1 cost subject to constraints $S_{1,1}$, $S_{1,2}$, and $S'_{1,2}$.

Note that the frame employed to define constraint $S'_{1,2}$ simply consists of an overcomplete Haar wavelet representation carried out over 1 resolution level. Other wavelet families have been tested without observing significant improvements.

Table 3.2: *Comparative results based on MAE for Teddy, Venus, and Cones stereo pairs. The red superscript numbers represent method ranking.*

Method	Teddy	Venus	Cones
NCC	1.169^{8}	0.411^{8}	1.134^{8}
Gc+Occ Kolmogorov and Zabih [2001]	1.166^{7}	0.303^{7}	0.585^{6}
ConvexTV Pock et al. [2008]	0.366^{1}	0.178^{1}	0.374^{1}
GlobalGCP Wang and Yang [2011]	0.489^{2}	0.272^{4}	0.443^{3}
HistoAggr Min et al. [2011]	0.788^{4}	0.278^{5}	0.434^{2}
CurveletSupWgt Mukherjee et al. [2010]	0.831^{5}	0.299^{6}	0.831^{7}
Subg. proj. Miled et al. [2009a]	0.890^{6}	0.207^{2}	0.551^{5}
PPXA+	0.666^{3}	0.211^{3}	0.487^{4}

The proposed approach appears to be competitive with respect to the other ones, although most of them are not able to deal with illumination variation. Moreover, as already mentioned, the proposed method can be implemented on a GPU architecture, thus reducing drastically the computation time. Indeed, the proximity operators and the projections onto the different convex sets can be computed in parallel. For example, with a GPU implementation, when considering the range values $S_{1,1}$ and the total variation constraints $S_{1,2}$, the following executing times per iteration were measured: for Teddy: 21.7ms, for Venus: 5.72ms, and for Cones: 21.98ms Gaetano et al. [2012].

Let us now turn our attention to more challenging scenarios including illumination variation both for grayscale and color images.

3.4.5 Gray level images with illumination variation

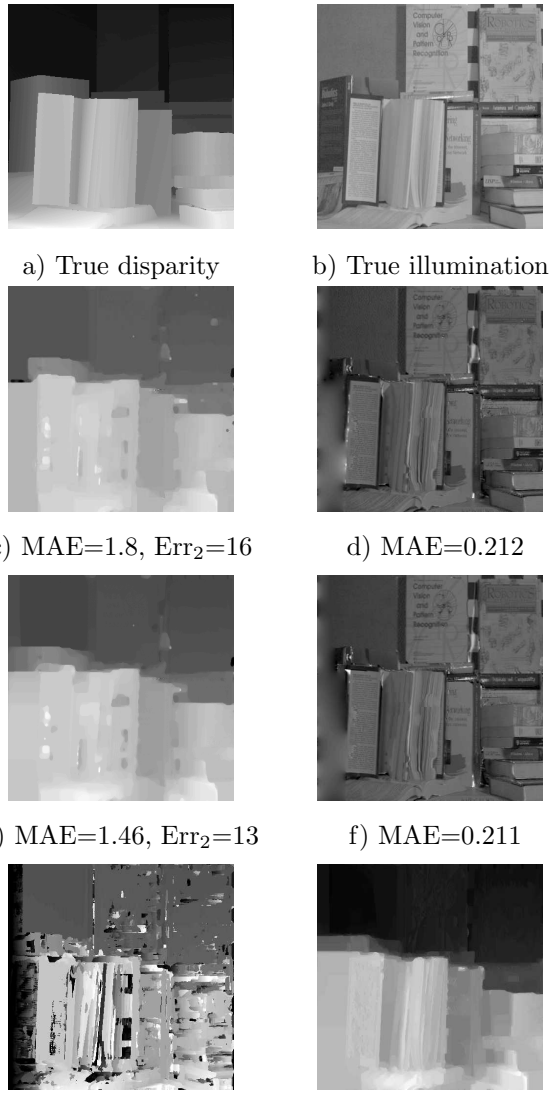
In this section, we evaluate our method on grayscale images ($K = 1$) in the presence of real and simulated illumination changes. The results are obtained on three stereo pairs depicted in Figure 3.6, named "Books", "Dolls", and "Parking meter".

Figure 3.6: *Stereo images: "Books" (left), "Dolls" (center), and "Parking meter" (right).*

In a first experiment, we estimate the disparity map and the illumination variation (real case) on the Books pair downloaded from MiddleBury website. We take here $\phi^{(1)} = |\cdot|$ for the cost function and consider the two convex constraints S_1 and S_2 for which the upper bounds were evaluated on the ground truth maps. In Figure 3.7, we compare our results with the subgradient projection approach proposed in Miled et al. [2009a], where a strictly convex quadratic function is used. We also compare our approach with two naive procedures for the disparity field estimation: the correlation based approach ignoring the illumination variation and a histogram equalization technique followed by a variational approach for disparity estimation, in the spirit of the approach we developed.

Concerning the choice of the criterion and of the constraint sets, we present in Figure 3.8 results obtained by using the following settings:

- for a)-b): S_1, S_2', and $\phi^{(1)} = |\cdot|$;
- for c)-d) same constraints as a)-b) and $\phi^{(1)} = (\cdot)^4$;
- for e)-f): S_1, S_2, S_3, and $\phi^{(1)} = |\cdot|$;
- for g)-h) same constraints as e)-f) and $\phi^{(1)} = |\cdot|^{\frac{3}{2}}$.
  By comparing the obtained results, it turns out that the ℓ_1 cost is a good choice, and that the use of second-order constraints leads to a marginal improvement.



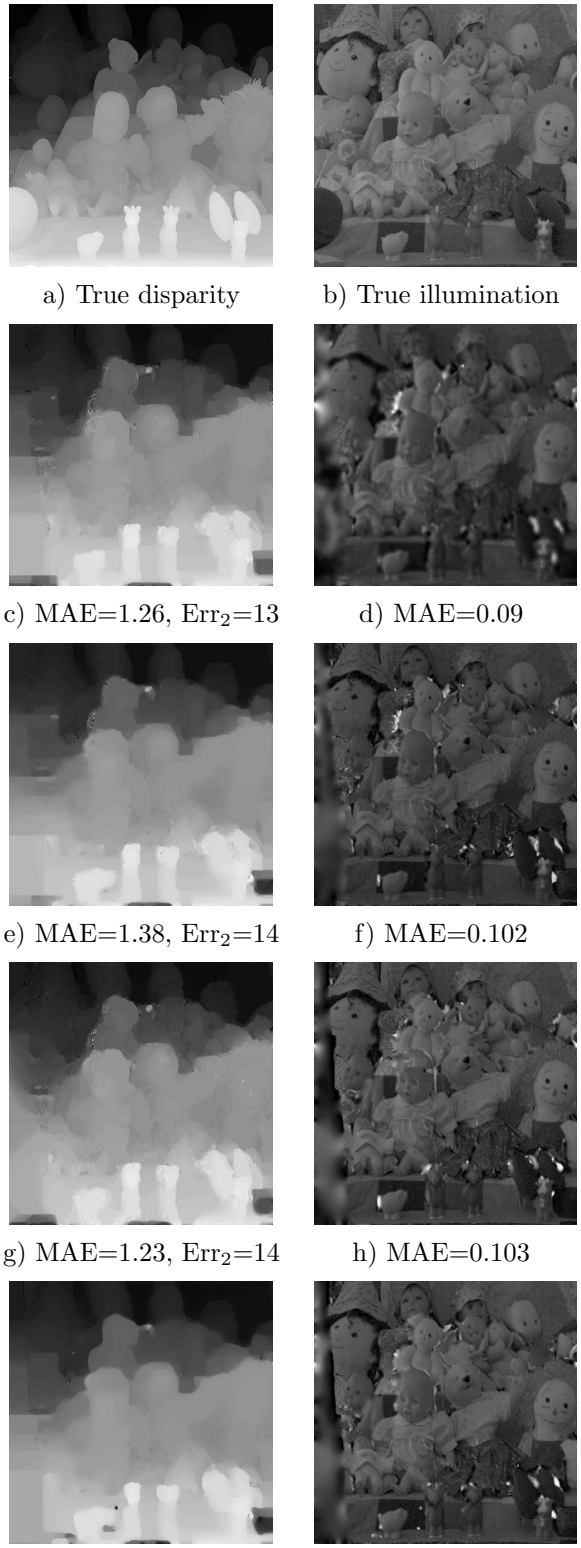
a) True disparity b) True illumination

c) MAE=1.8, Err_2=16 d) MAE=0.212

e) MAE=1.46, Err_2=13 f) MAE=0.211

g) MAE= 4.69, Err_2=26 h) MAE=1.55, Err_2=18

Figure 3.7: *Results for "Books" stereo pair: a)-b) ground truths, c)-d) subgradient projection method (c) disparity and d) illumination fields), e)-f) proposed approach with* $(u_{\min}, u_{\max}) = (20, 75)$,$(v_{\min}, v_{\max}) = (0.1, 1.1)$, $(\tau_2, \kappa_2) = (74000, 400)$ *(e) disparity and f) illumination fields), g) estimation of the disparity by NCC, h) estimation of the disparity after histogram equalization.*



a) True disparity b) True illumination

c) MAE=1.26, Err_2=13 d) MAE=0.09

e) MAE=1.38, Err_2=14 f) MAE=0.102

g) MAE=1.23, Err_2=14 h) MAE=0.103

i) MAE=1.23, Err_2= 14 j) MAE=0.103

Figure 3.8: *Results for Dolls stereo pair: proposed approach applied using c)-d)-ℓ_1-cost and $S_1 \cap S_2'$ with $(u_{\min}, u_{\max}) = (20, 75)$, $(v_{\min}, v_{\max}) = (0.1, 1.1)$, $(\tau_2', \kappa_2) = (92000, 230)$, e)-f)-$\ell_4$ cost and $S_1 \cap S_2'$, g)-h) ℓ_1 cost and $S_1 \cap S_2 \cap S_3$ with $(\tau_2, \kappa_2) = (34000, 230)$, $(\tau_3, \kappa_3) = (180000, 570)$, i)-j)-$\ell_{3/2}$ cost and $S_1 \cap S_2 \cap S_3$.*

In a third experiment (see Figure 3.10), we show the results provided by the proposed method (with $\phi^{(1)} = |\cdot|$) for a real image stereo pair from JISCT database, [3] taken under simulated illumination variation. To introduce a significant illumination variation, the left image is kept unchanged and the right image is modified by multiplying it with the Gaussian profile g depicted in Figure 3.9 and defined as:

$$(\forall (i_1, i_2) \in \mathcal{A})$$
$$g(i_1, i_2) = \alpha_1 \exp\left(-\frac{(i_1 - i_{1,0})^2 + (i_2 - i_{2,0})^2}{2\xi^2}\right) + \alpha_2 \quad (3.47)$$

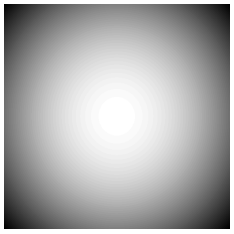
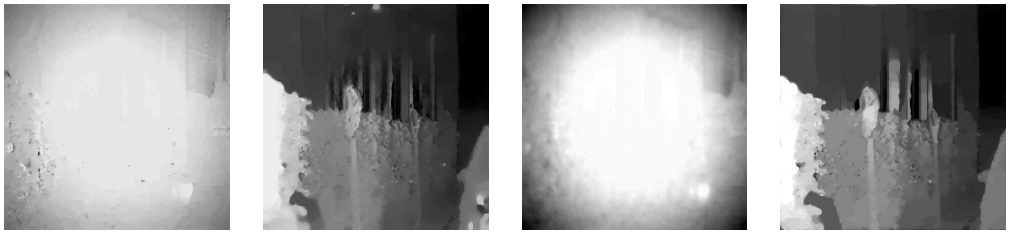


Figure 3.9: *Synthetic Gaussian profile.*

where $(i_{1,0}, i_{2,0})$ is the image center, $\xi = 512$ is the illumination deviation, α_1 and α_2 are given constants equal to 1.8 and -0.6 respectively.
In this case, the constraint bounds in S_1 and S_2 are not known. Hence, they are computed in the first place using the initial disparity, and then divided by a factor 2, thus assuming that the expected result should be smoother than the initial estimate.



a) MAE=0.033 b) c) MAE=0.031 d)

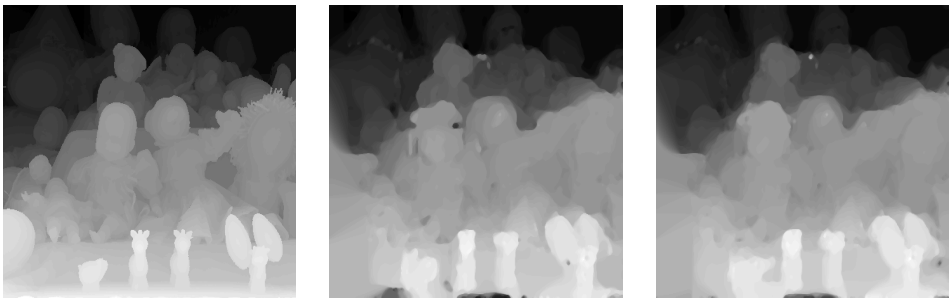
Figure 3.10: *Results for "Parking meter" stereo pair. a)-b) Subgradient projection method and c)-d) our approach with* $(u_{\min}, u_{\max}) = (1, 8)$, $(v_{\min}, v_{\max}) = (0.1, 1.1)$, $(\tau_2, \kappa_2) = (20000, 0.4)$. *a)-c) is the measured illumination field, b)-d) the measured disparity field.*

[3] http://vasc.ri.cmu.edu/idb/html/jisct/

3.4.6 Color images with illumination variation

For color images, we present tests performed on the Dolls pair also downloaded from Middlebury website. Constraint sets S_1 and S_2 were used, the associated bounds being computed on the ground truth fields. The results are provided in Figure 3.11 for YUV color images and the grayscale level images (Y component only) by using, for every $k \in \{1, \ldots, K\}$, $\phi^{(k)} = |\cdot|$. As expected, matching errors are reduced by using the proposed illumination variation model for color images.

a) left image b) right image



c) True disparity e) MAE=1.26, Err_2=13 g) MAE=1.10, Err_2=11

d) True illumination f) MAE=0.09 h) MAE=0.09

Figure 3.11: *Results for Dolls stereo pair: e)-f) results from gray scale images and g)-h) from color (YUV) images.* $((u_{\min}, u_{\max}) = (20, 75)$, $(v_{\min}, v_{\max}) = (0.1, 1.1)$, $(\tau_2, \kappa_2) = (92000, 240))$.

§ 3.5 Experiments with non local constraints

In this part, we exploit a recent extension of the total variation functional which is the Non Local TV (NLTV). Here, the nonlocality principal means that any point in the depth map can interact directly with any other point in the whole map. The main idea of nonlocal extension is based on the definition of nonlocal derivative operators, where the objective is to take into account additional prior information derived from the object itself. Such pixel grouping has already been considered in the field of image restoration, stereo matching, and motion estimation: when performing image restoration, the support weights are computed based on the fact that natural images exhibit a large amount of redundancy. This leads to the well-known neighborhood filters Buades et al. [2008] which belong to the best performing methods. A variational interpretation of these ideas leads to the nonlocal regularizers Gilboa and Osher [2009]; Peyré [2011]; Chierchia et al. [2013]; Couprie et al. [2013]. In the context of stereo matching, Yoon and Kweon Yoon and Kweon [2006] computed adaptive support weights to locally adapt the local matching window to the image structure, which leads to a significant improvement of the method. In Xiao et al. [2006], the authors proposed to incorporate the bilateral Gaussian filter into a variational motion estimation framework in order to improve the performance in occluded regions. In Werlberger et al. [2010], the nonlocal total variation regularization is introduced using the second order Taylor expansion of the data term applied in motion estimation.

The nonlocal total variation regularization incorporates this information directly into the energy definition yielding:

$$\mathrm{NLTV}(\mathbf{u}) = \sum_{s\in\mathcal{A}} \sqrt{\sum_{n\in\mathcal{N}_s\subset\mathcal{W}_s} \omega_{s,n}|\mathbf{u}(s)-\mathbf{u}(n)|^2}, \tag{3.48}$$

where $\mathcal{W}_s$ is a set of positions $n \in \mathcal{A}\setminus\{s\}$ located into a $Q\times Q$ window centered at s, where $Q\in\mathbb{N}$ is odd, and $\mathcal{N}_s$ is the actual neighborhood support of pixel s, which may contain a smaller subset of positions in $\mathcal{W}_s$.

The associated constraint is expressed as

$$S_{1,4} = \left\{\mathbf{u}\in\mathbb{R}^{|\mathcal{A}|} \;\middle|\; \mathrm{NLTV}(\mathbf{u}) \le \tau_4\right\} \tag{3.49}$$

with $\tau_4 > 0$.

Weight estimation for NLTV operator

For every $s\in\mathcal{A}$, we design the support $\mathcal{N}_s$ by selecting the pixels $n\in\mathcal{W}_s$ that are most similar to s according to their weights $\omega_{s,n}$. For every $n\in\mathcal{W}_s$, the weight $\omega_{s,n}$ depends on the similarity between patches built around the pixels s and n of the disparity map, as will be shown afterwards. In general,

such similarities can be estimated by following two different approaches. The first solution consists of computing the weights on the left image, assuming that object boundaries provide cues to localize discontinuities in the disparity map, as will be shown next. The alternative solution consists of a two-step approach, where an estimate of the disparity map is first computed with the TV regularizer and subsequently used to compute the weights. We follow the latter approach, as the former one suffers from the fact that similar patches in the left image may not necessarily correspond to regions with similar disparities.

In both cases, the weights are computed as follows

$$\omega_{s,n} = \widetilde{\omega}_s \exp\left(-\delta \, \|\widetilde{B}_s \widetilde{F}_s \widetilde{x} - \widetilde{B}_n \widetilde{F}_n \widetilde{x}\|^2\right), \tag{3.50}$$

where $\delta \in]0, +\infty[$, $\widetilde{\omega}_s \in (0, +\infty)$, $\widetilde{B}_s$ (resp. $\widetilde{B}_n$) selects a $\widetilde{Q} \times \widetilde{Q}$ patch centered at position s (resp. n), and $\widetilde{F}_s$ (resp. $\widetilde{F}_n$) is a linear processing of the image depending on the position s (resp. n). The constant $\widetilde{\omega}_s$ is set so as to normalize the weights (i.e. $\sum_{n \in \mathcal{N}_s} \omega_{s,n} = 1$). In the present work, we consider the *foveated self-similarity* measure recently introduced in Foi and Boracchi [2012], due to its better performance in denoising. This approach can be derived from (3.50) by setting $\widetilde{F}_s$ (resp. $\widetilde{F}_n$) to a set of low-pass Gaussian filters whose variances increase as the spatial distance from the patch center s (resp. n) grows.

For every $s \in \mathcal{A}$, the neighbourhood $\mathcal{N}_s$ is built according to the procedure described in Gilboa and Osher [2007]. In practice, we limit the size of the neighbourhood $|\mathcal{N}_s|$ to 14, and set $Q = 11$ and $\widetilde{Q}$=3.

3.5.1 Results

To emphasize the potential of our approach, we realized some experiments on the already introduced Middlebury benchmark database. For the NLTV constraint, we have used an epigraphical projection in order to handle the projection onto ℓ_{12} ball norm Chierchia et al. [2013]. Concerning the choice of the constraint sets, for NLTV (resp. TV) we use $S_1, S_{2,2}$, and S_4 (resp. S_1 and S_2). Note that the matrix inversion required by primal algorithms (such as PPXA+) can be performed by resorting to variable splitting as detailed in Peyré and Fadili [2011]. In this respect, primal-dual methods often constitute a better choice, as they require nothing but the linear operators and their associated adjoint operators. For this reason, in the following we employ the

M+LFBF algorithm(described in the previous chapter), which offers a good performance, robustness to numerical errors, and guaranteed convergence. The images used in our experiments are displayed in Figure 3.12.

(a) *Venus* **(b)** *Bull* **(c)** *Cloth* **(d)** *Wood*

Figure 3.12: *Left images of the stereo pairs.*

In Tables 3.3 and 3.4, we present the results obtained for two different cases: -Case 1: Gray images without illumination variation (Table 3.3). According to the percentage of bad pixels and the Mean Absolute Error, we can see that the NLTV approach performs better than the total variation one, with a gain of 0.83% in terms of bad pixels.

Table 3.3: *Disparity estimation results obtained for the gray level images Venus and Bull, using an ℓ_1 cost.*

	TV	NLTV	Parameters
Venus	MAE=0.21 Err_2=1.52	MAE=0.18 Err_2=0.69	τ_2=14000,τ_4=15600 $(u_{\min}, u_{\max})$ =(3,20)
Bull	MAE=0.20 Err_2=1.47	MAE= 0.18 Err_2= 1.16	τ_2=2533,τ_4=7458 $(u_{\min}, u_{\max})$ =(3,20)

-Case 2: Color images under illumination variation (Table 3.4). Also in this case, NLTV performs better than TV with a gain in the percentage of bad pixels of 0.51% and 2.01% for Cloth and Wood pairs respectively.

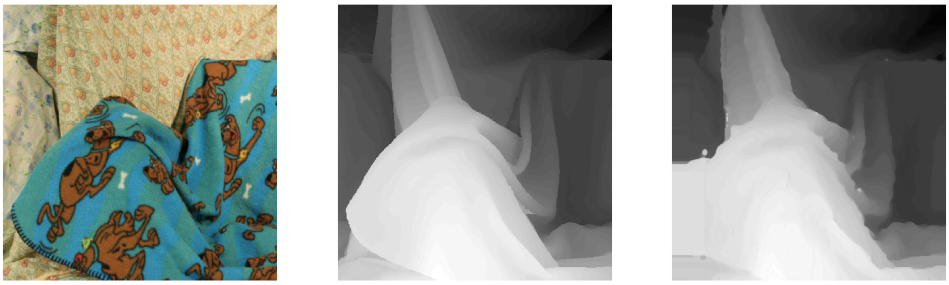
Table 3.4: *Disparity estimation results obtained for the color stereo pairs Cloth and Wood, under illumination variation using an ℓ_1 cost.*

	TV	NLTV	Parameters
Cloth	MAE=0.84 Err_2=2.44	MAE=0.82 Err_2=1.73	τ_2=45000, τ_4=18800 $(u_{\min}, u_{\max})$ =(15,55) $(v_{\min}, v_{\max})$=(0.1,0.6)
Wood	MAE=2.78 Err_2=10.47	MAE=2.53 Err_2=8.10	τ_2=58000, τ_4=21000 $(u_{\min}, u_{\max})$ =(21,71) $(v_{\min}, v_{\max})$=(0.1,0.6)

Finally, in Figure 3.13, we illustrate the recovered disparity maps using our method for color stereo pairs Cloth and Wood. These visual results clearly show the efficiency of our method in recovering the depth estimation in addition to illumination variations, for multicomponent images.

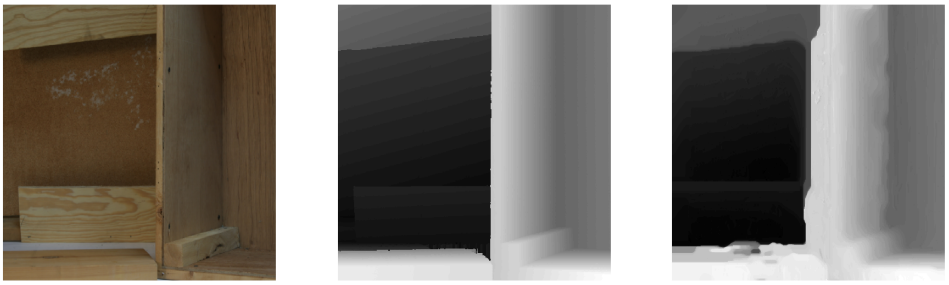
§ 3.6 Conclusion

In this chapter, we have investigated the application of a parallel proximal algorithm to the dense disparity estimation problem, for multicomponent (e.g. color) stereo matching, under illumination variation conditions. The proposed approach is flexible, as it allows us to consider various convex objective functions and constraints. It is also able to exploit the potentials offered by multicore/GPU parallel system architectures. However, one of the current limitations of the proposed method is that it is limited to convex cost functions, thus requiring a linearization of the original disparity model and making the estimation potentially sensitive to high matching error values. In the next chapter, we extend this approach to the multi-view context, but we resort to the convex relaxation proposed in Cremers et al. [2011], which allows us to consider nonconvex cost functions with no approximation in the computed solution (except for a quantization error).



(a) *Left image* **(b)** *True disparity* **(c)** *MAE=0.82, Err_2=1.73*

(d) *Right image* **(e)** *True illumination* **(f)** *MAE=0.067*



(g) *Left image* **(h)** *True disparity* **(i)** *MAE=2.53, Err_2=8.10*

(j) *Right image* **(k)** *True illumination* **(l)** *MAE=0.015*

Figure 3.13: *Visual results for with color images "Cloth" and "Wood". "Cloth" ((a) to (f)) with* $(u_{\min}, u_{\max}) = (15, 55)$,$(v_{\min}, v_{\max}) = (0.1, 0.6)$ *and* $(\tau_4, \kappa_2) = (18800, 135)$. *"Wood" ((g) to (l)) with* $(u_{\min}, u_{\max}) = (21, 71)$,$(v_{\min}, v_{\max}) = (0.1, 0.6)$ *and* $(\tau_4, \kappa_2) = (21000, 42)$.

- Chapter 4 -

Multi-view disparity estimation

"When you feel life at crossroads, you need higher perspective view."
`Toba Beta`

In the previous chapter, we discussed the problem of depth estimation from two views, under illumination variation. However, using two views, we have a limited field of view and we can only see a portion of a scene. In particular, a global description of objects often cannot be reconstructed from only two viewpoints due to occlusions. For this purpose, multi-view imaging has been used for many vision tasks to make the objects more visible.

This chapter addresses the problem of multi-view disparity estimation from rectified images. A sequence of disparity maps is estimated by using a multilabel optimization approach. For each of these maps, information from all the available views are exploited. The joint estimation problem is solved in a convex manner without adding much complexity with respect to the stereo matching case. The good performance of the proposed method is illustrated through experiments.

§ 4.1 Introduction

We consider the classical problem of estimating a dense 3D structure of a scene from a collection of calibrated views. Being one of the fundamental problems in computer vision, this issue has lately gained much attention in this field and remains an active research area. There are different types of approaches to tackle this problem, depending on the exploited image information. All these methods aim at reversing the image formation process. However, an image is essentially formed by projecting the 3D scene onto a 2D plane (which causes the loss of depth) and, in a mathematical sense, the inverse projection mapping does not exist, since all 3D points along a visual

ray are projected onto the same image 2D point. This makes the problem of multi-view 3D reconstruction ill-posed. Thus, additional constraints are needed in order to make the reconstruction problem well-posed. Multi-view is known to produce detailed reconstructions with a quality close to that of laser scanned models Seitz et al. [2006].

4.1.1 Related Work

The problem of 3D modeling from multiple images or multi-view stereopsis has been widely studied in the field of computer vision, computer graphics, and computational photography. A large panel of variational methods for multi-view 3D reconstruction inherits the active contour framework originally proposed for image segmentation Kass et al. [1988]. In this context, the problem consists of modeling a continuous two-dimensional surface as the minimizer of an appropriate energy functional, by using geodesic active contour models Caselles et al. [1995, 1997], level sets Faugeras and Keriven [1998], triangle meshes Hernandez and Schmitt [2004]; Duan et al. [2004] or graph cuts Vogiatzis et al. [2007]. A generalization of this approach has been developed in Pons et al. [2007], which allows to replace the classical photo-consistency estimation with a global matching score on the entire image domain.

On the other side, the Patch-based Multi-view Stereo algorithm proposed by Furukawa and Ponce [2010] appears as the state-of-the-art method among those based on feature extraction and matching. However, methods based on feature extraction and matching may show significant errors caused by severe occlusions and highly reflective surfaces. Probabilistic models have been proposed as an alternative approach to handle complicated scene geometry. These models do not make hard decisions about surface geometry or appearance: they explicitly represent uncertainties by assigning probabilities to multiple hypotheses. Early works along this line of reasoning Broadhurst et al. [2001]; Bhotika et al. [2002] can be regarded as extensions of Space Carving Kutulakos and Seitz [1999] and, more recently, algorithms based on generative models for the reverse image formation process have been introduced Gargallo et al. [2007]; Liu and Cooper [2011]. Using a Bayesian approach, these algorithms infer the maximum a posteriori estimation from the joint probability of all the images.

A different approach amounts at recovering a set of consistent view-dependent depth maps from an image sequence, instead of reconstructing a complete 3D model. This is mainly motivated by applications such as view synthesis, depth-based segmentation, and video enhancement. Early approaches in depth estimation Okutomi and Kanade [1993]; Collins [1996] used local window-based methods and employed a local *winner-takes-all* strategy. Our work is somewhat related to that of Kang and Szeliski [2004]; Gargallo and Sturm [2005], which also aims at inferring consistent depth maps from

multiple images. The authors in Gargallo and Sturm [2005] formulated 3D modeling from images as a Bayesian MAP problem, and solved it using the expectation-maximization algorithm. They used the estimated depth map to determine the visibility prior. A multiple-depth-map prior was finally used to smooth and merge the depths while preserving discontinuities. Larsen et al. [2007] presented an approach for 3D reconstruction from multiple synchronized video streams. In order to improve the final reconstruction quality, they used the optical flow to find corresponding pixels in the subsequent frames acquired by the same camera, and enforced the temporal consistency in reconstructing successive frames. With the observation that the depth error in conventional stereo methods grows quadratically with depth, Gallup et al. [2008] proposed a multi-baseline and multi-resolution stereo method to achieve constant depth accuracy by varying the baseline and resolution proportionally to depth.

In summary, although many approaches have been proposed to model 3D objects or to estimate depths using multiple input images, the problem of how to appropriately extract information and recover consistent depths from an image sequence remains challenging. In the present work, a global multilabel approach is proposed for multi-view disparity estimation, which is grounded on the use of a convex optimization strategy. A global solution to the problem is provided by minimizing a regularized similarity measure. The regularization enforces the smoothness of the sought disparity maps while preserving coherent contours in the set of images. The proposed formulation handles possibly nonconvex similarity measures. We also point out that the additional complexity required by the application of our method to the multi-view context is small with respect to the stereo case.

The chapter is organized as follows. In Section 4.2, we provide a general formulation of the multi-view disparity estimation problem. In Section 4.3, we describe a convex relaxation method based on the multilabel approach and we propose an algorithm to solve the resulting convex optimization problem. In Section 4.4, we describe an alternative convex relaxation method based on an improved first-order-Taylor approximation around an initial estimate. In Section 4.5, we present some numerical comparisons of the proposed multi-view disparity estimation methods and we show the effectiveness of the first one with respect to the standard stereoscopic approach (i.e. based on two views) and other multi-view estimation approach. Some conclusions are given in Section 4.6.

§ 4.2 Multi-view disparity estimation

Although two images are enough to estimate the depth of a scene from point correspondences, having additional views presents two significant benefits. Firstly, the estimation process becomes more accurate: since there is more

information about the same 3D point, this redundancy can be exploited in order to improve the depth estimation quality. Secondly, occluded pixels (i.e. pixels that have no match in some of the images) can be handled using the other images where a correspondence exists. The uncertainty related to occluded pixels can thus be reduced in multi-view systems.

This work investigates the disparity estimation problem for an arbitrary number of views acquired by rectified cameras Hartley and Zisserman [2004]; Nozick [2011], with the camera centers aligned.

4.2.1 Problem formulation

We suppose that N rectified views $(I_n)_{1 \leq n \leq N}$ are taken along a straight line with equal distance between any two consecutive camera positions (Figure 4.1).

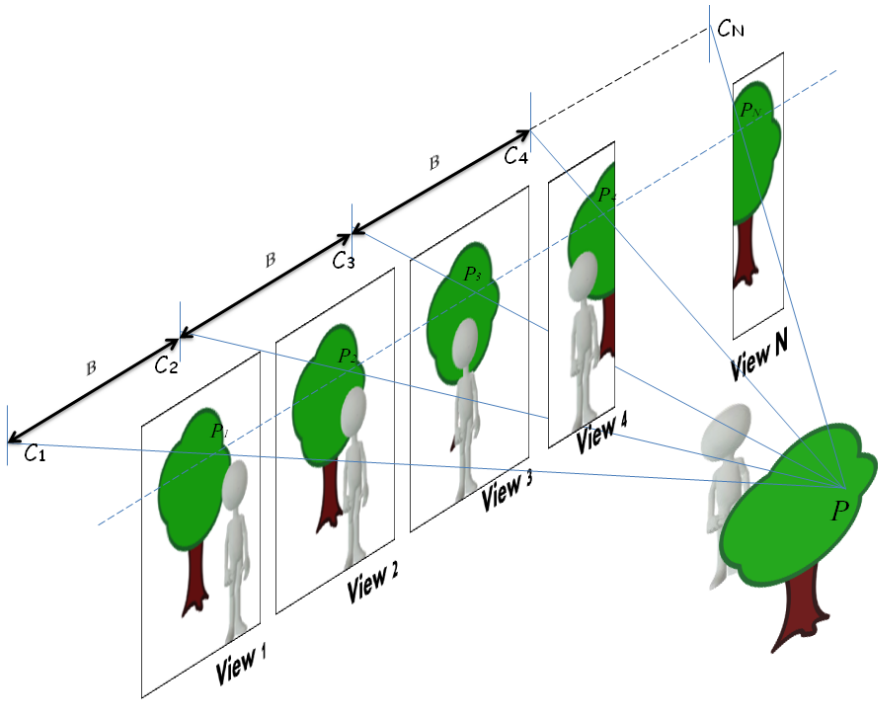


Figure 4.1: *N images (view 1, view 2, ...) captured by N aligned and equidistant cameras (C_1, C_2, ...) at N different positions, resulting in a set of stereo pairs with equal baselines B.*

Each pair of views (I_n, I_m), with $(n, m) \in \{1, \ldots, N\}^2$ and $n \neq m$, is associated with a disparity field $u_{n,m}$. By invoking Equation (3.6), the disparity values $(u^{(\mathbf{s})}_{n,m})_{\mathbf{s} \in \mathcal{A}}$ are proportional to the distance between the two

cameras, the focal length f, and to the inverse depths of the objects $(Z^{(s)})$ in the scene:

$$(\forall s \in \mathcal{A}), \quad u_{n,m}^{(s)} = \frac{f\ (m-n)B}{Z^{(s)}}. \tag{4.1}$$

Using the same notation as in the previous chapter, $\mathbf{s} = (i_1, i_2) \in \mathcal{A}$ denotes a pixel coordinate and $\mathcal{A} \subset \mathbb{R}^2$ is the image support. Under classical epipolar geometry assumptions (see Chapter 3), the intensities of corresponding pixels in the rectified images are related by

$$I_n^{(i_1,i_2)} = I_m^{(i_1 - u_{n,m}^{(\mathbf{s})}, i_2)}. \tag{4.2}$$

The overall number of disparity fields is equal to $N(N-1)/2$, but these are obviously related since they correspond to the same 3D scene. In particular, for every $n \in \{1, \dots, N\}$, if $k_n \neq n$ is some given index in $\{1, \dots, N\}$ (typically, $k_n = n+1$ is the right image index, when $n < N$ and $k_n = N-1$ is the left image index, when $n = N$), we have

$$(\forall (n,m) \in \{1, \dots, N\}^2, n \neq m) \qquad u_{n,m} = \alpha_{n,m} u_{n,k_n} \tag{4.3}$$

where $\alpha_{n,m}$ is a real-valued constant depending on the relative positions of the cameras. According to Equation (4.1), when the cameras are equidistant, $\alpha_{n,m} = \frac{m-n}{k_n - n}$.

Hence, by using (4.3), the number of disparity fields to be estimated reduces to N. By taking advantage of the N views, the u_{n,k_n} field can be estimated with a variational approach that minimizes a criterion of the form

$$f_n(u_{n,k_n}) = \sum_{\substack{m=1 \\ m \neq n}}^{N} \sum_{\mathbf{s} \in \mathcal{D}_{n,m}} \psi(I_n^{(\mathbf{s})}, I_m^{(i_1 - \alpha_{n,m} u_{n,k_n}^{(\mathbf{s})}, i_2)}) \tag{4.4}$$

where $\mathcal{D}_{n,m}$ denotes some finite subset of $\mathcal{A}$ of unoccluded pixels (i.e. pixels in the n-th view that can be matched with pixels in the m-th view) and $\psi : \mathbb{R}^2 \to \mathbb{R}$ is a similarity measure. As explained in the following sections, the proposed approach is applicable to any kind of function of interest (ℓ_2 norm, truncated ℓ_1 cost function, etc...).

§ 4.3 Approach based on multilabeling

4.3.1 Convex representation

In order to estimate u_{n,k_n}, we propose to adopt a multilabel approach Pock et al. [2008]; Hiltunen et al. [2012]. More specifically, the disparity u_{n,k_n} is quantized over $Q+1$ quantization levels $r_0, r_1, \dots, r_Q$, such that $r_0 < r_1 < \cdots < r_Q$. We employ a bijection that maps the disparity map

$u_{n,k_n} \in \{r_0, \dots, r_Q\}^P$, with $P = \text{card}(\mathcal{A})$, to a multivariate binary field $\theta_n = (\theta_{n,1}, \dots, \theta_{n,Q}) \in \mathcal{B}$ with components

$$(\forall\ q \in \{1, \dots, Q\})(\forall\ \mathbf{s} \in \mathcal{A}) \qquad \theta_{n,q}^{(\mathbf{s})} = \begin{cases} 1 & \text{if } u_{n,k_n}^{(\mathbf{s})} \geq r_q \\ 0 & \text{otherwise} \end{cases}$$

where

$$\mathcal{B} = \{\theta_n \in (\{0,1\}^P)^Q |\ (\forall\ \mathbf{s} \in \mathcal{A})\ \ 1 \geq \theta_{n,1}^{(\mathbf{s})} \geq \cdots \geq \theta_{n,Q}^{(\mathbf{s})} \geq 0\}.$$

More precisely, there exists a bijection between the set of disparity fields $u_{n,k_n} \in \{r_0, \dots, r_Q\}^P$ and the set of multivariate label images $\theta_{n,q}$ as defined above. We have then

$$(\forall\ \mathbf{s} \in \mathcal{A}) \qquad u_{n,k_n}^{(\mathbf{s})} = r_0 + \sum_{q=1}^{Q} (r_q - r_{q-1}) \theta_{n,q}^{(\mathbf{s})} \tag{4.5}$$

The estimation of the quantized disparity u_{n,k_n} is thus equivalent to finding the associated multivariate binary field $\theta_n \in \mathcal{B}$. It can be further observed that the global error measure can be re-expressed as a linear function of θ_n as follows:

$$\widetilde{f}_n(\theta_n) = \sum_{\substack{m=1 \\ m \neq n}}^{N} \sum_{\mathbf{s} \in \mathcal{D}_{n,m}} \sum_{q=0}^{Q} \psi(I_n^{(\mathbf{s})}, I_m^{(i_1 - \alpha_{n,m} r_q, i_2)}) (\theta_{n,q}^{(\mathbf{s})} - \theta_{n,q+1}^{(\mathbf{s})}) \tag{4.6}$$

where $\theta_{n,0} = 1$ and $\theta_{n,Q+1} = 0$. Consequently, the minimization problem can be expressed as:

$$\underset{\theta_n \in \mathcal{B}}{\text{minimize}}\ \ \widetilde{f}_n(\theta_n) + \rho(\theta_n) \tag{4.7}$$

where ρ is some regularization function used to promote the spatial regularity of the binary images. A typical choice for ρ is the isotropic discrete total variation which is denoted by tv. Then, our problem becomes:

$$\underset{\theta_n \in \mathcal{B}}{\text{minimize}}\ \ \widetilde{f}_n(\theta_n) + \mu \sum_{q=1}^{Q} (r_q - r_{q-1})\, \mathsf{tv}(\theta_{n,q}), \quad \mu > 0. \tag{4.8}$$

Note that $\widetilde{f}_n$ and tv are convex functions, but the problem is nonconvex due to the nonconvexity of the set $\mathcal{B}$. We thus use a convex relaxation based on the approach of Cremers et al. [2011]. The problem is then equivalent to

$$\underset{\theta_n \in \mathcal{R}}{\text{minimize}}\ \ \widetilde{f}_n(\theta_n) + \mu \sum_{q=1}^{Q} (r_q - r_{q-1})\, \mathsf{tv}(\theta_{n,q}) \tag{4.9}$$

where

$$\mathcal{R} = \{\theta_n \in ([0,1]^P)^Q | (\forall\ \mathbf{s} \in \mathcal{A})\ \ 1 \geq \theta_{n,1}^{(\mathbf{s})} \geq \cdots \geq \theta_{n,Q}^{(\mathbf{s})} \geq 0\}.$$

Although the convex function $\widetilde{f}_n(\theta_n) + \rho(\theta_n)$ is minimized over a convex set $\mathcal{R}$, the uniqueness of a solution is not guaranteed Ekeland and Téman [1999]. In order to solve Problem (4.9), we need to introduce some auxiliary functions.

Firstly, let $g_n : \mathcal{R} \to \mathbb{R}$ be a linear function defined as

$$(\forall\ \theta_n \in \mathcal{R}) \qquad g_n(\theta_n) = \widetilde{f}_n(\theta_n) - \sum_{\substack{m=1 \\ m \neq n}}^{N} \sum_{\mathbf{s} \in \mathcal{D}_{n,m}} \psi\Big(I_n^{(\mathbf{s})}, I_m^{(i_1 - \alpha_{n,m} r_0, i_2)}\Big) \tag{4.10}$$

$$= \sum_{q=1}^{Q} \sum_{\mathbf{s} \in \mathcal{A}} {\theta_{n,q}}^{\top} \varsigma_{n,q} = {\theta_n}^{\top} \varsigma_n, \tag{4.11}$$

where ${\cdot}^{\top}\cdot$ is the standard scalar product in the space $(\mathbb{R}^P)^Q$ and $\varsigma_n = (\varsigma_{n,1}, \ldots, \varsigma_{n,Q}) \in (\mathbb{R}^P)^Q$. For every $q \in \{1, \ldots, Q\}$, the components of vector $\varsigma_{n,q}$ are given by

$$(\forall\ \mathbf{s} \in \mathcal{A})\ \ \varsigma_{n,q}^{(\mathbf{s})} = \sum_{\substack{m=1 \\ m \neq n}}^{N} 1_{n,m}(\mathbf{s}) \big(\psi(I_n^{(\mathbf{s})}, I_m^{(i_1 - \alpha_{n,m} r_q, i_2)}) - \psi(I_n^{(\mathbf{s})}, I_m^{(i_1 - \alpha_{n,m} r_{q-1}, i_2)})\big)$$

where, for every $m \in \{1, \ldots, N\}$ with $n \neq m$, $1_{n,m}(\mathbf{s}) = 1$ if $\mathbf{s} \in \mathcal{D}_{n,m}$ and 0 otherwise.

In addition, the constraint set $\mathcal{R}$ can be decomposed as the intersection of two more tractable sets:

$$\theta_n \in \mathcal{R} \Leftrightarrow (\theta_n \in E_1 \text{ and } L\theta_n \in E_2) \tag{4.12}$$

where $E_1 = ([0,1]^P)^Q$, $E_2 = ([0, +\infty[^P)^{Q-1}$, and $L : (\mathbb{R}^P)^Q \to (\mathbb{R}^P)^{Q-1}$ is a linear operator, calculating the successive differences between the Q components of θ_n, defined as

$$(\forall\ \theta_n \in (\mathbb{R}^P)^Q) \qquad L\theta_n = (\zeta_{n,1}, \ldots, \zeta_{n,Q-1}), \tag{4.13}$$

and for all $\mathbf{s} \in \mathcal{A}$,

$$\begin{bmatrix} \zeta_{n,1}(\mathbf{s}) \\ \vdots \\ \zeta_{n,Q-1}(\mathbf{s}) \end{bmatrix} = L \begin{bmatrix} \theta_{n,1}(\mathbf{s}) \\ \vdots \\ \theta_{n,Q}(\mathbf{s}) \end{bmatrix} \tag{4.14}$$

where $L \in \mathbb{R}^{(Q-1)\times Q}$, is such that

$$L = \begin{bmatrix} 1 & -1 & 0 & \dots & \dots & 0 \\ 0 & 1 & -1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 & -1 \end{bmatrix}. \tag{4.15}$$

Finally, the total variation function can be expressed as

$$\rho(\theta_n) = \mu\, h(D\theta_n) \tag{4.16}$$

where $D : (\mathbb{R}^P)^Q \to (\mathbb{R}^{2P})^Q$ is a block diagonal operator whose q-th diagonal term with $q \in \{1, \dots, Q\}$ corresponds to a discrete gradient operator applied to $\theta_{n,q}$ computed in the horizontal and vertical directions. D is a linear operator given by

$$D = \begin{bmatrix} D_1 & 0 & \dots & & 0 \\ D_2 & 0 & & & \vdots \\ 0 & D_1 & & & \\ 0 & D_2 & & & \vdots \\ \vdots & \vdots & & & 0 \\ 0 & 0 & & 0 & D_1 \\ 0 & 0 & \dots & 0 & D_2 \end{bmatrix} \tag{4.17}$$

where D_1 and D_2 are the spatial gradient operators operating in the horizontal/vertical directions with periodic extension at boundaries. These operators correspond to 2D filters with frequency responses $1 - \exp(-2\pi\jmath\nu_1)$ and $1 - \exp(-2\pi\jmath\nu_2)$, where (ν_1, ν_2) are the 2D horizontal/vertical frequency variables ($\jmath^2 = -1$). The above function h is a sum of $\ell_{1,2}$ norms:[1]
$\big(\forall\ \delta_n = (\delta_{n,1}, \dots, \delta_{n,Q}) \in (\mathbb{R}^{2P})^Q\big)$

$$h(\delta_n) = \sum_{q=1}^{Q} (r_q - r_{q-1}) \|\delta_{n,q}\|_{1,2}.$$

With these definitions, the optimization problem reads

$$\underset{\theta_n \in (\mathbb{R}^P)^Q}{\text{minimize}}\ g_n(\theta_n) + \mu\, h(D\theta_n) \quad \text{s.t.} \quad \begin{cases} \theta_n \in E_1, \\ L\theta_n \in E_2, \end{cases} \tag{4.18}$$

and the corresponding constrained formulation reads

$$\underset{\theta_n \in (\mathbb{R}^P)^Q}{\text{minimize}}\ g_n(\theta_n) \quad \text{s.t.} \quad \begin{cases} \theta_n \in E_1, \\ L\theta_n \in E_2, \\ h(D\theta_n) \le \eta, \end{cases} \tag{4.19}$$

where $\eta \ge 0$.

[1] See Section 6.1 for the definition of the $\ell_{1,2}$-norm.

4.3.2 Proximal solution

Guidelines for addressing the above optimization problems are provided in Chambolle and Pock [2011]; Briceño-Arias and Combettes [2011]; Combettes and Pesquet [2012]; Setzer et al. [2010]; Parikh and Boyd [2013] by employing algorithms involving computation of proximity operators.

Proximity operators

Since g_n is a linear function, it can be further deduced from the standard properties of the proximity operator Combettes and Pesquet [2011] that the proximity operator of $\gamma g_n + \iota_{E_1}$ with $\gamma \in 0, +\infty]$ is given by:

$$\text{prox}_{\gamma g_n + \iota_{E_1}} = \text{P}_{E_1}(\cdot - \gamma \varsigma_n) \tag{4.20}$$

For the proximity operator of h, needed to solve Problem (4.18), we have: $\big(\forall\ \delta_n = (\delta_{n,1}, \ldots, \delta_{n,Q}) \in (\mathbb{R}^{2P})^Q\big)$

$$\text{prox}_{\mu h} \delta_n = \xi_n \tag{4.21}$$

where, $\xi_n = (\xi_{n,1}, \ldots, \xi_{n,Q}) \in (\mathbb{R}^{2P})^Q$, for every $q \in \{1, \ldots, Q\}$ and $\mathbf{s} \in \mathcal{A}$,

$$\xi_{n,q}^{(\mathbf{s})} = \max\left\{0, 1 - \frac{\mu(r_q - r_{q-1})}{\|\delta_{n,q}^{(\mathbf{s})}\|_{1,2}}\right\} \delta_{n,q}^{(\mathbf{s})}. \tag{4.22}$$

For the projection onto the lower level set of h, needed to solve Problem (4.19), we use the projection onto the ℓ_1 ball proposed by Van Den Berg and Friedlander [2008]. In our experiments, we focus on the constrained formulation (4.19) since the regularization parameter η may be easier to estimate.

Remark 4.3.1 The projection onto the polyhedron $\mathcal{R}$ may be computed in a finite number of iterations by solving the dual problem, which consists of a quadratic criterion subject to a positivity constraint. However, since at each iteration we need to compute a large number of projections (for each pixel in the image), this approach turns out to be slower than splitting the constraint $\mathcal{R}$ into $E_1 + E_2$.

M+LFBF algorithm

We propose here to use a primal-dual proximal algorithm whose main advantage is the absence of any matrix inversion. More precisely, we use the M+LFBF algorithm proposed in Combettes and Pesquet [2012]. The algorithm is initialized by setting the constants β and γ, and the variables $v_1^{[0]}$ and $v_2^{[0]}$ associated with the lower-level set constraint on h and E_2.

The main loop consists of computing the proximity operator given in (4.20), as well as the direct projections onto the different convex sets modeling prior information.

Algorithm 4 M+LFBF Combettes and Pesquet [2012]

$\beta = \left(\|D\|^2 + \|L\|^2\right)^{1/2}$
$\epsilon \in]0, (1+\beta)^{-1}[, \quad \gamma \in [\epsilon, (1-\epsilon)/\beta]$
$v_1^{[0]} \in (\mathbb{R}^{2P})^Q$, $v_2^{[0]} \in (\mathbb{R}^P)^{Q-1}$
$\theta_n^{[0]} \in (\mathbb{R}^P)^Q$

`for` $t = 0, 1, \ldots$
$p^{[t]} = D^\top v_1^{[t]} + L^\top v_2^{[t]}$
$\widehat{\theta}_n^{[t]} = \mathrm{P}_{E_1}\left(\theta_n^{[t]} - \gamma\varsigma_n - \gamma p^{[t]}\right)$

$\widehat{v}_1^{[t]} = v_1^{[t]} + \gamma D\theta_n^{[t]} - \gamma \mathrm{P}_{\mathrm{lev}_{\leq \eta} h}\left(\frac{v_1^{[t]} + \gamma D\theta_n^{[t]}}{\gamma}\right)$
$\widehat{v}_2^{[t]} = v_2^{[t]} + \gamma L\theta_n^{[t]} - \gamma \mathrm{P}_{E_2}\left(\frac{v_2^{[t]} + \gamma L\theta_n^{[t]}}{\gamma}\right)$

$v_1^{[t+1]} = \widehat{v}_1^{[t]} + \gamma D(\widehat{\theta}_n^{[t]} - \theta_n^{[t]})$
$v_2^{[t+1]} = \widehat{v}_2^{[t]} + \gamma L(\widehat{\theta}_n^{[t]} - \theta_n^{[t]})$

$\widehat{p}^{[t]} = D^\top \widehat{v}_1^{[t]} + L^\top \widehat{v}_2^{[t]}$
$\theta_n^{[t+1]} = \widehat{\theta}_n^{[t]} - \gamma\left(\widehat{p}^{[t]} - p^{[t]}\right)$

Remark 4.3.2

1- If $\gamma \in [\epsilon, \frac{(1-\epsilon)}{\beta}]$ where $\beta = \left(\|D\|^2 + \|L\|^2\right)^{1/2}$ and $\epsilon \in]0, (1+\beta)^{-1}[$, then the sequence $(\theta_n^{[t]})_{t\in\mathbb{N}}$ generated by the Algorithm 4 converges to a solution $\overline{\theta}_n$ to Problem 4.19 (Combettes and Pesquet [2012]).

2- In practice, to accelerate the convergence, the step-size γ has to be chosen as large as possible (for example, $\epsilon = 0.001/(\beta+1)$ and $\gamma = (1-\epsilon)/\beta$).

3- Due to the forms of operators L and D, we have

$$\|L\| = 2 \tag{4.23}$$

(that is the maximum singular value of matrix L). In addition, according

to (4.17), we have

$$\begin{aligned}\|D\| &= \|D^*D\|^{\frac{1}{2}} = \|D_1^*D_1 + D_2^*D_2\|^{\frac{1}{2}} \\ &= \max_{\nu_1,\nu_2}(|1-\exp(2\pi\jmath\nu_1)|^2 + |1-\exp(2\pi\jmath\nu_2)|^2)^{\frac{1}{2}} = 2\sqrt{2}. \quad (4.24)\end{aligned}$$

This shows that $\beta = 2\sqrt{3}$.

§ 4.4 Improved Taylor approximation

An alternative way to deal with multi-view disparity estimation consists of relaxing the data fidelity term by resorting to a linearization around an initial estimate. In Chapter 3, we presented an approach based on the first-order Taylor approximation. In the present section, we improve this approach by using a second-order Taylor approximation Malis [2004], which allows us to tighten the relaxation of the data fidelity term in the absence of illumination variation.

Recall that $(I_n)_{1\leq n\leq N}$ denote the N rectified views, and the intensities of corresponding pixels between two views are related by

$$I_n^{(i_1,i_2)} = I_m^{(i_1-u_{n,m}^{(\mathbf{s})},i_2)}. \quad (4.25)$$

When the magnitude of the difference between the fields $u_{n,m}$ and the initial estimate $\bar{u}_{n,m}$ is small enough, we obtain the following expression:
$\big(\forall(n,m)\in\{1,\dots,N\}^2,\ n\neq m\big)$ and $\big(\forall\mathbf{s}=(i_1,i_2)\in\mathcal{D}_{n,m}\big)$

$$\begin{aligned}I_m^{(i_1-u_{n,m}^{(\mathbf{s})},i_2)} \simeq & I_m^{(i_1-\bar{u}_{n,m}^{(\mathbf{s})},i_2)} - (u_{n,m}^{(\mathbf{s})}-\bar{u}_{n,m}^{(\mathbf{s})})\nabla^{(1)}I_m^{(i_1-\bar{u}_{n,m}^{(\mathbf{s})},i_2)} \\ & + \underbrace{\frac{1}{2}(u_{n,m}^{(\mathbf{s})}-\bar{u}_{n,m}^{(\mathbf{s})})^2\nabla^{(1,1)}I_m^{(i_1-\bar{u}_{n,m}^{(\mathbf{s})},i_2)} + o(|u_{n,m}^{(\mathbf{s})}-\bar{u}_{n,m}^{(\mathbf{s})}|^2)}_{O(|u_{n,m}^{(\mathbf{s})}-\bar{u}_{n,m}^{(\mathbf{s})}|^2)},\end{aligned} \quad (4.26)$$

where $\nabla^{(1)}$ (resp. $\nabla^{(1,1)}$) denotes the first-order (resp. the second order) derivative along the first space dimension (see Chapter 3, Equation 3.22).

By deriving the two members of Equation (4.25), we obtain:

$$\nabla^{(1)}I_n^{(\mathbf{s})} = \nabla^{(1)}(I_m^{(i_1-u_{n,m}^{(\mathbf{s})},i_2)}) = \nabla^{(1)}I_m^{(i_1-u_{n,m}^{(\mathbf{s})},i_2)}(1-\nabla^{(1)}u_{n,m}^{(\mathbf{s})}). \quad (4.27)$$

Since $u_{n,m}^{(\mathbf{s})}$ is locally smooth, it appears reasonable to assume that $\nabla^{(1)}u_{n,m}^{(\mathbf{s})} \simeq 0$. The results in Section 4.5 confirm the validity of this assumption. By performing a second term Taylor expansion around the initial estimate $\bar{u}$, we

obtain:

$$\begin{aligned}\nabla^{(1)} I_m^{(i_1-u_{n,m}^{(\mathbf{s})},i_2)} &\simeq \nabla^{(1)} I_m^{(i_1-\bar{u}_{n,m}^{(\mathbf{s})},i_2)} \\ &\quad - (u_{n,m}^{(\mathbf{s})} - \bar{u}_{n,m}^{(\mathbf{s})})\nabla^{(1,1)} I_m^{(i_1-\bar{u}_{n,m}^{(\mathbf{s})},i_2)} \\ &\quad + o(|u_{n,m}^{(\mathbf{s})} - \bar{u}_{n,m}^{(\mathbf{s})}|^2). \end{aligned} \tag{4.28}$$

Then, combining Equation (4.26) and (4.28), we obtain

$$\begin{aligned} I_m^{(i_1-u_{n,m}^{(\mathbf{s})},i_2)} &\simeq I_m^{(i_1-\bar{u}_{n,m}^{(\mathbf{s})},i_2)} \\ &\quad - \frac{1}{2}(u_{n,m}^{(\mathbf{s})} - \bar{u}_{n,m}^{(\mathbf{s})})\nabla^{(1)}(I_m^{(i_1-\bar{u}_{n,m}^{(\mathbf{s})},i_2)} + I_n^{(\mathbf{s})}) \\ &\quad + o(|u_{n,m}^{(\mathbf{s})} - \bar{u}_{n,m}^{(\mathbf{s})}|^2) \end{aligned} \tag{4.29}$$

Following the approach in Chapter 3, the criterion is reformulated as:

$$\begin{aligned} J_n(u_{n,k_n}) &= \sum_{\substack{m=1\\m\neq n}}^{N} \sum_{\mathbf{s}\in\mathcal{D}_{n,m}} \psi\Big(I_n^{(\mathbf{s})} - I_m^{(i_1-\alpha_{n,m}u_{n,k_n}^{(\mathbf{s})},i_2)}\Big) \\ &= \sum_{\substack{m=1\\m\neq n}}^{N} \sum_{\mathbf{s}\in\mathcal{D}_{n,m}} \psi\Big(T_{n,m}^{(\mathbf{s})} u_{n,k_n}^{(\mathbf{s})} - r_{n,m}^{(\mathbf{s})}\Big) \end{aligned} \tag{4.30}$$

where, by using (4.3), (4.25), and (4.29), for every $\mathbf{s} = (i_1, i_2) \in \mathcal{D}_{n,m}$,

$$\begin{cases} T_{n,m}^{(\mathbf{s})} = \frac{1}{2}\alpha_{n,m}\nabla^{(1)}\Big(I_m^{(i_1-\alpha_{n,m}\bar{u}_{n,k_n}^{(\mathbf{s})},i_2)} + I_n^{(\mathbf{s})}\Big), \\ r_{n,m}^{(\mathbf{s})} = I_m^{(i_1-\alpha_{n,m}\bar{u}_{n,k_n}(\mathbf{s}),i_2)} + \bar{u}_{n,k_n}^{(\mathbf{s})} T_{n,m}^{(\mathbf{s})} - I_n^{(\mathbf{s})}. \end{cases} \tag{4.31}$$

To minimize the criterion J_n, we need to incorporate additional prior information on the disparity field (see Chapter 3, Section 3.2.2.1), which yields a minimization problem expressed as:

$$\underset{u_{n,k_n}\in\mathbb{R}^P}{\text{minimize}} \;\; J_n(u_{n,k_n}) \quad \text{s. t.} \quad \begin{cases} u_{n,k_n} & \in S_{1,1}, \\ u_{n,k_n} & \in S_{1,2}, \end{cases} \tag{4.32}$$

where

$$\begin{aligned} S_{1,1} &= \{u_{n,k_n} \in \mathbb{R}^P \mid (\forall s \in \mathbb{R}^P) \;\; u_{\min} \leq u_{n,k_n}^{(s)} \leq u_{\max}\}, \\ S_{1,2} &= \{u_{n,k_n} \in \mathbb{R}^P \mid \mathrm{TV}(u_{n,k_n}) \leq \tau_2\}. \end{aligned}$$

To solve Problem 4.32, we will use Algorithm 5, which is a modified version of PPXA+ employed in Chapter 3.[2]

[2] In this section, the optimization is performed w.r.t. one variable instead of two. Recall that $S_{1,1}$ (resp. $S_{1,2}$) can be expressed as $L_1(C_{1,1})$ (resp. $L_2(C_{1,2})$), where L_1 is the identity matrix and L_2 the gradient operator matrix in $\mathbb{R}^{2P}$.

Algorithm 5 PPXA+

- *Initialization*

`For` $m = 1, \dots, N-1$

$L_{m+2} = I$

$(\omega_1, \omega_2, \dots, \omega_{N+1}) \in]0, +\infty[^{N+1}$

$(\mathbf{z}_i^{[0]})_{1 \leq i \leq N+1} \in \mathbb{R}^P \times \mathbb{R}^{2P} \times \mathbb{R}^P \cdots \times \mathbb{R}^P$

$Q_1 = \Big(\sum_{i=1}^{N+1} \omega_i (L_i)^\top L_i \Big)^{-1}$

$\mathbf{u}_{n,k_n}^{[0]} = Q_1 \Big(\sum_{i=1}^{N+1} \omega_i (L_i)^\top \mathbf{z}_{i,0} \Big)$

`For` $l = 0, 1, \dots$

$\mathrm{p}_1^{[l]} = P_{C_{1,1}}(\mathbf{z}_1^{[l]})\,, \quad \mathrm{p}_2^{[l]} = P_{C_{1,2}}(\mathbf{z}_2^{[l]})$

`For` $m = 1, \dots, N-1$

$\mathrm{p}_{m+2}^{[l]} = \mathrm{prox}_{\frac{J_n^{(m)}}{\omega_{m+2}}}(\mathbf{z}_{m+2}^{[l]})$

$\mathrm{c}^{[l]} = Q_1 \big(\sum_{i=1}^{N+1} \omega_i (L_i)^\top \mathrm{p}_i^{[l]} \big)$

$\mathbf{z}_1^{[l+1]} = \mathbf{z}_1^{[l]} + \lambda^{[l]} (L_1 (2\mathrm{c}^{[l]} - \mathbf{u}_{n,k_n}^{[l]}) - \mathrm{p}_1^{[l]})$

$\mathbf{z}_2^{[l+1]} = \mathbf{z}_2^{[l]} + \lambda^{[l]} (L_2 (2\mathrm{c}^{[l]} - \mathbf{u}_{n,k_n}^{[l]}) - \mathrm{p}_2^{[l]})$

`For` $m = 1, \dots, N-1$

$\mathbf{z}_{m+2}^{[l+1]} = \mathbf{z}_{m+2}^{[l]} + \lambda^{[l]} (2\mathrm{c}^{[l]} - \mathbf{u}_{n,k_n}^{[l]} - \mathrm{p}_{m+2}^{[l]})$

$\mathbf{u}_{n,k_n}^{[l+1]} = \mathbf{u}_{n,k_n}^{[l]} + \lambda^{[l]} (\mathrm{c}^{[l]} - \mathbf{u}_{n,k_n}^{[l]})$

§ 4.5 Numerical results

In this section, we present some numerical examples to illustrate the performance of the proposed method for multi-view disparity estimation. In this regard, we will show the improvement resulting from the use of more than two images for estimating the disparity map and the flexibility of our algorithm to handle both convex and nonconvex similarity measures ψ. The quality of the results is evaluated by computing the mean absolute error (MAE) and the percentage of unoccluded badly estimated pixels between the computed disparity $\widehat{u}_{n,k_n}$ and the ground truth u_{n,k_n}. A pixel is considered as unoccluded in I_n if it appears at least in one image of the considered image sequence $(I_m)_{1 \leq m \leq N, m \neq n}$. Thus, the percentage of false pixels reads

$$\mathrm{Err}_T(u_{n,k_n}, \widehat{u}_{n,k_n}) = \frac{1}{\Big| \bigcup\limits_{\substack{m=1 \\ m \neq n}}^{N} \mathcal{D}_{n,m} \Big|} \sum_{\mathbf{s} \in \bigcup\limits_{\substack{m=1 \\ m \neq n}}^{N} \mathcal{D}_{n,m}} 1_{\{|u_{n,k_n}^{(\mathbf{s})} - \hat{u}_{n,k_n}^{(\mathbf{s})}| > T\}}, \tag{4.33}$$

where T is a predefined threshold, as in Chapter 3. Since ground truth disparity maps for image sequences are hard to find, a set of experiments are conducted to evaluate the proposed framework.

4.5.1 Disparity map estimation from multiple images

Firstly, we limit ourselves to the dataset available at the Middlebury stereo vision website. Each dataset consists of 9 or 7 rectified views taken from equidistant cameras along a line. Ground truth disparity maps, created by using the structured lighting technique of Scharstein and Szeliski [2003], are only provided for two viewpoints in each dataset. In the case when we have 9 views, we estimate the disparity between view 2 and view 4, which is then evaluated using the provided ground truth divided by two. Finally, by adding more images, we demonstrate the efficiency of our method in recovering depth information from multiple images. Our analysis is conducted by estimating the disparity map $u_{2,4}$ on three image sequences (*Sawtooth*, *Teddy*, and *Cones* - Figure 4.2) and by comparing different similarity measures in the optimization criterion.

Multilabel approach

The main advantage of this method is the possibility to handle both convex and nonconvex error similarity measures ψ, namely the ℓ_1-norm ($\psi{=}|\cdot|$), the truncated ℓ_1-norm ($\psi{=}\ \mathbf{min}\{|\cdot|,\epsilon\}$), the $\ell_{\frac{1}{2}}$-norm ($\psi{=}|\cdot|^{\frac{1}{2}}$), and the truncated $\ell_{\frac{1}{2}}$-pseudo norm ($\psi{=}\ \mathbf{min}\{|\cdot|^{\frac{1}{2}},\epsilon\}$).

Sawtooth Teddy Cones

Figure 4.2: *Tested image sequences*

The first line of Tables 4.1, 4.2 and 4.3 illustrates the MAE and Err_1 indices of the estimated disparity map $u_{2,4}$, in the case when only two images are used in the estimation process (the standard stereo scenario with left and right images). The second line of these tables (resp. the third line) collects the MAE and Err_1 indices in the case when we estimate the disparity map $u_{2,4}$ by using three (resp. five) multi-view images.[3]

[3]Note that, in Tables 4.1, 4.2, and 4.3, the five multi-view images (3rd line) contain the three multi-view images used in the previous experiment (2nd line), which in turn contain the left/right stereo images (1st line).

Table 4.1: *The* (MAE, Err_1) *results on Sawtooth image, with* Q*=14,* r_0*=2,* r_Q*=9 and* $(\forall\, q \in \{2,\dots,Q\})\ r_q - r_{q-1} = 0.5$*, and* $\eta = 6.9 \times 10^3$.

	ℓ_1	$\ell_{1,\epsilon}$	$\ell_{\frac{1}{2}}$	$\ell_{\frac{1}{2},\epsilon}$
$(I_2,\ I_4)$	(0.17, 1.96)	ϵ=9, **(0.17, 1.23)**	(0.16, 1.84)	ϵ=6, (0.19, 1.31)
(I_0, I_2, I_4)	(0.15, 1.76)	ϵ=8, **(0.15, 1.16)**	(0.14, 1.66)	ϵ=6, (0.18, 1.28)
$(I_0, I_1, I_2, I_3, I_4)$	(0.19, 1.56)	ϵ=8, **(0.18, 1.08)**	(0.18, 1.47)	ϵ=6, (0.16, 1.27)

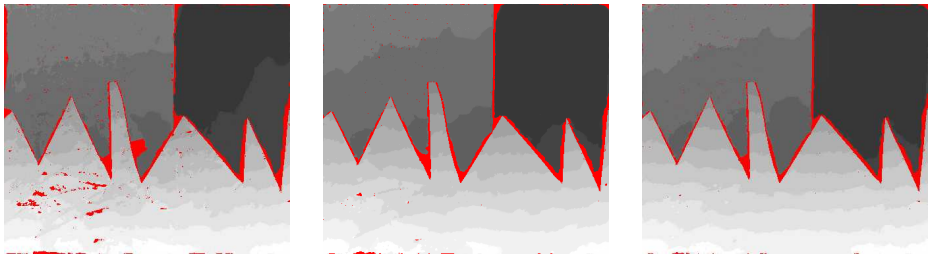
Table 4.2: *The* (MAE, Err_1) *results on Teddy image, with* Q*=20,* r_0*=6,* r_Q*=26 and* $(\forall\, q \in \{2,\dots,Q\})\ r_q - r_{q-1} = 1$*, and* $\eta = 23 \times 10^3$.

	ℓ_1	$\ell_{1,\epsilon}$	$\ell_{\frac{1}{2}}$	$\ell_{\frac{1}{2},\epsilon}$
$(I_2,\ I_4)$	(0.64, 5.70)	ϵ=25, **(0.56, 4.29)**	(0.60, 5.37)	ϵ=19, (0.64, 5.90)
(I_0, I_2, I_4)	(0.52, 5.33)	ϵ=23, **(0.48, 4.08)**	(0.50, 5.02)	ϵ=19, (0.61, 5.67)
$(I_0, I_1, I_2, I_3, I_4)$	(0.46, 4.59)	ϵ=23, **(0.48, 3.82)**	(0.44, 4.33)	ϵ=19, (0.61, 5.64)

Table 4.3: *The* (MAE, Err_1) *results on Cones image, with* Q*=26,* r_0*=2,* r_Q*=28 and* $(\forall\, q \in \{2,\dots,Q\})\ r_q - r_{q-1} = 1$*, and* $\eta = 27 \times 10^3$.

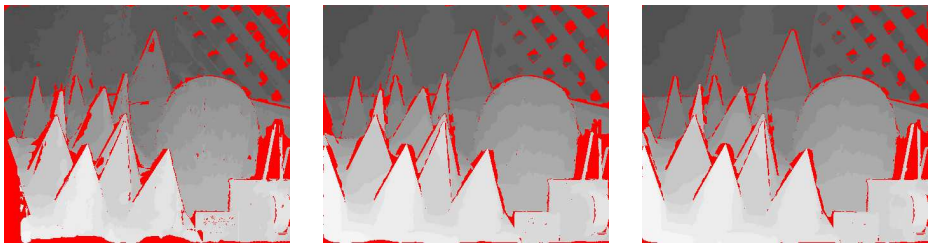
	ℓ_1	$\ell_{1,\epsilon}$	$\ell_{\frac{1}{2}}$	$\ell_{\frac{1}{2},\epsilon}$
$(I_2,\ I_4)$	(0.49, 7.34)	ϵ=27, **(0.41, 5.24)**	(0.49, 6.96)	ϵ=25, (0.59, 7.9)
(I_0, I_2, I_4)	(0.49, 4.71)	ϵ=27, **(0.45, 3.72)**	(0.42, 4.44)	ϵ=25, (0.49, 5.23)
$(I_0, I_1, I_2, I_3, I_4)$	(0.44, 4.04)	ϵ=27, **(0.43, 3.68)**	(0.43, 3.81)	ϵ=25, (0.48, 5.19)

Moreover, we illustrate in Figure 4.3 the estimated disparity maps. The red regions indicate the pixels associated with a wrong disparity value.



(0.17, 1.23) (0.15, 1.16) (0.18, 1.08)

(0.56, 4.29) (0.48, 4.08) (0.48, 3.82)



(0.41, 5.24) (0.45, 3.72) (0.43, 3.68)

Figure 4.3: *The disparity map $u_{2,4}$ and the indices* (MAE, Err_1) obtained by using the proposed method with the truncated ℓ_1-norm. From left to right: stereo, 3 views, 5 views. From top to button: Sawtooth, Teddy, and Cones.

Improved First-Order Taylor approximation To investigate the performance of the improved first-order Taylor approximation, we firstly compare it to the standard Taylor approximation (presented in Chapter 3) for two different error measures : the ℓ_1-norm ($\psi(\cdot) = |\cdot|$) and the squared ℓ_2-norm ($\psi(\cdot) = |\cdot|^2$). As shown in Table 4.4, the improved first-order expansion reduces the error introduced by the Taylor approximation around the same initial estimate. Note that the improved approximation can only be applied to the disparity estimation problem for gray level images without illumination variation. This is the reason why it was not employed in Chapter 3, where the considered images were affected by illumination variations.

Table 4.4: (MAE, Err_1) *results on* $u_{2,4}$ *for three different stereo pairs using* $\psi = |\cdot|$ *and* $\psi = |\cdot|^2$*. Comparison is performed between the first-Order Taylor linearization and the improved one.*

Stereo pair	τ_2	$[u_{\min}, u_{\max}]$	$\psi = \lvert\cdot\rvert^p$	1st-order	Improved 1st-order
Sawtooth	14×10^3	[3, 18]	$p=1$	(0.29, 2.81)	(0.26, 2.79)
			$p=2$	(0.34, 3.12)	(0.30, 3.10)
Teddy	46×10^3	[0, 26]	$p=1$	(0.61, 7.81)	(0.60, 7.64)
			$p=2$	(0.62, 7.93)	(0.62, 7.74)
Cones	68×10^3	[0, 27]	$p=1$	(0.63, 7.84)	(0.66, 7.41)
			$p=2$	(0.69, 7.80)	(0.62, 7.61)

Performance comparison of different methods

In this section, we compare the approaches based on multi-labeling and Taylor approximation with the method in Woodford et al. [2009] based on graph-cuts.

To do so, we employ some standard datasets taken from the Middlebury Database ("Sawtooth", "Teddy" and "Cones", Figure 4.2) and a synthetic dataset called "Sponza", which consists of five images with the corresponding ground-truth disparity maps (Figure 4.4) generated by using the photo-realistic rendering Embree software.[4] Each disparity map is computed by exploiting depth information and image characteristics (Chapter 3, Equation 3.6).

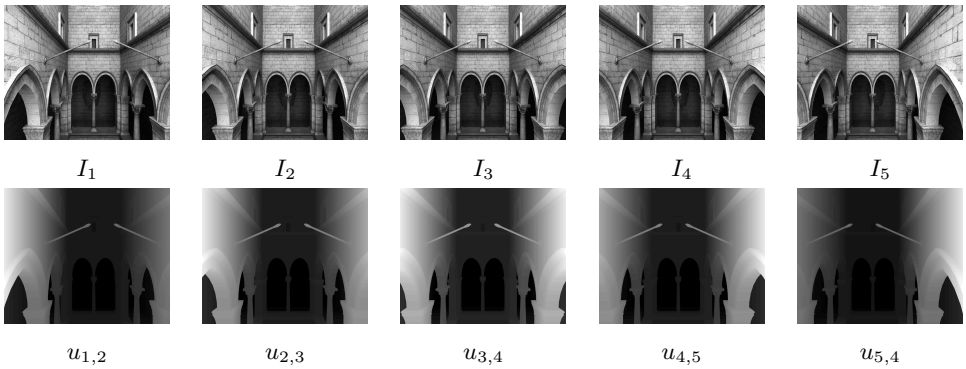
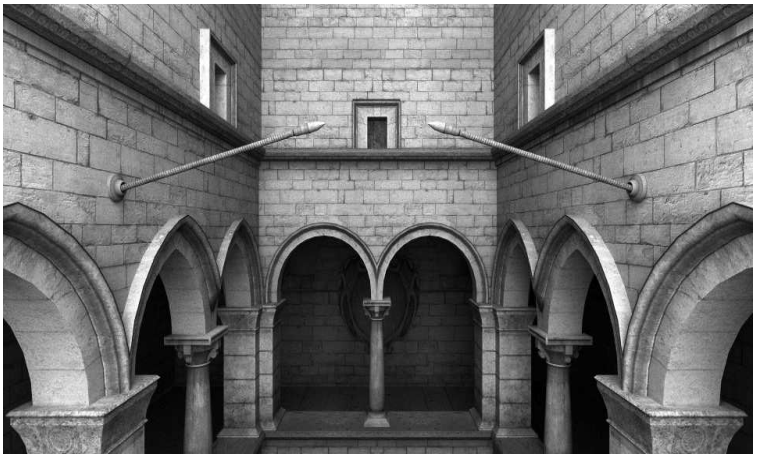


Figure 4.4: *"Sponza" sequence. Each image is of size* 480×960 *(PPM grayscale format). Calibration parameters for the equidistant cameras are known.*

[4] http://software.intel.com/en-us/articles/embree-photo-realistic-ray-tracing-kernels-0



I_3

$u_{3,4}$

Figure 4.5: *Image 3 of "Sponza" sequence.*

Table 4.5 collects the percentage of error pixels obtained with the improved-first-order Taylor approximation, the multilabel approach, and the method in Woodford et al. [2009]. One can see that the multilabel approach performs better than the others, as it employs a tighter convex relaxation of the data fidelity term. This can be also observed in Figure 4.6, where the disparity map of "Sponza" dataset estimated with the aforementioned methods is displayed. Nonetheless, the Taylor-approximation approach is much faster (see Chapter 3, Section 3.4), i.e. it trades off the precision with the speed.

Implementation parameters To perform the above experiments, we used the following setups for the considered algorithms:

- **Multilabel.** For Sawtooth, Teddy, and Cones, we fix the same parameters as in Tables 4.1, 4.2, and 4.3. For Sponza, sequence, we have: $\psi = \ell_{1,\epsilon}$ (ϵ=40), Q=80, $r_0 = 11$, $r_Q = 91$, $(\forall\ q \in \{2, \dots, Q\}), r_q - r_{q-1} = 1$, and $\eta = 54 \times 10^3$.

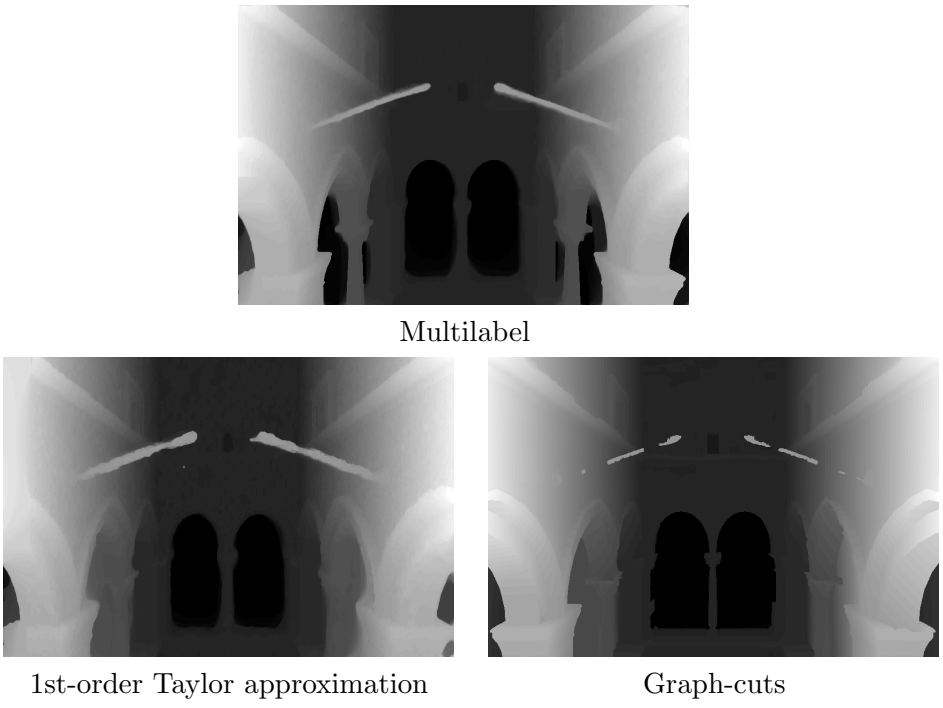
- **Improved Taylor approximation** For Sponza we have $\psi = \ell_1$, $\tau_2 = 11 \times 10^4$, and [$u_{\min}$, $u_{\max}$]=[11,91]. (For Sawtooth, Teddy, and Cones, see Table 4.4).

- **Graph cuts** Woodford et al. [2009] The authors incorporated the second-order smoothness priors for a global optimization of the stereo reconstruction. They employed triple cliques to estimate depth by using graph cuts. We have used the code publicly available[5] for gray level images. The default setting has been unchanged (except for Sponza).

Table 4.5: *Percentage of bad pixels obtained with the considered methods.*

$u_{2,4}$		$(I_2,\ I_4)$	(I_0, I_2, I_4)	$(I_0, I_1, I_2, I_3, I_4)$
Sawtooth	Multilabel	**1.23**	**1.16**	**1.08**
	Improved 1st-order	2.79	2.13	1.86
	Graph cuts	1.75	1.66	1.70
Teddy	Multilabel	**4.29**	**4.08**	**3.82**
	Improved 1st-order	7.64	5.43	5.00
	Graph cuts	4.57	4.46	3.83
Cones	Multilabel	**5.24**	**3.72**	**3.68**
	Improved 1st-order	7.41	5.4	4.65
	Graph cuts	7.02	6.02	4.65
$u_{3,4}$		$(I_3,\ I_4)$	(I_2, I_3, I_4)	$(I_1, I_2, I_3, I_4, I_5)$
Sponza	Multilabel	**5.4**	**4.2**	**3.10**
	Improved 1st-order	7.71	5.93	5.49
	Graph cuts	6.50	5.47	4.84

[5] `http://www.robots.ox.ac.uk/~ojw/software.htm`



Multilabel

1st-order Taylor approximation

Graph-cuts

Figure 4.6: *Visual comparison of "Sponza" disparity map $u_{3,4}$ estimated by the considered methods using 5 images.*

4.5.2 Disparity map sequence estimation

In this section, we show the reconstruction result of a 3D scene from a sequence of disparity maps. Figure 4.7 displays such a disparity sequence, where five images were used to estimate each map. The improvement is obvious with respect to the stereo case, using the ℓ_1 similarity measure.

Finally, an extended disparity map is obtained from the disparity map sequence (see Figure 4.8) as well as a 3D reconstruction of the "Sponza" sequence (see Figure 4.9) performed from the estimated maps. Five views, of the 3D scene, are presented from different camera position: Left, Down, Center, Up and Right.

§ 4.6 Conclusion

In this chapter, we have proposed a multilabel optimization approach for performing multi-view disparity estimation. Rather than estimating a single depth map, we associate a depth map with each input image (or a subset of those images). We have proposed two different approaches to estimate the depth map of each chosen reference frame, which differ in the type of convex relaxation employed. The first approach employs a relaxation based

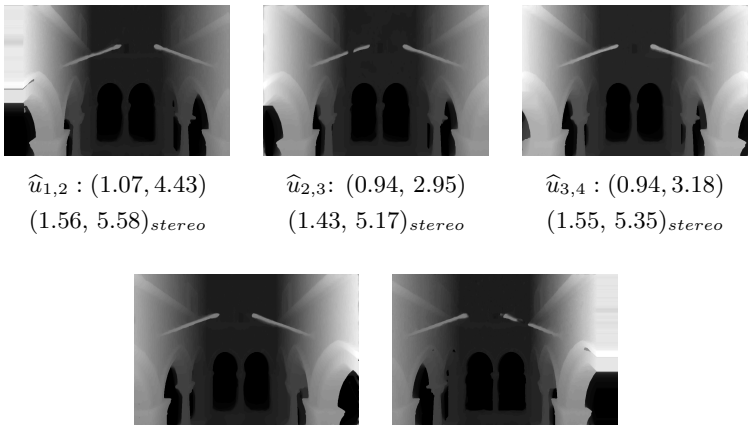


Figure 4.7: (MAE, Err_2) *obtained using five images for each disparity map estimation,* Q*=80,* r_0*=11,* r_Q*=91 and* $(\forall\ q \in \{2,\ldots,Q\})$, $r_q - r_{q-1} = 1$. *For comparison, we indicate the values of* $(\mathrm{MAE}, \mathrm{Err}_2)_{\mathrm{stereo}}$ *obtained in the stereo case when* u_{n,k_n} *is estimated from (*I_n*,*I_{k_n}*).* ψ *is the* ℓ_1*-norm.*

on multi-labeling, with the possibility of using nonconvex similarity measures. So doing, there is no more dependence of the method performance on the initial estimate, at the cost of an increased complexity. The second approach employs a relaxation based on an improved first-order Taylor approximation around an initial estimate. The numerical issues related to the computation of our disparity sequence have been addressed by using a proximal algorithm. The proposed method leads to improved results compared with other methods. It is also worth emphasizing that this method can be efficiently implemented on multicore architectures.

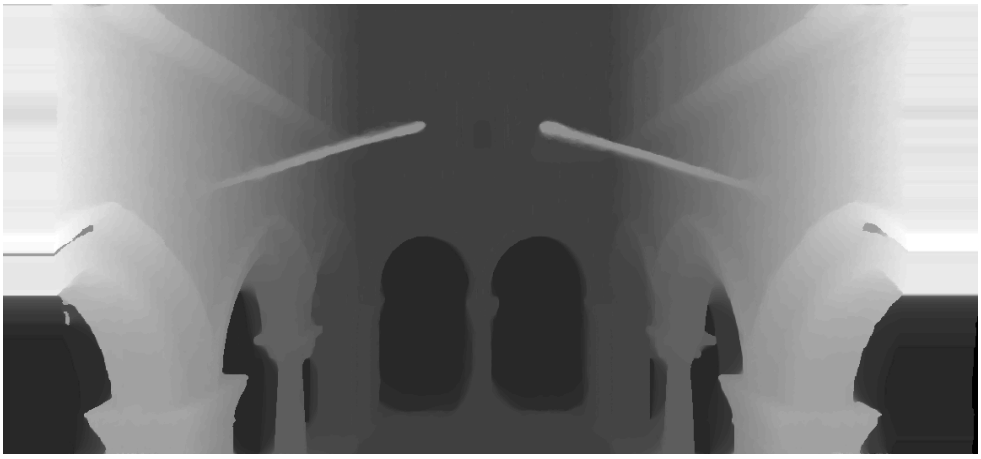


Figure 4.8: *The extended disparity map constructed using the sequence of disparity maps:* $(u_{1,2}, u_{2,3}, u_{3,4}, u_{4,5}, u_{5,4})$

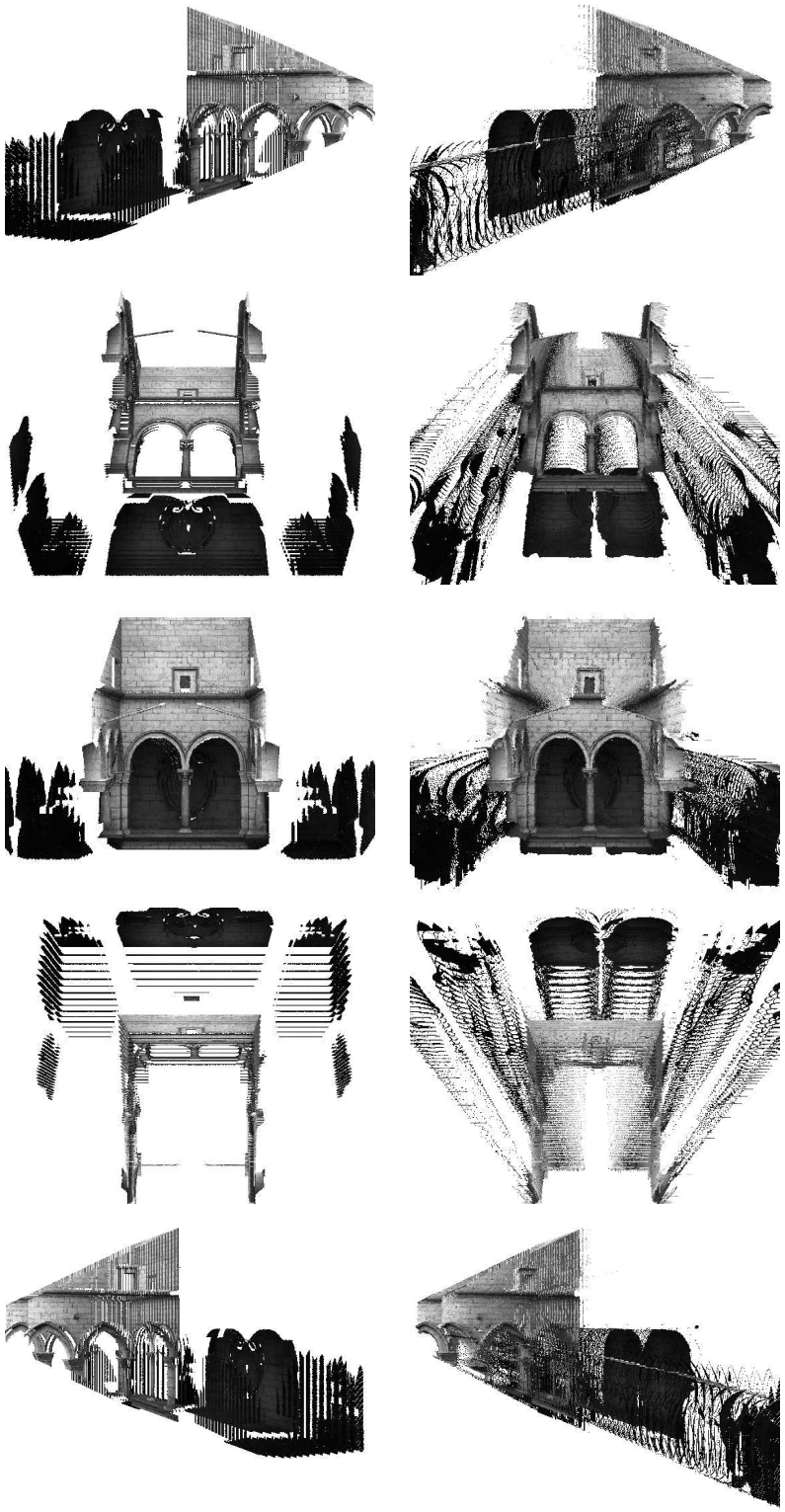


Figure 4.9: *3D reconstruction from the true disparity sequence (Left column) and from the proposed method (Right column). The camera position from the top to the buttom: Left, Down, Center, Up, and Right.*

- Chapter 5 -

Proximity operators of discrete information divergences

"Knowledge is of two kinds. We know a subject ourselves, or we know where we can find information upon it. "
`Samuel Johnson`

The notion of information measure plays a central role in the fields of probability, statistics, and information theory, as it allows one to assess how close two distributions are from each other. Among the large panel of available measures, a special attention has been paid to φ-divergences, e.g. Kullback-Leibler, Jeffreys-Kullback, Hellinger, Chi-Square, I_α, and Renyi divergences. Indeed, these divergences have been studied extensively in many different contexts, including convex optimization. In the latter context, the solution of optimization problems may be challenging.

One of the main limitations of existing convex optimization methods is that they restrict the use of φ-divergences to problems in which the minimization is performed w.r.t. one of the divergence arguments. A possible solution to overcome this issue is to resort to alternating minimization approaches, which however require specific assumptions to be valid Tseng [1997]; Bauschke et al. [2006]. In the present chapter, we derive closed-form expressions of the proximity operators of the aforementioned φ-divergences. These novel results enrich the list of functions for which the proximity operator can be easily computed. Building upon these results, we are able to address a wider range of convex optimization problems by means of proximal methods.

§ 5.1 INTRODUCTION

Claude Shannon [1948], in his seminal paper "A Mathematical Theory of Communication", introduced a powerful mathematical framework to quantify the notion of information, laying the foundations for the field of information theory and originating a major revolution in communications and related fields. The power of the framework introduced by Shannon is reflected in his two main results known as the Source Coding Theorem and the Channel Coding Theorem.

In the same spirit as Shannon [1948], it was natural to investigate functionals which allow one to quantify how much information is shared between two probability distributions. As pointed out by Liese and Vajda [1987], the origins of these ideas can be traced back to the early twentieth century literature in the works by Pearson [1900] and Hellinger [1909], although research in this area became much more prolific after the publication of Shannon's paper Shannon [1948]. Motivated by this work, Kullback and Leibler [1951] introduced the information measure now known as the Kullback-Leibler divergence within the context of hypothesis testing.

Yet a higher level of generalization in the area of probabilistic divergences was achieved by the work by Csiszár [1963] (and independently also Ali and Silvey [1966]; Cover and Thomas [1991]), who introduced the notion of φ-divergences for probability distributions, a framework which encompasses a vast number of information measures currently used in the literature, including the Kullback-Leibler (KL), Jeffreys-Kullback (JK), Hellinger (Hel), Chi square, and I_α divergences. The works by Arimoto [1971]; LeCam [1973]; Burbea and Rao [1982]; Basseville [1989]; Vajda [1989]; Barron et al. [1990]; Liese and Vajda [2006]; Amari [2009]; Österreicher [2013]; Nielsen and Nock [2013]; Basseville [2013] have enriched this formalism and its applications in an extensive manner.

The KL divergence is known to play a prominent role in the computation of channel capacity and rate-distortion functions. One can address these problems with the celebrated alternating minimization algorithm proposed by Blahut [1972]; Arimoto [1972]. However, other approaches based on geometric programming Chiang and Boyd [2004] may provide more efficient numerical solutions. The generalized KL divergence is also known as the I-divergence. It plays an important role in inverse problems for recovering the signal of interest in the presence of Poisson noise. In such a case, the I-divergence is usually employed as a data fidelity term. For instance, an alternating projection technique was proposed in Byrne [1993], where both the data fidelity term and the regularization term are based on the KL divergence. The problem was formulated in a similar manner in Richardson [1972], whereas in Fessler [1995]; Piro et al. [2008]; Dupé et al. [2009]; Zanella et al. [2009];

Piro et al. [2010]; Pustelnik et al. [2011]; Teuber et al. [2012]; Carlavan and Blanc-Féraud [2012]; Anthoine et al. [2012] more general forms of the regularization functions are considered. In particular, the developments in Dupé et al. [2009]; Pustelnik et al. [2011]; Teuber et al. [2012]; Anthoine et al. [2012] are grounded on proximal splitting methods. These proximal tools were recently shown to offer both efficient and flexible solutions to a wide class of possibly nonsmooth convex minimization problems (see Combettes and Pesquet [2011]; Parikh and Boyd [2013] and references therein). Note that, in all of the aforementioned works, one of the two variables of the KL divergence is fixed.

As the KL divergence is a Bregman distance, optimization problems involving this function can also be addressed by using the alternating minimization approach proposed in Bauschke et al. [2006]. However, the required optimization steps may be difficult to implement and the convergence of the algorithm is only guaranteed under restrictive conditions. Moreover, a Kullback-Leibler proximal algorithm generalizing the EM algorithm was investigated in Chrétien and Hero [2008]. The KL divergence is then used as a metric for the maximization of a log-likelihood function, rather than being one of the terms of the objective function. Recently, the authors in Vemuri et al. [2011]; Meizhu et al. [2010] defined a new similarity measure called total Kullback Leibler divergence, which has the potential benefit of being invariant to axis rotations. On the other side, the classical symmetrization of Kullback-Leibler divergence (also known as Jeffreys-Kullback divergence Jeffreys [1946]) has been recently used in the k-means algorithm (as a replacement of the squared difference) Nielsen and Nock [2009]; Nielsen [2013], yielding analytical expressions of the centroids in terms of the Lambert W function.

The Hellinger divergence was named after Ernst Hellinger [1909] and later rediscovered in Beran [1977]; Devijver and Kittler [1982]; Gibbs and Su [2002]; Liese and Vajda [2006]; Rauber et al. [2008] under different names (such as Jeffreys-Matusita distance). In the field of information theory, the Hellinger divergence is commonly used in the analysis of nonparametric density estimation LeCam [1973]; van de Geer [1993], statistics and Data Analytics Chang-Hwan [2012], and machine learning Cieslak et al. [2012].

The Chi-square divergence can be traced back to Pearson [1900], where it was used to quantitatively assess whether an observed phenomenon tends to confirm or deny a given hypothesis. This work heavily contributed to the development of modern statistics: in 1984, the journal *Science* referred to Pearson [1900] as "one of the 20 most important scientific breakthroughs". Moreover, Chi-square divergence was also successfully applied in different contexts, such as information theory and signal processing, as a dissimilarity measure between two probability distributions Cover and Thomas [1991];

Park et al. [2011].

The I_α-divergence was originally proposed in Chernoff [1952] to statistically evaluate the efficiency of an hypothesis test. Subsequently, it was recognized as an instance of more general divergence classes Ali and Silvey [1966], such as the class of φ-divergences Csiszár [1974] and the class of Bregman divergences Amari [2009]. Indeed, I_α-divergence has been extended by many researchers as in Liese and Vajda [1987]; Zhang [2004]; Minka [2005]; Amari [2009], it has been considered in the context of Non-negative Matrix Factorization where the hyper-parameter α is associated with characteristics of a learning machine Cichocki et al. [2008].

Rényi divergence was introduced in Rényi [1961] as a measure of information related to Rényi entropy. According to the definition given in Harremoës [2006], Rényi divergence measures "how much a probabilistic mixture of two codes can be compressed". It has been studied and applied in many areas Vajda [1989]; Liese and Vajda [1987, 2006] including image registration problems Hero et al. [2002].

To the best of our knowledge, existing approaches for optimizing convex criteria involving these divergences are often restricted to problems where the minimization is performed w.r.t. one of the arguments of the divergence, or they are based on an alternating minimization process which requires specific assumptions to be valid. In the context of proximal methods, this lacuna can be explained by the fact that there exist few results concerning the proximity operator of non-separable convex functions (as opposed to separable ones, for which many results are available Chaux et al. [2007]; Combettes and Pesquet [2011]). Some examples of such non-separable functions are: a quadratic function composed by a linear operator Combettes and Pesquet [2008], a separable function composed with an orthonormal or semi-orthogonal linear operator Combettes and Pesquet [2008], the ℓ_2-norm Combettes and Pesquet [2008], the quadratic-over-linear function Benamou and Brenier [2000]; Benamou et al. [2002], the indicator function of a closed convex Combettes and Pesquet [2008] or the epigraph of a lower semi continuous convex function Chierchia et al. [2013], and a convex function composed with the distance to a closed convex set Combettes and Pesquet [2008].

In this work, we develop a novel proximal method that allows us to address more general forms of optimization problems. The expressions we derive for these divergences enrich the list of functions for which such proximity operators can be easily computed. In addition to its flexibility, the proposed approach leads to parallel proximal algorithms that can be efficiently implemented on multicore architectures.

The remaining of the chapter is organized as follows. In Section 5.2, we present the general form of the addressed optimization problem and introduce the notation used in this work. In Section 5.3, we study the proximity operators of φ-divergences and some of their properties. The

aforementioned divergence proximity operators are described in Section 5.4. In Section 5.5, we explain how the forms of these proximity operators can be employed to deduce the expression of the projection onto the epigraph of some convex functions. Finally, Section 5.6 concludes the chapter.

§ 5.2 Problem formulation

5.2.1 Optimization problem

The objective of this chapter is to address convex minimization problems involving a discrete information divergence D, which can be formulated as follows:

Problem 5.2.1 Let D be a function in $\Gamma_0(\mathbb{R}^P \times \mathbb{R}^P)$. Let A and B be matrices in $\mathbb{R}^{P\times N}$, and let u and v be vectors in $\mathbb{R}^P$. For every $s \in \{1,\dots,S\}$, let R_s be a function in $\Gamma_0(\mathbb{R}^{K_s})$ and $T_s \in \mathbb{R}^{K_s\times N}$. We want to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; D(Ax+u, Bx+v) + \sum_{s=1}^{S} R_s(T_s x). \tag{5.1}$$

Hereabove $\Gamma_0(\mathcal{H})$ denotes the class of convex functions f defined on a real Hilbert space $\mathcal{H}$ and taking their values in $]-\infty,+\infty]$ which are lower-semicontinuous and proper (i.e. their domain $\operatorname{dom} f$ on which they take finite values is nonempty). Allowing functions to take the $+\infty$ value makes Problem 5.2.1 a versatile one. For example, the involved functions may take finite values only for nonnegative-valued arguments. This also allows us to include in this formulation convex constrained problems by letting R_s with $s \in \{1,\dots,S\}$ be equal to the indicator function ι_C of some nonempty closed convex set C. Recall that the indicator of a nonempty closed convex subset C of $\mathcal{H}$ is defined as

$$(\forall x \in \mathcal{H}) \qquad \iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases} \tag{5.2}$$

In applications to inverse problems, R_s may also be some regularization function serving to enforce the smoothness of the sought solution or to model some additional prior information, e.g. its sparsity after some appropriate linear transform T_s.

A special case of interest in information theory is obtained when $u = v = 0$, $N = 2P$, and

$$A = [I_P \;\; 0], \qquad B = [0 \;\; I_P] \tag{5.3}$$

where I_P is the $P \times P$ identity matrix. By decomposing the vector x as

$$x = [p^\top q^\top]^\top \tag{5.4}$$

where $(p, q) \in (\mathbb{R}^P)^2$, and by setting

$$(\forall s \in \{1, \ldots, S\}) \qquad T_s = [U_s \;\; V_s] \tag{5.5}$$

where U_s and V_s are matrices in $\mathbb{R}^{K_s \times P}$, Problem 5.2.1 becomes:

Problem 5.2.2 Let D be a function in $\Gamma_0(\mathbb{R}^P \times \mathbb{R}^P)$. For every $s \in \{1, \ldots, S\}$, let R_s be a function in $\Gamma_0(\mathbb{R}^{K_s})$, and let U_s and V_s be matrices in $\mathbb{R}^{K_s \times P}$. We want to

$$\underset{p \in \mathbb{R}^P,\, q \in \mathbb{R}^P}{\text{minimize}} \;\; D(p, q) + \sum_{s=1}^{S} R_s(U_s p + V_s q). \tag{5.6}$$

When the variables p and q correspond to probability mass distributions, a standard example of convex constraint subset of $\mathbb{R}^{2P}$ is the product of two unit simplexes of $\mathbb{R}^P$:

$$C = \Big\{ (p^{(1)}, \ldots, p^{(P)}, q^{(1)}, \ldots, q^{(P)})^\top \in [0,1]^{2P} \;\Big|\; \sum_{i=1}^{P} p^{(i)} = 1 \text{ and } \sum_{i=1}^{P} q^{(i)} = 1 \Big\}. \tag{5.7}$$

5.2.2 Considered class of divergences

We will focus on additive information measures of the form

$$\big(\forall p = (p^{(i)})_{1 \le i \le P} \in \mathbb{R}^P\big)\big(\forall q = (q^{(i)})_{1 \le i \le P} \in \mathbb{R}^P\big) \quad D(p, q) = \sum_{i=1}^{P} \Phi(p^{(i)}, q^{(i)}) \tag{5.8}$$

where Φ is defined as follows

$$\big(\forall (\upsilon, \xi) \in \mathbb{R}^2\big) \quad \Phi(\upsilon, \xi) = \begin{cases} \xi\, \varphi\Big(\dfrac{\upsilon}{\xi}\Big) & \text{if } \upsilon \in [0, +\infty[\text{ and } \xi \in \,]0, +\infty[\\ \upsilon \lim\limits_{\zeta \to +\infty} \dfrac{\varphi(\zeta)}{\zeta} & \text{if } \upsilon \in \,]0, +\infty[\text{ and } \xi = 0 \\ 0 & \text{if } \upsilon = \xi = 0 \\ +\infty & \text{otherwise} \end{cases} \tag{5.9}$$

and $\varphi \colon \mathbb{R} \to [0, +\infty]$ belongs to $\Gamma_0(\mathbb{R})$ and is twice differentiable on $]0, +\infty[$.[1] So, Φ is the perspective function Bauschke and Combettes [2011] of φ on $[0, +\infty[\times \,]0, +\infty[$ and D is a function in $\Gamma_0(\mathbb{R}^{2P})$.
If φ is strictly convex such that

$$\varphi(1) = \varphi'(1) = 0 \tag{5.10}$$

[1] The existence of $\lim_{\zeta \to +\infty} \varphi(\zeta)/\zeta$ in $[0, +\infty]$ is guaranteed [Hiriart-Urruty and Lemaréchal, 1996, Section 2.3].

D belongs to the celebrated class of φ-divergences Csiszár [1963, 1967]. Then,

$$\big(\forall (p,q) \in [0,+\infty[^P \times [0,+\infty[^P\big) \qquad D(p,q) \geq 0 \tag{5.11}$$

$$D(p,q) = 0 \quad \Leftrightarrow \quad p = q. \tag{5.12}$$

Examples of φ-divergences will be provided in Sections 5.4.1, 5.4.2, 5.4.3, 5.4.4 and 5.4.6. For a thorough investigation of the rich properties of φ-divergences, the reader is refered to Csiszár [1963]; Ali and Silvey [1966]; Basseville [1989]. Other divergences (e.g. Rényi divergence) are expressed as

$$\big(\forall (p,q) \in \mathbb{R}^P \times \mathbb{R}^P\big) \qquad D_g(p,q) = g\big(D(p,q)\big) \tag{5.13}$$

where g is some increasing function. Then, provided that $g\big(\varphi(1)\big) = 0$, for every $(p,q) \in ([0,+\infty[^P)^2$ such that $[p^\top \ q^\top]^\top \in C$ where C is given by (5.7), we have $D_g(p,q) \geq 0$. Note that, from an optimization standpoint, minimizing D or D_g (possibly subject to constraints) makes no difference, so that we will only address problems involving D in the rest of this chapter.

5.2.3 Proximity operators

Proposition 5.2.3 Moreau [1965]; Bauschke and Combettes [2011] *Let $f \in \Gamma_0(\mathcal{H})$. Then,*

(i) *for every $\overline{x} \in \mathcal{H}$, $\operatorname{prox}_f \overline{x} \in \operatorname{dom} f$.*

(ii) *For every $(x,\overline{x}) \in \mathcal{H}^2$, $x = \operatorname{prox}_f \overline{x} \ \Leftrightarrow \ \overline{x} - x \in \partial f(x)$, where $\partial f(x)$ denotes the subdifferential of f at x. If f is Gâteaux differentiable at x, $\partial f(x) = \{\nabla f(x)\}$ where $\nabla f(x)$ denotes the gradient of f at x.*

(iii) *For every $(\overline{x},z) \in \mathcal{H}^2$,*

$$\operatorname{prox}_{f(\cdot+z)} \overline{x} = \operatorname{prox}_f(\overline{x}+z) - z. \tag{5.14}$$

(iv) *For every $(\overline{x},z) \in \mathcal{H}^2$ and for every $\alpha \in \mathbb{R}$,*

$$\operatorname{prox}_{f+\cdot^\top z+\alpha} \overline{x} = \operatorname{prox}_f(\overline{x}-z). \tag{5.15}$$

(v) *Let f^* be the conjugate function of f ($f^* \in \Gamma_0(\mathcal{H})$ defined by $f^* : u \mapsto \sup_{x\in\mathcal{H}}(\langle x \mid u\rangle - f(x))$). For every $\overline{x} \in \mathcal{H}$,*

$$\operatorname{prox}_{f^*} \overline{x} = \overline{x} - \operatorname{prox}_f \overline{x}. \tag{5.16}$$

(vi) *Let $\mathcal{G}$ be a real Hilbert space and let $T\colon \mathcal{G} \to \mathcal{H}$ be a bounded linear operator, the adjoint of which is denoted by T^*. If $T \circ T^* = \kappa \operatorname{Id}$ where $\kappa \in \left]0,+\infty\right[$, then*

$$\operatorname{prox}_{f\circ T} = \operatorname{Id} + \frac{1}{\kappa} T^* \circ (\operatorname{prox}_{\kappa f} - \operatorname{Id}) \circ T. \tag{5.17}$$

Numerous additional properties of proximity operators are mentioned in Combettes and Pesquet [2011]; Parikh and Boyd [2013].

In this chapter, we will be mainly concerned with the determination of the proximity of D as introduced in Section 5.2.2 (in this case, $\mathcal{H} = \mathbb{R}^P \times \mathbb{R}^P$). The next result shows that the problem reduces to the determination of the proximity operator of a real function of two variables.

Proposition 5.2.4 *Let D be defined by (5.8) where $\Phi \in \Gamma_0(\mathbb{R}^2)$ and let $\gamma \in]0, +\infty[$. Let $u = (u^{(i)})_{1\leq i\leq P} \in \mathbb{R}^P$ and $v = (v^{(i)})_{1\leq i\leq P} \in \mathbb{R}^P$. Then, for every $\overline{p} = (\overline{p}^{(i)})_{1\leq i\leq P} \in \mathbb{R}^P$ and for every $\overline{q} = (\overline{q}^{(i)})_{1\leq i\leq P} \in \mathbb{R}^P$,*

$$\operatorname{prox}_{\gamma D(\cdot+u,\cdot+v)}(\overline{p},\overline{q}) = \big((p^{(i)} - u^{(i)})_{1\leq i\leq P}, (q^{(i)} - v^{(i)})_{1\leq i\leq P}\big) \tag{5.18}$$

where

$$(\forall i \in \{1,\dots,P\}) \qquad (p^{(i)}, q^{(i)}) = \operatorname{prox}_{\gamma\Phi}(\overline{p}^{(i)} + u^{(i)}, \overline{q}^{(i)} + v^{(i)}). \tag{5.19}$$

Proof. The result is a straightforward consequence of Combettes and Pesquet [2011] and Proposition 5.2.3(iii) (by setting $f = D$ and $z = (u, v)$). □

Note that, although an extensive list of proximity operators of one-variable real functions can be found in Combettes and Pesquet [2011], few results are available for real functions of two variables Benamou and Brenier [2000]; Benamou et al. [2002]; Combettes and Pesquet [2008]; Chierchia et al. [2013]. An example of such a result is provided below.

Proposition 5.2.5 *Let $\varphi \in \Gamma_0(\mathbb{R})$ be an even differentiable function on $\mathbb{R}\setminus\{0\}$. Let $\Phi\colon \mathbb{R}^2 \to]-\infty, +\infty] : (\nu,\xi) \mapsto \varphi(\nu-\xi) + \iota_{[0,+\infty[}(\nu) + \iota_{[0,+\infty[}(\xi)$. Then, for every $(\overline{\nu},\overline{\xi}) \in \mathbb{R}^2$,*
$\operatorname{prox}_\Phi(\overline{\nu},\overline{\xi}) =$

$$\begin{cases} \frac{1}{2}\big(\overline{\nu}+\overline{\xi}+\operatorname{prox}_{2\varphi}(\overline{\nu}-\overline{\xi}), \overline{\nu}+\overline{\xi}-\operatorname{prox}_{2\varphi}(\overline{\nu}-\overline{\xi})\big) & \text{if } \operatorname{prox}_{2\varphi}(\overline{\nu}-\overline{\xi}) < |\overline{\nu}+\overline{\xi}| \\ (0, \operatorname{prox}_\varphi \overline{\xi}) & \text{if } \operatorname{prox}_\varphi \overline{\xi} > 0 \\ & \text{and } \operatorname{prox}_\varphi \overline{\xi} \geq \overline{\nu}+\overline{\xi} \\ (\operatorname{prox}_\varphi \overline{\nu}, 0) & \text{if } \operatorname{prox}_\varphi \overline{\nu} > 0 \\ & \text{and } \operatorname{prox}_\varphi \overline{\nu} \geq \overline{\nu}+\overline{\xi} \\ (0,0) & \text{otherwise.} \end{cases} \tag{5.20}$$

Proof. Let $(\overline{\nu},\overline{\xi}) \in \mathbb{R}^2$. From Proposition 5.2.3(i), we know that $\operatorname{prox}_\Phi(\overline{\nu},\overline{\xi}) \in [0,+\infty[^2$. By using Proposition 5.2.3(ii), we have the following equivalences:

$$\begin{cases} (\nu,\xi) \in]0,+\infty[^2 \\ (\nu,\xi) = \operatorname{prox}_\Phi(\overline{\nu},\overline{\xi}) \end{cases} \Leftrightarrow \begin{cases} (\nu,\xi) \in]0,+\infty[^2 \\ \overline{\nu}-\nu \in \partial\varphi(\nu-\xi) \\ \overline{\xi}-\xi \in -\partial\varphi(\nu-\xi) \end{cases}$$
$$\Leftrightarrow \begin{cases} (\nu,\xi) \in]0,+\infty[^2 \\ (\nu,\xi) = \operatorname{prox}_{\tilde{\Phi}}(\overline{\nu},\overline{\xi}) \end{cases} \tag{5.21}$$

where $\tilde{\Phi}\colon (\overline{\nu},\overline{\xi}) \mapsto \varphi(\overline{\nu}-\overline{\xi})$. By using now Proposition 5.2.3(vi), we get

$$
\begin{aligned}
(5.21) &\Leftrightarrow \begin{cases} (\nu,\xi) \in]0,+\infty[^2 \\ (\nu,\xi) = (\overline{\nu},\overline{\xi}) + \frac{1}{2}\big(\mathrm{prox}_{2\varphi}(\overline{\nu}-\overline{\xi}) - \overline{\nu} + \overline{\xi}\big)(1,-1) \end{cases} \\
&\Leftrightarrow \begin{cases} (\nu,\xi) \in]0,+\infty[^2 \\ \nu = \frac{1}{2}\big(\overline{\nu}+\overline{\xi} + \mathrm{prox}_{2\varphi}(\overline{\nu}-\overline{\xi})\big) \\ \xi = \frac{1}{2}\big(\overline{\nu}+\overline{\xi} - \mathrm{prox}_{2\varphi}(\overline{\nu}-\overline{\xi})\big) \end{cases} \\
&\Leftrightarrow \begin{cases} \mathrm{prox}_{2\varphi}(\overline{\nu}-\overline{\xi}) < |\overline{\nu}+\overline{\xi}| \\ \nu = \frac{1}{2}\big(\overline{\nu}+\overline{\xi} + \mathrm{prox}_{2\varphi}(\overline{\nu}-\overline{\xi})\big) \\ \xi = \frac{1}{2}\big(\overline{\nu}+\overline{\xi} - \mathrm{prox}_{2\varphi}(\overline{\nu}-\overline{\xi})\big). \end{cases} \qquad (5.22)
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\begin{cases} \nu = 0, \xi \in]0,+\infty[\\ (\nu,\xi) = \mathrm{prox}_{\Phi}(\overline{\nu},\overline{\xi}) \end{cases} &\Leftrightarrow \begin{cases} \nu = 0, \xi \in]0,+\infty[\\ \overline{\nu} - \varphi'(-\xi) \in \partial \iota_{[0,+\infty[}(0) =]-\infty, 0] \\ \xi - \overline{\xi} - \varphi'(-\xi) = 0 \end{cases} \\
&\Leftrightarrow \begin{cases} \nu = 0, \xi \in]0,+\infty[\\ \xi \geq \overline{\nu} + \overline{\xi} \\ \xi = \mathrm{prox}_{\varphi(-\cdot)}\overline{\xi} \end{cases} \\
&\Leftrightarrow \begin{cases} \mathrm{prox}_{\varphi(-\cdot)}\overline{\xi} \in]0,+\infty[\cap [\overline{\nu}+\overline{\xi}, +\infty[\\ \nu = 0 \\ \xi = \mathrm{prox}_{\varphi(-\cdot)}\overline{\xi} \end{cases} \\
&\Leftrightarrow \begin{cases} \mathrm{prox}_{\varphi}\overline{\xi} \in]0,+\infty[\cap [\overline{\nu}+\overline{\xi}, +\infty[\\ \nu = 0 \\ \xi = \mathrm{prox}_{\varphi}\overline{\xi}. \end{cases} \qquad (5.23)
\end{aligned}
$$

where the last equivalence results from the assumption that φ is even. Symmetrically,

$$
\begin{cases} \nu \in]0,+\infty[\,, \xi = 0 \\ (\nu,\xi) = \mathrm{prox}_{\Phi}(\overline{\nu},\overline{\xi}) \end{cases} \Leftrightarrow \begin{cases} \mathrm{prox}_{\varphi}\overline{\nu} \in]0,+\infty[\cap [\overline{\nu}+\overline{\xi}, +\infty[\\ \nu = \mathrm{prox}_{\varphi}\overline{\nu} \\ \xi = 0. \end{cases} \qquad (5.24)
$$

□

The above two propositions provide a simple characterization of the proximity operators of some distances defined for nonnegative-valued vectors. However, the assumptions made in Proposition 5.2.5 are not satisfied by the class of functions Φ considered in Section 5.2.2.[2] In the next section, we will propose two algorithms for solving a general class of convex problems involving these functions Φ.

[2]Indeed, none of the considered φ-divergences can be expressed as a function of the difference between the two arguments.

5.2.4 Proximal splitting algorithms

As soon as we know how to calculate the proximity operators of the functions involved in Problem 5.2.1, various proximal methods can be employed to solve it numerically. Two examples of such methods are given subsequently.

The first algorithm is PPXA+ Pesquet and Pustelnik [2012] which constitutes an extension of PPXA (Parallel ProXimal Agorithm) proposed in Combettes and Pesquet [2008]. As can be seen in Afonso et al. [2011]; Setzer et al. [2010], PPXA+ is also strongly related to augmented Lagrangian methods [Pesquet and Pustelnik, 2012, Section 6]. In the form provided below, this algorithm requires $A^\top A + B^\top B + \sum_{s=1}^{S} T_s^\top T_s$ to be invertible.[3]

Algorithm 6 PPXA+

`Initialization`

$(\omega_0, \ldots, \omega_S) \in]0, +\infty[^{S+1}$, $t_{0,0} \in \mathbb{R}^P, t_{1,0} \in \mathbb{R}^P, t_{2,0} \in \mathbb{R}^{K_1}, \ldots, t_{S+1,0} \in \mathbb{R}^{K_S}$

$x_0 =$

$(\omega_0 A^\top A + \omega_0 B^\top B + \sum_{s=1}^{S} \omega_s T_s^\top T_s)^{-1}(\omega_0 A^\top t_{0,0} + \omega_0 B^\top t_{1,0} + \sum_{s=1}^{S} \omega_s T_s^\top t_{s+1,0})$

`for` $n = 0, 1, \ldots$

$(r_{0,n}, r_{1,n}) = \mathrm{prox}_{\omega_0^{-1} D(\cdot+u,\cdot+v)}(t_{0,n}, t_{1,n}) + e_{0,n}$

`for` $s = 1, \ldots, S$

$r_{s+1,n} = \mathrm{prox}_{\omega_s^{-1} R_s} t_{s+1,n} + e_{s,n}$

$y_n =$

$(\omega_0 A^\top A + \omega_0 B^\top B + \sum_{s=1}^{S} \omega_s T_s^\top T_s)^{-1}(\omega_0 A^\top r_{0,n} + \omega_0 B^\top r_{1,n} + \sum_{s=1}^{S} \omega_s T_s^\top r_{s+1,n})$

$\lambda_n \in]0, 2[$

$t_{0,n+1} = t_{0,n} + \lambda_n\big(A(2y_n - x_n) - r_{0,n}\big)$

$t_{1,n+1} = t_{1,n} + \lambda_n\big(B(2y_n - x_n) - r_{1,n}\big)$

`for` $s = 1, \ldots, S$

$t_{s+1,n+1} = t_{s+1,n} + \lambda_n\big(T_s(2y_n - x_n) - r_{s+1,n}\big)$

$x_{n+1} = x_n + \lambda_n(y_n - x_n)$.

In this algorithm, $\omega_0, \ldots, \omega_S$ are weighting factors and $(\lambda_n)_{n\geq 0}$ relaxation factors. Typical values for these parameters will be indicated later. For every $n \geq 0$, the variables $e_{0,n} \in (\mathbb{R}^P)^2$, $e_{1,n} \in \mathbb{R}^{K_1}, \ldots, e_{S,n} \in \mathbb{R}^{K_S}$ model possible errors in the computation of the proximity operators. For instance, these errors arise when the proximity operator is not available in a closed form

[3] A slightly more complex variant of PPXA+ exists where this assumption is not required Attouch and Soueycatt [2009]; Pesquet and Pustelnik [2012], Prop. 5.4.

and one needs to compute it through inner iterations. Under some technical conditions, the convergence of PPXA+ is guaranteed.

Proposition 5.2.6 [Pesquet and Pustelnik, 2012, Corollary 5.3] *Suppose that the following assumptions hold.*

(i) $A^\top A + B^\top B + \sum_{s=1}^{S} T_s^\top T_s$ *is invertible.*

(ii) *There exists* $\check{x} \in \mathbb{R}^N$ *such that*

$$\begin{aligned} &A\check{x} + u \in \left]0, +\infty\right[^P, \quad B\check{x} + v \in \left]0, +\infty\right[^P \\ &(\forall s \in \{1, \ldots, S\}) \qquad T_s\check{x} \in \mathrm{ri}(\mathrm{dom}\, R_s). \end{aligned} \tag{5.25}$$

(iii) *There exists* $\underline{\lambda} \in]0, 2[$ *such that* $(\forall n \in \mathbb{N})\ \underline{\lambda} \leq \lambda_{n+1} \leq \lambda_n < 2$.

(iv) $(\forall s \in \{0, \ldots, S\})\ \sum_{n\in\mathbb{N}} \|e_{s,n}\| < +\infty$.

If the set of solutions to Problem (5.2.1) *is nonempty, then any sequence* $(x_n)_{n\in\mathbb{N}}$ *generated by Algorithm* (6) *converges to an element of this set.*

It can be noticed that, at each iteration n, PPXA+ requires to solve a linear system in order to compute the intermediate variable y_n. The computational cost of this operation may be high when N is large. Proximal primal-dual approaches Chen and Teboulle [1994]; Esser et al. [2010]; Chambolle and Pock [2011]; Briceño-Arias and Combettes [2011]; Combettes and Pesquet [2012]; Vũ [2013]; Condat [2013] allow us to circumvent this difficulty. An example of such an approach is the M+LFBF (Monotone+Lipschitz Forward Backward Forward) method Combettes and Pesquet [2012] which takes the following form:

Algorithm 7 M+LFBF

Initialization

$t_{0,0} \in \mathbb{R}^P, t_{1,0} \in \mathbb{R}^P, t_{2,0} \in \mathbb{R}^{K_1}, \ldots, t_{S+1,0} \in \mathbb{R}^{K_S}$

$x_0 \in \mathbb{R}^N,\ \beta = \left(\|A\|^2 + \|B\|^2 + \sum_{s=1}^S \|T_s\|^2\right)^{1/2},\ \varepsilon \in \left]0, 1/(\beta+1)\right[$

for $n = 0, 1, \ldots$

$\gamma_n \in [\varepsilon, (1-\varepsilon)/\beta]$

$\widehat{x}_n = x_n - \gamma_n(A^\top t_{0,n} + B^\top t_{1,n} + \sum_{s=1}^S T_s^\top t_{s+1,n})$

$\widehat{t}_{0,n} = t_{0,n} + \gamma_n A x_n,\ \widehat{t}_{1,n} = t_{1,n} + \gamma_n B x_n$

$(r_{0,n}, r_{1,n}) = (\widehat{t}_{0,n}, \widehat{t}_{1,n}) - \gamma_n \mathrm{prox}_{\gamma_n^{-1} D(\cdot+u,\cdot+v)}(\gamma_n^{-1}\widehat{t}_{0,n}, \gamma_n^{-1}\widehat{t}_{1,n}) + e_{0,n}$

$\widetilde{t}_{0,n} = r_{0,n} + \gamma_n A \widehat{x}_n,\ \widetilde{t}_{1,n} = r_{1,n} + \gamma_n B \widehat{x}_n$

$t_{0,n+1} = t_{0,n} - \widehat{t}_{0,n} + \widetilde{t}_{0,n},\ t_{1,n+1} = t_{1,n} - \widehat{t}_{1,n} + \widetilde{t}_{1,n}$

for $s = 1, \ldots, S$

$\widehat{t}_{s+1,n} = t_{s+1,n} + \gamma_n T_s x_n$

$r_{s+1,n} = \widehat{t}_{s+1,n} - \gamma_n \mathrm{prox}_{\gamma_n^{-1} R_s}(\gamma_n^{-1}\widehat{t}_{s+1,n}) + e_{s,n}$

$\widetilde{t}_{s+1,n} = r_{s+1,n} + \gamma_n T_s \widehat{x}_n$

$t_{s+1,n+1} = t_{s+1,n} - \widehat{t}_{s+1,n} + \widetilde{t}_{s+1,n}$

$\widetilde{x}_n = \widehat{x}_n - \gamma_n(A^\top r_{0,n} + B^\top r_{1,n} + \sum_{s=1}^S T_s^\top r_{s+1,n})$

$x_{n+1} = x_n - \widehat{x}_n + \widetilde{x}_n.$

In this algorithm, $(\gamma_n)_{n\geq 0}$ is a sequence of step-sizes, and $e_{0,n} \in (\mathbb{R}^P)^2$, $e_{1,n} \in \mathbb{R}^{K_1}, \ldots, e_{S,n} \in \mathbb{R}^{K_S}$ still correspond to possible errors in the computation of the proximity operators. The convergence of the algorithm is secured by the following result.

Proposition 5.2.7 [Combettes and Pesquet, 2012, Theorem 4.2] *Suppose that the following assumptions hold.*

(i) *There exists* $\check{x} \in \mathbb{R}^N$ *such that*

$$\begin{aligned} &A\check{x} + u \in \left]0, +\infty\right[^P, \quad B\check{x} + v \in \left]0, +\infty\right[^P \\ &(\forall s \in \{1, \ldots, S\}) \qquad T_s\check{x} \in \mathrm{ri}(\mathrm{dom}\, R_s). \end{aligned} \tag{5.26}$$

(ii) $(\forall s \in \{0, \ldots, S\})\ \sum_{n\in\mathbb{N}} \|e_{s,n}\| < +\infty$.

If the set of solutions to Problem (5.2.1) *is nonempty, then any sequence* $(x_n)_{n\in\mathbb{N}}$ *generated by Algorithm* (10) *converges to an element of this set.*

It is worth highlighting that these two algorithms share two interesting features: at first, many operations (e.g. the loops on s) can be implemented in

parallel. Second, they are tolerant with respect to errors in the computations of the proximity operators. [4]

§ 5.3 Main result

As shown by Proposition 5.2.4, we need to compute the proximity operator of a scaled version of a function $\Phi \in \Gamma_0(\mathbb{R}^2)$ as defined in (5.9). In the following, Θ denotes a primitive on $]0,+\infty[$ of the function $\zeta \mapsto \zeta\varphi'(\zeta^{-1})$. The following functions will subsequently play an important role:

$$\vartheta_- \colon]0,+\infty[\to \mathbb{R} \colon \zeta \mapsto \varphi'(\zeta^{-1}) \tag{5.27}$$

$$\vartheta_+ \colon]0,+\infty[\to \mathbb{R} \colon \zeta \mapsto \varphi(\zeta^{-1}) - \zeta^{-1}\varphi'(\zeta^{-1}). \tag{5.28}$$

A first technical result is as follows:

Lemma 5.3.1 *Let $\gamma \in]0,+\infty[$, let $(\overline{v},\overline{\xi}) \in \mathbb{R}^2$, and define*

$$\chi_- = \inf\big\{\zeta \in]0,+\infty[\;\big|\; \vartheta_-(\zeta) < \gamma^{-1}\overline{v}\big\} \tag{5.29}$$

$$\chi_+ = \sup\big\{\zeta \in]0,+\infty[\;\big|\; \vartheta_+(\zeta) < \gamma^{-1}\overline{\xi}\big\} \tag{5.30}$$

(with the usual convention $\inf \varnothing = +\infty$ and $\sup \varnothing = -\infty$). If $\chi_- \neq +\infty$, the function

$$\psi \colon]0,+\infty[\to \mathbb{R} \colon \zeta \mapsto \zeta\varphi(\zeta^{-1}) - \Theta(\zeta) + \frac{\gamma^{-1}\overline{v}}{2}\zeta^2 - \gamma^{-1}\overline{\xi}\zeta \tag{5.31}$$

is strictly convex on $]\chi_-,+\infty[$. In addition, if

(i) $\chi_- \neq +\infty$ *and* $\chi_+ \neq -\infty$

(ii) $\lim\limits_{\substack{\zeta\to\chi_-\\ \zeta>\chi_-}} \psi'(\zeta) < 0$

(iii) $\lim_{\zeta\to\chi_+} \psi'(\zeta) > 0$

then ψ admits a unique minimizer $\widehat{\zeta}$ on $]\chi_-,+\infty[$, and $\widehat{\zeta} < \chi_+$.

Proof. The derivative of ψ is

$$\begin{aligned}(\forall \zeta \in]0,+\infty[) \qquad \psi'(\zeta) &= \varphi(\zeta^{-1}) - (\zeta+\zeta^{-1})\varphi'(\zeta^{-1}) + \gamma^{-1}\overline{v}\zeta - \gamma^{-1}\overline{\xi}\\ &= \zeta\big(\gamma^{-1}\overline{v} - \vartheta_-(\zeta)\big) + \vartheta_+(\zeta) - \gamma^{-1}\overline{\xi}.\end{aligned} \tag{5.32}$$

The function ϑ_- is decreasing as the convexity of φ yields

$$(\forall \zeta \in]0,+\infty[) \qquad \vartheta_-'(\zeta) = -\zeta^{-2}\varphi''(\zeta^{-1}) \leq 0. \tag{5.33}$$

[4] Note also that we have employed PPXA+ (resp. M+LFBF) in Chapter 3 (resp. Chapter 4), where the optimization was performed with respect to one variable.

This allows us to deduce that, if $\left\{\zeta \in]0,+\infty[\;\middle|\; \vartheta_-(\zeta) < \gamma^{-1}\overline{v}\right\} \neq \varnothing$,

$$]\chi_-,+\infty[= \left\{\zeta \in]0,+\infty[\;\middle|\; \vartheta_-(\zeta) < \gamma^{-1}\overline{v}\right\}. \tag{5.34}$$

Similarly, the function ϑ_+ is increasing as the convexity of φ yields

$$(\forall \zeta \in]0,+\infty[) \qquad \vartheta'_+(\zeta) = \zeta^{-3}\varphi''(\zeta^{-1}) \geq 0 \tag{5.35}$$

which allows us to deduce that, if $\left\{\zeta \in]0,+\infty[\;\middle|\; \vartheta_+(\zeta) < \gamma^{-1}\overline{\xi}\right\} \neq \varnothing$,

$$]0,\chi_+[= \left\{\zeta \in]0,+\infty[\;\middle|\; \vartheta_+(\zeta) < \gamma^{-1}\overline{\xi}\right\}. \tag{5.36}$$

If $(\chi_-,\chi_+) \in]0,+\infty[^2$, then (5.32) leads to

$$\psi'(\chi_-) = \vartheta_+(\chi_-) - \gamma^{-1}\overline{\xi} \tag{5.37}$$

$$\psi'(\chi_+) = \chi_+\left(\gamma^{-1}\overline{v} - \vartheta_-(\chi_+)\right). \tag{5.38}$$

So, Conditions (ii) and (iii) are equivalent to

$$\vartheta_+(\chi_-) - \gamma^{-1}\overline{\xi} < 0 \tag{5.39}$$

$$\chi_+\left(\gamma^{-1}\overline{v} - \vartheta_-(\chi_+)\right) > 0. \tag{5.40}$$

In view of (5.34) and (5.36), these inequalities are satisfied if and only if $\chi_- < \chi_+$. The inequality is also obviously satisfied if $\chi_- = 0$ or $\chi_+ = +\infty$.

In addition, we have

$$(\forall \zeta \in]0,+\infty[) \qquad \psi''(\zeta) = \gamma^{-1}\overline{v} - \vartheta_-(\zeta) + \zeta^{-1}(1+\zeta^{-2})\varphi''(\zeta^{-1}). \tag{5.41}$$

When $\zeta > \chi_- \neq +\infty$, $\gamma^{-1}\overline{v} - \vartheta_-(\zeta) > 0$, and the convexity of φ yields $\psi''(\zeta) > 0$. This shows that ψ is strictly convex on $]\chi_-,+\infty[$.

If Conditions (i)-(iii) are satisfied, due to the continuity of ψ', there exists $\widehat{\zeta} \in]\chi_-,\chi_+[$ such that $\psi'(\widehat{\zeta}) = 0$. Because of the strict convexity of ψ on $]\chi_-,+\infty[$, $\widehat{\zeta}$ is the unique minimizer of ψ on this interval. □

The required assumptions in the previous lemma can often be simplified as stated below.

Lemma 5.3.2 *Let $\gamma \in]0,+\infty[$ and $(\overline{v},\overline{\xi}) \in \mathbb{R}^2$. If $(\chi_-,\chi_+) \in]0,+\infty[^2$, then Conditions (ii) and (iii) in Lemma 5.3.1 are equivalent to: $\chi_- < \chi_+$. If $\chi_- \in]0,+\infty[$ and $\chi_+ = +\infty$ (resp. $\chi_- = 0$ and $\chi_+ \in]0,+\infty[$), Conditions (ii)-(iii) are satisfied if and only if $\lim_{\zeta\to+\infty}\psi'(\zeta) > 0$ (resp. $\lim_{\substack{\zeta\to 0\\ \zeta>0}}\psi'(\zeta) < 0$).*

Proof. If $(\chi_-, \chi_+) \in]0, +\infty[^2$, we have already shown that Conditions (ii) and (iii) are satisfied if and only $\chi_- < \chi_+$.

If $\chi_- \in]0, +\infty[$ and $\chi_+ = +\infty$ (resp. $\chi_- = 0$ and $\chi_+ \in]0, +\infty[$), we still have

$$\psi'(\chi_-) = \vartheta_+(\chi_-) - \gamma^{-1}\overline{\xi} < 0. \tag{5.42}$$

$$(\text{resp. } \psi'(\chi_+) = \chi_+ \left(\gamma^{-1}\overline{\upsilon} - \vartheta_-(\chi_+)\right) > 0) \tag{5.43}$$

which shows that Condition (ii) (resp. Condition (iii)) is always satisfied. □

By using the same expressions of χ_- and χ_+ as in the previous lemmas, we obtain the following characterization of the proximity operator of any scaled version of Φ:

Proposition 5.3.3 *Let $\gamma \in]0, +\infty[$ and $(\overline{\upsilon}, \overline{\xi}) \in \mathbb{R}^2$. $\operatorname{prox}_{\gamma\Phi}(\overline{\upsilon}, \overline{\xi}) \in]0, +\infty[^2$ if and only if Conditions (i)-(iii) in Lemma 5.3.1 are satisfied. When these conditions hold,*

$$\operatorname{prox}_{\gamma\Phi}(\overline{\upsilon}, \overline{\xi}) = \left(\overline{\upsilon} - \gamma\,\vartheta_-(\widehat{\zeta}), \overline{\xi} - \gamma\,\vartheta_+(\widehat{\zeta})\right) \tag{5.44}$$

where $\widehat{\zeta} < \chi_+$ is the unique minimizer of ψ on $]\chi_-, +\infty[$.

Proof. For every $(\overline{\upsilon}, \overline{\xi}) \in \mathbb{R}^2$, such that Conditions (i)-(iii) in Lemma 5.3.1 hold, let

$$\upsilon = \overline{\upsilon} - \gamma\,\vartheta_-(\widehat{\zeta}) \tag{5.45}$$

$$\xi = \overline{\xi} - \gamma\,\vartheta_+(\widehat{\zeta}) \tag{5.46}$$

where the existence of $\widehat{\zeta} \in]\chi_-, \chi_+[$ is guaranteed by Lemma 5.3.1. As consequences of (5.34) and (5.36), υ and ξ are positive. In addition, since

$$\begin{aligned} & \psi'(\widehat{\zeta}) = 0 \\ \Leftrightarrow \quad & \widehat{\zeta}\left(\gamma^{-1}\overline{\upsilon} - \vartheta_-(\widehat{\zeta})\right) = \gamma^{-1}\overline{\xi} - \vartheta_+(\widehat{\zeta}) \end{aligned} \tag{5.47}$$

we derive from (5.45) and (5.46) that $\widehat{\zeta} = \xi/\upsilon > 0$. This allows us to re-express (5.45) and (5.46) as

$$\upsilon - \overline{\upsilon} + \gamma\varphi'\left(\frac{\upsilon}{\xi}\right) = 0 \tag{5.48}$$

$$\xi - \overline{\xi} + \gamma\left(\varphi\left(\frac{\upsilon}{\xi}\right) - \frac{\upsilon}{\xi}\varphi'\left(\frac{\upsilon}{\xi}\right)\right) = 0 \tag{5.49}$$

that is

$$\upsilon - \overline{\upsilon} + \gamma\frac{\partial\Phi}{\partial\upsilon}(\upsilon, \xi) = 0 \tag{5.50}$$

$$\xi - \overline{\xi} + \gamma\frac{\partial\Phi}{\partial\xi}(\upsilon, \xi) = 0. \tag{5.51}$$

The latter equations are satisfied if and only if $(\upsilon, \xi) = \operatorname{prox}_{\gamma\Phi}(\overline{\upsilon}, \overline{\xi})$ Combettes and Pesquet [2011].

Conversely, for every $(\overline{\upsilon}, \overline{\xi}) \in \mathbb{R}^2$, let $(\upsilon, \xi) = \operatorname{prox}_{\gamma\Phi}(\overline{\upsilon}, \overline{\xi})$. If $(\upsilon, \xi) \in]0, +\infty[^2$, (υ, ξ) satisfies (5.48) and (5.49). By setting $\widetilde{\zeta} = \xi/\upsilon > 0$, after simple calculations, we find

$$\upsilon = \overline{\upsilon} - \gamma\, \vartheta_-(\widetilde{\zeta}) > 0 \tag{5.52}$$

$$\xi = \overline{\xi} - \gamma\, \vartheta_+(\widetilde{\zeta}) > 0 \tag{5.53}$$

$$\psi'(\widetilde{\zeta}) = 0. \tag{5.54}$$

According to (5.34) and (5.36), (5.52) and (5.53) imply that $\chi_- \neq +\infty$, $\chi_+ \neq -\infty$, and $\widetilde{\zeta} \in]\chi_-, \chi_+[$. In addition, according to Lemma 5.3.1, ψ' is strictly increasing on $]\chi_-, +\infty[$ (since ψ is strictly convex on this interval). Hence, ψ' has a limit at χ_- (which may be equal to $-\infty$ when $\chi_- = -\infty$), and Condition (ii) is satisfied. Similarly, ψ' has a limit at χ_+ (possibly equal to $+\infty$ when $\chi_+ = +\infty$), and Condition (iii) is satisfied. □

Remark 5.3.4 In (5.9), a special case arises when

$$(\forall \zeta \in]0, +\infty[) \qquad \varphi(\zeta) = \widetilde{\varphi}(\zeta) + \zeta\widetilde{\varphi}(\zeta^{-1}) \tag{5.55}$$

where $\widetilde{\varphi}$ is a twice differentiable convex function on $]0, +\infty[$. Then Φ takes a symmetric form, so giving birth to $\mathcal{L}$-divergences. It can then be deduced from (5.28) that

$$(\forall \zeta \in]0, +\infty[) \qquad \vartheta_-(\zeta) = \vartheta_+(\zeta^{-1}) = \widetilde{\varphi}(\zeta) + \widetilde{\varphi}'(\zeta^{-1}) - \zeta\widetilde{\varphi}'(\zeta). \tag{5.56}$$

§ 5.4 Examples

5.4.1 Kullback-Leibler divergence

Let us now apply the results in the previous section to the function

$$\Phi\colon (\upsilon, \xi) \mapsto \begin{cases} \upsilon \ln\left(\frac{\upsilon}{\xi}\right) + \xi - \upsilon & \text{if } (\upsilon, \xi) \in]0, +\infty[^2 \\ \xi & \text{if } \upsilon = 0 \text{ and } \xi \in [0, +\infty[\\ +\infty & \text{otherwise.} \end{cases} \tag{5.57}$$

This is a function in $\Gamma_0(\mathbb{R}^2)$ satisfying (5.9) with

$$(\forall \zeta \in]0, +\infty[) \qquad \varphi(\zeta) = \zeta \ln \zeta - \zeta + 1. \tag{5.58}$$

Proposition 5.4.1 *The proximity operator of* $\gamma\Phi$ *with* $\gamma \in]0,+\infty[$ *is* $(\forall(\overline{v},\overline{\xi}) \in \mathbb{R}^2)$

$$\operatorname{prox}_{\gamma\Phi}(\overline{v},\overline{\xi}) = \begin{cases} \big(\overline{v}+\gamma\ln\widehat{\zeta},\overline{\xi}+\gamma(\widehat{\zeta}^{-1}-1)\big) & \text{if } \exp(\overline{v}/\gamma) > 1-\gamma^{-1}\overline{\xi} \\ (0,0) & \text{otherwise} \end{cases} \tag{5.59}$$

where, in the first case, $\widehat{\zeta}$ *is the unique minimizer on* $]\exp(-\overline{v}/\gamma),+\infty[$ *of the function:*

$$\psi\colon\ ]0,+\infty[\to \mathbb{R}\colon \zeta \mapsto \Big(\frac{\zeta^2}{2}-1\Big)\ln\zeta + \frac{1}{2}\Big(\gamma^{-1}\overline{v}-\frac{1}{2}\Big)\zeta^2 + (1-\gamma^{-1}\overline{\xi})\zeta. \tag{5.60}$$

Proof. For every $(\overline{v},\overline{\xi}) \in \mathbb{R}^2$, $(v,\xi) = \operatorname{prox}_{\gamma\Phi}(\overline{v},\overline{\xi})$ is such that $(v,\xi) \in \operatorname{dom}\Phi$ Moreau [1962]. Let us first note that

$$v \in]0,+\infty[\Leftrightarrow (v,\xi) \in]0,+\infty[^2\,. \tag{5.61}$$

We are now able to apply Proposition 5.3.3, where ψ is given by (5.60) and

$$(\forall\zeta \in]0,+\infty[) \qquad \Theta(\zeta) = \frac{\zeta^2}{2}\left(\frac{1}{2}-\ln\zeta\right)-1 \tag{5.62}$$

$$\vartheta_-(\zeta) = -\ln\zeta \tag{5.63}$$

$$\vartheta_+(\zeta) = 1-\zeta^{-1}. \tag{5.64}$$

In addition,

$$\chi_- = \exp(-\overline{v}/\gamma) \tag{5.65}$$

$$\chi_+ = \begin{cases} (1-\gamma^{-1}\overline{\xi})^{-1} & \text{if } \overline{\xi} < \gamma \\ +\infty & \text{otherwise.} \end{cases} \tag{5.66}$$

According to (5.61) and Proposition 5.3.3, $v \in]0,+\infty[$ if and only if Conditions (i)-(iii) in Lemma 5.3.1 hold. Since $\chi_- \in]0,+\infty[$ and $\lim_{\zeta\to+\infty}\psi'(\zeta) = +\infty$, Lemma 5.3.2 shows that these conditions are satisfied if and only if

$$\overline{\xi} < \gamma \quad \text{and} \quad \exp(-\overline{v}/\gamma) < (1-\gamma^{-1}\overline{\xi})^{-1} \tag{5.67}$$

$$\text{or} \quad \overline{\xi} \geq \gamma \tag{5.68}$$

or, equivalently

$$\exp(\overline{v}/\gamma) > 1-\gamma^{-1}\overline{\xi}. \tag{5.69}$$

Under this assumption, Proposition 5.3.3 leads to the following expressions of the proximity operator:

$$v = \overline{v}+\gamma\ln\widehat{\zeta} \tag{5.70}$$

$$\xi = \overline{\xi}+\gamma\left(\frac{1}{\widehat{\zeta}}-1\right) \tag{5.71}$$

where $\widehat{\zeta}$ is the unique minimizer on $]\exp(-\overline{\upsilon}/\gamma), +\infty[$ of the function ψ.

We have shown that $\upsilon > 0 \Leftrightarrow$ (5.69). So, $\upsilon = 0$ when (5.69) is not satisfied. Then, the expression of ξ simply reduces to the asymmetric soft-thresholding rule Combettes and Pesquet [2007a]:

$$\xi = \begin{cases} \overline{\xi} - \gamma & \text{if } \overline{\xi} > \gamma \\ 0 & \text{otherwise.} \end{cases} \tag{5.72}$$

However, $\exp(\overline{\upsilon}/\gamma) \leq 1 - \gamma^{-1}\overline{\xi} \Rightarrow \overline{\xi} < \gamma$, so that ξ is necessarily equal to 0. □

Remark 5.4.2

(i) It can be noticed that

$$\begin{aligned} & \psi'(\widehat{\zeta}) = \widehat{\zeta} \ln \widehat{\zeta} + \gamma^{-1}\overline{\upsilon}\widehat{\zeta} - \frac{1}{\widehat{\zeta}} + 1 - \gamma^{-1}\overline{\xi} = 0 \\ \Leftrightarrow \quad & \widehat{\zeta}^{-1} \exp\left(\widehat{\zeta}^{-1}(\widehat{\zeta}^{-1} + \gamma^{-1}\overline{\xi} - 1)\right) = \exp(\gamma^{-1}\overline{\upsilon}). \end{aligned} \tag{5.73}$$

In the case where $\overline{\xi} = \gamma$, the above equation reduces to

$$\begin{aligned} & 2\widehat{\zeta}^{-2} \exp\left(2\widehat{\zeta}^{-2}\right) = 2\exp(2\gamma^{-1}\overline{\upsilon}) \\ \Leftrightarrow \quad & \widehat{\zeta} = \left(\frac{2}{W(2e^{2\overline{\upsilon}/\gamma})}\right)^{1/2} \end{aligned} \tag{5.74}$$

where W is the Lambert W function Corless et al. [1996]. When $\overline{\xi} \neq \gamma$, although a closed form expression of (5.73) is not available, efficient numerical methods to compute $\widehat{\zeta}$ can be developed as shown in the following.

(ii) For a given $(\overline{\upsilon}, \overline{\xi}) \in \mathbb{R}^2$, minimizing ψ over $]\exp(-\overline{\upsilon}/\gamma), +\infty[$ amounts to minimizing $\psi_1 + \psi_2$ over $\mathbb{R}$, where

$$(\forall \zeta \in \mathbb{R}) \qquad \psi_1(\zeta) = \begin{cases} \frac{\zeta^2}{2}\ln\zeta + \frac{1}{2}\left(\gamma^{-1}\overline{\upsilon} - \frac{1}{2}\right)\zeta^2 & \text{if } \zeta \geq \exp(-\overline{\upsilon}/\gamma) \\ +\infty & \text{otherwise} \end{cases} \tag{5.75}$$

and

$$(\forall \zeta \in \mathbb{R}) \qquad \psi_2(\zeta) = \begin{cases} -\ln\zeta + (1 - \gamma^{-1}\overline{\xi})\zeta & \text{if } \zeta > 0 \\ +\infty & \text{otherwise.} \end{cases} \tag{5.76}$$

We have then

$$(\forall \zeta > \exp(-\overline{\upsilon}/\gamma)) \qquad \psi_1''(\zeta) = \ln\zeta + \gamma^{-1}\overline{\upsilon} + 1 > 0 \tag{5.77}$$

and

$$(\forall \zeta > 0) \qquad \psi_2''(\zeta) = \frac{1}{\zeta^2} > 0 \tag{5.78}$$

which shows that ψ_1 and ψ_2 are strictly convex functions.

As ψ_1 and ψ_2 belong to $\Gamma_0(\mathbb{R})$, a possibility for performing this minimization is to employ the Douglas-Rachford algorithm. This requires calculating the proximity operators of $\tau\psi_1$ and $\tau\psi_2$ where $\tau > 0$.
For every $\overline{\zeta} \in \mathbb{R}$, $\operatorname{prox}_{\tau\psi_1}\overline{\zeta} \in [\exp(-\overline{v}/\gamma), +\infty[$ Moreau [1962]. We have then

$$\begin{cases} \zeta > \exp(-\overline{v}/\gamma) \\ \zeta = \operatorname{prox}_{\tau\psi_1}\overline{\zeta} \end{cases} \tag{5.79}$$

$$\Leftrightarrow \quad \begin{cases} \zeta > \exp(-\overline{v}/\gamma) \\ \zeta - \overline{\zeta} + \tau\Big(\zeta \ln \zeta + \gamma^{-1}\overline{v}\zeta\Big) = 0. \end{cases} \tag{5.80}$$

Since $\zeta \mapsto \zeta + \tau(\zeta \ln \zeta + \gamma^{-1}\overline{v}\zeta)$ is strictly increasing (hence, bijective) from $[\exp(-\overline{v}/\gamma), +\infty[$ to $[\exp(-\overline{v}/\gamma), +\infty[$, this is also equivalent to

$$\begin{cases} \overline{\zeta} > \exp(-\overline{v}/\gamma) \\ \zeta - \overline{\zeta} + \tau(\zeta \ln \zeta + \gamma^{-1}\overline{v}\zeta) = 0. \end{cases} \tag{5.81}$$

In addition, the latter equality can be rewritten as

$$\begin{aligned} & \frac{\overline{\zeta}}{\tau\zeta} \exp\left(\frac{\overline{\zeta}}{\tau\zeta}\right) = \frac{\overline{\zeta}}{\tau} \exp(\tau^{-1} + \gamma^{-1}\overline{v}) \\ \Leftrightarrow \quad & \zeta = \frac{\overline{\zeta}}{\tau W\Big(\tau^{-1}\overline{\zeta} e^{\tau^{-1}+\gamma^{-1}\overline{v}}\Big)}. \end{aligned} \tag{5.82}$$

In summary, (5.79) is equivalent to $\overline{\zeta} > \exp(-\overline{v}/\gamma)$ and (5.82). We deduce that

$$\operatorname{prox}_{\tau\psi_1}\overline{\zeta} = \begin{cases} \dfrac{\overline{\zeta}}{\tau W\Big(\tau^{-1}\overline{\zeta} e^{\tau^{-1}+\gamma^{-1}\overline{v}}\Big)} & \text{if } \overline{\zeta} > \exp(-\overline{v}/\gamma) \\ \exp(-\overline{v}/\gamma) & \text{otherwise.} \end{cases} \tag{5.83}$$

The proximity operator of $\tau\psi_2$ follows from standard results on the calculation of proximity operators Pustelnik et al. [2011]:

$$\operatorname{prox}_{\tau\psi_2}\overline{\zeta} = \frac{1}{2}\Big(\overline{\zeta} + \tau(\gamma^{-1}\overline{\xi} - 1) + \sqrt{|\overline{\zeta} + \tau(\gamma^{-1}\overline{\xi} - 1)|^2 + 4\tau}\Big). \tag{5.84}$$

The Douglas-Rachford algorithm to compute the solution is then provided by:

Algorithm 8 Douglas-Rachford algorithm

```
Fix  ε ∈ ]0,1[,  τ > 0,  ζ̂_{-1/2} ∈ ]0,+∞[
for  n = 0, 1, ...
    ζ̂_n = prox_{τψ_2} ζ̂_{n-1/2}
    μ_n ∈ [ε, 2 − ε]
    ζ̂_{n+1/2} = ζ̂_{n-1/2} + μ_n (prox_{τψ_1}(2ζ̂_n − ζ̂_{n-1/2}) − ζ̂_n).
```

(iii) More generally, we can derive the proximity operator of

$$\widetilde{\Phi}\colon (\upsilon,\xi)\mapsto\begin{cases}\upsilon\ln\left(\frac{\upsilon}{\xi}\right)+\kappa(\xi-\upsilon) & \text{if } (\upsilon,\xi)\in\left]0,+\infty\right[^2\\ \kappa\xi & \text{if } \upsilon=0 \text{ and } \xi\in[0,+\infty[\\ +\infty & \text{otherwise}\end{cases} \tag{5.85}$$

where $\kappa\in\mathbb{R}$. Of particular interest in the literature is the case when $\kappa=0$ Blahut [1972]; Arimoto [1972]; Dupé et al. [2009]; Pustelnik et al. [2011]. From Proposition 5.2.3(iv), we get, for every $\gamma\in\left]0,+\infty\right[$,

$$\big(\forall(\overline{\upsilon},\overline{\xi})\in\mathbb{R}^2\big)\qquad \operatorname{prox}_{\gamma\widetilde{\Phi}}(\overline{\upsilon},\overline{\xi})=\operatorname{prox}_{\gamma\Phi}(\overline{\upsilon}+\gamma\kappa-\gamma,\overline{\xi}-\gamma\kappa+\gamma) \tag{5.86}$$

where $\operatorname{prox}_{\gamma\Phi}$ is provided by Proposition 5.4.1.

5.4.2 Jeffreys-Kullback divergence

Let us now consider the symmetrized form of (5.57) given by

$$\Phi\colon(\upsilon,\xi)\mapsto\begin{cases}(\upsilon-\xi)\big(\ln\upsilon-\ln\xi\big) & \text{if } (\upsilon,\xi)\in\left]0,+\infty\right[^2\\ 0 & \text{if } \upsilon=\xi=0\\ +\infty & \text{otherwise.}\end{cases} \tag{5.87}$$

This function belongs to $\Gamma_0(\mathbb{R}^2)$ and satisfies (5.9) and (5.55) with

$$(\forall\zeta\in\left]0,+\infty\right[)\qquad \widetilde{\varphi}(\zeta)=-\ln\zeta. \tag{5.88}$$

Proposition 5.4.3 *The proximity operator of $\gamma\Phi$ with $\gamma\in\left]0,+\infty\right[$ is* $(\forall(\overline{\upsilon},\overline{\xi})\in\mathbb{R}^2)$

$$\operatorname{prox}_{\gamma\Phi}(\overline{\upsilon},\overline{\xi})=\begin{cases}\big(\overline{\upsilon}+\gamma\big(\ln\widehat{\zeta}+\widehat{\zeta}-1\big),\overline{\xi}-\gamma\big(\ln\widehat{\zeta}-\widehat{\zeta}^{-1}+1\big)\big)\\ \qquad\qquad\qquad\qquad \textit{if } W(e^{1-\gamma^{-1}\overline{\upsilon}})W(e^{1-\gamma^{-1}\overline{\xi}})<1\\ (0,0)\qquad\qquad \textit{otherwise}\end{cases} \tag{5.89}$$

where, in the first case, $\widehat{\zeta}$ *is the unique minimizer on* $]W(e^{1-\gamma^{-1}\overline{v}}),+\infty[$ *of the function:*

$$\psi\colon\ ]0,+\infty[\to\mathbb{R}\colon\zeta\mapsto\Big(\frac{\zeta^2}{2}+\zeta-1\Big)\ln\zeta+\frac{\zeta^3}{3}+\frac{1}{2}\Big(\gamma^{-1}\overline{v}-\frac{3}{2}\Big)\zeta^2-\gamma^{-1}\overline{\xi}\zeta. \tag{5.90}$$

Proof. We apply Proposition 5.3.3 where ψ is given by (5.90) and

$$(\forall\zeta\in]0,+\infty[)\qquad \Theta(\zeta)=\zeta^2\Big(\frac{3}{4}-\frac{\zeta}{3}-\frac{1}{2}\ln\zeta\Big) \tag{5.91}$$

$$\vartheta_-(\zeta)=\vartheta_+(\zeta^{-1})=-\ln\zeta-\zeta+1. \tag{5.92}$$

The above equalities have been derived from (5.55) and (5.56). It can be deduced from (5.29), (5.30) and (5.92) that

$$\chi_-+\ln\chi_-=1-\gamma^{-1}\overline{v} \tag{5.93}$$

$$\chi_+^{-1}+\ln(\chi_+^{-1})=1-\gamma^{-1}\overline{\xi} \tag{5.94}$$

that is

$$\chi_-=W(e^{1-\gamma^{-1}\overline{v}}) \tag{5.95}$$

$$\chi_+=\big(W(e^{1-\gamma^{-1}\overline{\xi}})\big)^{-1}. \tag{5.96}$$

According to Proposition 5.3.3, $\operatorname{prox}_{\gamma\Phi}(\overline{v},\overline{\xi})\in]0,+\infty[^2$ if and only if Conditions (i)-(iii) in Lemma 5.3.1 hold. Lemma 5.3.2 shows that these conditions are satisfied if and only if

$$W(e^{1-\gamma^{-1}\overline{v}})W(e^{1-\gamma^{-1}\overline{\xi}})<1. \tag{5.97}$$

Under this assumption, the expression of the proximity operator follows from Proposition 5.3.3 and (5.92).
We have shown that $\operatorname{prox}_{\gamma\Phi}(\overline{v},\overline{\xi})\in]0,+\infty[^2\Leftrightarrow$ (5.97). Since $\operatorname{prox}_{\gamma\Phi}(\overline{v},\overline{\xi})\in\operatorname{dom}\Phi$, we necessarily get $\operatorname{prox}_{\gamma\Phi}(\overline{v},\overline{\xi})=(0,0)$, when (5.97) is not satisfied. □

Remark 5.4.4 To minimize ψ, we need to find the zero on $[\chi_-,\chi_+]$ of the function:

$$\psi'(\zeta)=(\zeta+1)\log\zeta+\frac{\zeta}{2}+1-\frac{1}{\zeta}+\zeta^2+\Big(\gamma^{-1}\overline{v}-\frac{3}{2}\Big)\zeta-\gamma^{-1}\overline{\xi}. \tag{5.98}$$

Thus can be performed by Algorithm 9, the convergence of which is proved in Section 5.7.

Algorithm 9 Newton algorithm

Fix $\widehat{\zeta}_0\in[\chi_-,\chi_+]$
for $n=0,1,\dots$
$\widehat{\zeta}_{n+1}=\mathsf{P}_{[\chi_-,\chi_+]}\big(\widehat{\zeta}_n-\psi'(\widehat{\zeta}_n)/\psi''(\widehat{\zeta}_n)\big)$.

5.4.3 Hellinger divergence

Let us now consider the function of $\Gamma_0(\mathbb{R}^2)$ given by

$$\Phi\colon (\upsilon,\xi)\mapsto\begin{cases}(\sqrt{\upsilon}-\sqrt{\xi})^2 & \text{if } (\upsilon,\xi)\in[0,+\infty[^2\\ +\infty & \text{otherwise.}\end{cases}\tag{5.99}$$

This symmetric function satisfies (5.9) and (5.55) with

$$(\forall\zeta\in\,]0,+\infty[)\qquad \widetilde{\varphi}(\zeta)=\zeta-\sqrt{\zeta}.\tag{5.100}$$

Proposition 5.4.5 *The proximity operator of $\gamma\Phi$ with $\gamma\in\,]0,+\infty[$ is*

$(\forall(\overline{\upsilon},\overline{\xi})\in\mathbb{R}^2)$

$$\operatorname{prox}_{\gamma\Phi}(\overline{\upsilon},\overline{\xi})=\begin{cases}\left(\overline{\upsilon}+\gamma(\rho-1),\overline{\xi}+\gamma\left(\frac{1}{\rho}-1\right)\right) & \text{if } (\overline{\upsilon}<\gamma \text{ and}\\ & (1-\gamma^{-1}\overline{\upsilon})(1-\gamma^{-1}\overline{\xi})<1)\\ & \text{or } \overline{\upsilon}\geq\gamma\\ (0,0) & \text{otherwise}\end{cases}\tag{5.101}$$

where, in the first case, ρ is the unique solution on $]\max(1-\gamma^{-1}\overline{\upsilon},0),+\infty[$ of the equation:

$$\rho^4+(\gamma^{-1}\overline{\upsilon}-1)\rho^3+(1-\gamma^{-1}\overline{\xi})\rho-1=0.\tag{5.102}$$

Proof. For every $(\overline{\upsilon},\overline{\xi})\in\mathbb{R}^2$, $(\upsilon,\xi)=\operatorname{prox}_{\gamma\Phi}(\overline{\upsilon},\overline{\xi})$ is such that $(\upsilon,\xi)\in[0,+\infty[^2$ Moreau [1962]. By using the notation of Proposition 5.3.3 and by using Remark 5.3.4, we have:

$$(\forall\zeta\in\,]0,+\infty[)\qquad \Theta(\zeta)=\frac{\zeta^2}{2}-\frac{2}{5}\zeta^{5/2}+1\tag{5.103}$$

$$\vartheta_-(\zeta)=\vartheta_+(\zeta^{-1})=1-\sqrt{\zeta}\tag{5.104}$$

and

$$\chi_-=\begin{cases}(1-\gamma^{-1}\overline{\upsilon})^2 & \text{if } \overline{\upsilon}<\gamma\\ 0 & \text{otherwise}\end{cases}\tag{5.105}$$

$$\chi_+=\begin{cases}(1-\gamma^{-1}\overline{\xi})^{-2} & \text{if } \overline{\xi}<\gamma\\ +\infty & \text{otherwise.}\end{cases}\tag{5.106}$$

According to Proposition 5.3.3, $(\upsilon,\xi)\in\,]0,+\infty[^2$ if and only if Conditions (i)-(iii) in Lemma 5.3.1 hold. Under these conditions, Proposition 5.3.3 leads to

$$\upsilon=\overline{\upsilon}+\gamma(\widehat{\zeta}^{1/2}-1)\tag{5.107}$$

$$\xi=\overline{\xi}+\gamma\left(\widehat{\zeta}^{-1/2}-1\right)\tag{5.108}$$

where $\widehat{\zeta}$ is the unique minimizer on $]\chi_-,+\infty[$ of the function:

$$\psi\colon\ ]0,+\infty[\to\mathbb{R}\colon\zeta\mapsto\frac{2}{5}\zeta^{5/2}-2\zeta^{1/2}+\frac{\gamma^{-1}\overline{\upsilon}-1}{2}\zeta^2+(1-\gamma^{-1}\overline{\xi})\zeta. \quad (5.109)$$

This means that $\widehat{\zeta}$ is the unique solution on $]\chi_-,+\infty[$ of the equation:

$$\psi'(\widehat{\zeta})=\widehat{\zeta}^{3/2}-\widehat{\zeta}^{-1/2}+(\gamma^{-1}\overline{\upsilon}-1)\widehat{\zeta}+1-\gamma^{-1}\overline{\xi}=0. \quad (5.110)$$

By setting $\rho=\widehat{\zeta}^{1/2}$, (5.102) is obtained.

Since $\lim_{\substack{\zeta\to 0\\ \zeta>0}}\psi'(\zeta)=-\infty$ and $\lim_{\zeta\to+\infty}\psi'(\zeta)=+\infty$, Lemma 5.3.2 shows that Conditions (i)-(iii) are satisfied if and only if

$$\begin{aligned}&\overline{\upsilon}<\gamma,\ \ \overline{\xi}<\gamma,\ \ \text{and}\ \ (1-\gamma^{-1}\overline{\upsilon})^2<(1-\gamma^{-1}\overline{\xi})^{-2}\\ \text{or}\ \ &\overline{\upsilon}<\gamma\ \ \text{and}\ \ \overline{\xi}\geq\gamma\\ \text{or}\ \ &\overline{\upsilon}\geq\gamma\ \ \text{and}\ \ \overline{\xi}<\gamma\\ \text{or}\ \ &\overline{\upsilon}\geq\gamma\ \ \text{and}\ \ \overline{\xi}\geq\gamma\end{aligned} \quad (5.111)$$

or, equivalently

$$\begin{aligned}&\overline{\upsilon}<\gamma\ \ \text{and}\ \ (1-\gamma^{-1}\overline{\upsilon})(1-\gamma^{-1}\overline{\xi})<1\\ \text{or}\ \ &\overline{\upsilon}\geq\gamma.\end{aligned} \quad (5.112)$$

In turn, when (5.112) is not satisfied, we necessarily have $\upsilon=0$ or $\xi=0$. In the first case, the expression of ξ is simply given by the asymmetric soft-thresholding rule in (5.72). Similarly, in the second case, we have

$$\upsilon=\begin{cases}\overline{\upsilon}-\gamma & \text{if } \overline{\upsilon}>\gamma\\ 0 & \text{otherwise.}\end{cases} \quad (5.113)$$

However, when $\overline{\upsilon}>\gamma$ or $\overline{\xi}>\gamma$, (5.111) is always satisfied, so that $\upsilon=\xi=0$.

Altogether, the above results yield the expression of the proximity operator in (5.101). □

5.4.4 Chi square divergence

Let us now consider the function of $\Gamma_0(\mathbb{R}^2)$ given by

$$\Phi\colon(\upsilon,\xi)\mapsto\begin{cases}\dfrac{(\upsilon-\xi)^2}{\xi} & \text{if } \upsilon\in[0,+\infty[\ \text{and}\ \xi\in]0,+\infty[\\ 0 & \text{if } \upsilon=\xi=0\\ +\infty & \text{otherwise.}\end{cases} \quad (5.114)$$

This function satisfies (5.9) with

$$(\forall\zeta\in]0,+\infty[)\qquad\varphi(\zeta)=(\zeta-1)^2. \quad (5.115)$$

Proposition 5.4.6 *The proximity operator of* $\gamma\Phi$ *with* $\gamma \in]0,+\infty[$ *is* $(\forall(\overline{\upsilon},\overline{\xi}) \in \mathbb{R}^2)$

$$\operatorname{prox}_{\gamma\Phi}(\overline{\upsilon},\overline{\xi}) = \begin{cases} \big(\overline{\upsilon}+2\gamma(1-\rho),\overline{\xi}+\gamma(\rho^2-1)\big) & \text{if } \overline{\upsilon} > -2\gamma \\ & \text{and } \overline{\xi} > -\overline{\upsilon}\big(1+(4\gamma)^{-1}\overline{\upsilon}\big) \\ \big(0,\max(\overline{\xi}-\gamma,0)\big) & \text{otherwise} \end{cases} \tag{5.116}$$

where, in the first case, ρ *is the unique solution on* $]0,1+\gamma^{-1}\overline{\upsilon}/2[$ *of the equation:*

$$\rho^3 + (1+\gamma^{-1}\overline{\xi})\rho = 2 + \gamma^{-1}\overline{\upsilon}. \tag{5.117}$$

Proof. By proceeding similarly to the proof of Proposition 5.4.5. we have:

$$(\forall \zeta \in]0,+\infty[) \quad \Theta(\zeta) = 2\zeta - \zeta^2 \tag{5.118}$$

$$\vartheta_-(\zeta) = 2(\zeta^{-1}-1) \tag{5.119}$$

$$\vartheta_+(\zeta) = 1-\zeta^{-2} \tag{5.120}$$

and

$$\chi_- = \begin{cases} \dfrac{2}{2+\gamma^{-1}\overline{\upsilon}} & \text{if } \overline{\upsilon} > -2\gamma \\ +\infty & \text{otherwise} \end{cases} \tag{5.121}$$

$$\chi_+ = \begin{cases} \dfrac{1}{\sqrt{1-\gamma^{-1}\overline{\xi}}} & \text{if } \overline{\xi} < \gamma \\ +\infty & \text{otherwise.} \end{cases} \tag{5.122}$$

According to Proposition 5.3.3, $(\upsilon,\xi) \in]0,+\infty[^2$ if and only if Conditions (i)-(iii) in Lemma 5.3.1 hold. Then, $(\upsilon,\xi) = \operatorname{prox}_{\gamma\Phi}(\overline{\upsilon},\overline{\xi})$ is such that

$$\upsilon = \overline{\upsilon} + 2\gamma(1-\widehat{\zeta}^{-1}) \tag{5.123}$$

$$\xi = \overline{\xi} + \gamma(\widehat{\zeta}^{-2}-1) \tag{5.124}$$

where $\widehat{\zeta}$ is the unique minimizer on $]\chi_-,+\infty[$ of the function:

$$\psi\colon\ ]0,+\infty[\to \mathbb{R}\colon \zeta \mapsto \Big(1+\frac{\gamma^{-1}\overline{\upsilon}}{2}\Big)\zeta^2 - (1+\gamma^{-1}\overline{\xi})\zeta - 2 + \zeta^{-1}. \tag{5.125}$$

Thus, $\widehat{\zeta}$ is the unique solution on $]\chi_-,+\infty[$ of the equation:

$$\psi'(\widehat{\zeta}) = (2+\gamma^{-1}\overline{\upsilon})\widehat{\zeta} - 1 - \gamma^{-1}\overline{\xi} - \widehat{\zeta}^{-2} = 0. \tag{5.126}$$

By setting $\rho = \widehat{\zeta}^{-1}$, (5.117) is obtained.

Lemma 5.3.2 shows that Conditions (ii) and (iii) are satisfied iff

$$\begin{aligned} &\overline{\upsilon} > -2\gamma, \;\; \overline{\xi} < \gamma, \;\; \text{and} \;\; \frac{2}{2+\gamma^{-1}\overline{\upsilon}} < \frac{1}{\sqrt{1-\gamma^{-1}\overline{\xi}}} \\ &\text{or} \;\; \overline{\upsilon} > -2\gamma \;\; \text{and} \;\; \overline{\xi} \geq \gamma \end{aligned} \tag{5.127}$$

or, equivalently,

$$\begin{aligned} &\overline{\upsilon} > -2\gamma, \;\; \overline{\xi} < \gamma, \;\; \text{and} \;\; 1 - \frac{\overline{\xi}}{\gamma} < \left(1 + \frac{\overline{\upsilon}}{2\gamma}\right)^2. \\ &\text{or} \;\; \overline{\upsilon} > -2\gamma \;\; \text{and} \;\; \overline{\xi} \geq \gamma. \end{aligned} \tag{5.128}$$

When (5.128) does not hold, we necessarily have $\upsilon = 0$. The end of the proof is similar to that of Proposition 5.4.5. □

5.4.5 Renyi divergence

Let $\alpha \in]1, +\infty[$ and let us now consider the function of $\Gamma_0(\mathbb{R}^2)$ given by

$$\Phi\colon (\upsilon, \xi) \mapsto \begin{cases} \upsilon^\alpha \xi^{1-\alpha} & \text{if } \upsilon \in [0, +\infty[\text{ and } \xi \in]0, +\infty[\\ 0 & \text{if } \upsilon = \xi = 0 \\ +\infty & \text{otherwise} \end{cases} \tag{5.129}$$

which corresponds to the case when

$$(\forall \zeta \in]0, +\infty[) \qquad \varphi(\zeta) = \zeta^\alpha. \tag{5.130}$$

Note that the above function Φ allows us to generate the Rényi divergence up to a log transform and a multiplicative constant.

Proposition 5.4.7 *The proximity operator of $\gamma\Phi$ with $\gamma \in]0, +\infty[$ is*

$$\begin{aligned} &(\forall (\overline{\upsilon}, \overline{\xi}) \in \mathbb{R}^2) \\ &\operatorname{prox}_{\gamma\Phi}(\overline{\upsilon}, \overline{\xi}) = \begin{cases} \big(\overline{\upsilon} - \gamma\alpha\widehat{\zeta}^{1-\alpha}, \overline{\xi} + \gamma(\alpha-1)\widehat{\zeta}^{-\alpha}\big) & \text{if } \overline{\upsilon} > 0 \quad \text{and} \\ & \frac{\gamma^{\frac{1}{\alpha-1}}\overline{\xi}}{1-\alpha} < \left(\frac{\overline{\upsilon}}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} \\ \big(0, \max(\overline{\xi}, 0)\big) & \text{otherwise} \end{cases} \end{aligned} \tag{5.131}$$

where, in the first case, $\widehat{\zeta}$ is the unique solution on $](\alpha\gamma\overline{\upsilon}^{-1})^{\frac{1}{\alpha-1}}, +\infty[$ of the equation:

$$\gamma^{-1}\overline{\upsilon}\,\widehat{\zeta}^{1+\alpha} = \gamma^{-1}\overline{\xi}\,\widehat{\zeta}^{\alpha} + \alpha\widehat{\zeta}^2 + \alpha - 1. \tag{5.132}$$

Proof. We proceed similarly to the previous examples by noticing that

$$(\forall \zeta \in]0,+\infty[) \qquad \Theta(\zeta) = \begin{cases} \frac{\alpha}{3-\alpha}\zeta^{3-\alpha} & \text{if } \alpha \neq 3 \\ \alpha \ln \zeta & \text{if } \alpha = 3 \end{cases} \tag{5.133}$$

$$\vartheta_-(\zeta) = \alpha\zeta^{1-\alpha} \tag{5.134}$$

$$\vartheta_+(\zeta) = (1-\alpha)\zeta^{-\alpha} \tag{5.135}$$

$$\psi'(\zeta) = (1-\alpha)\zeta^{-\alpha} - \alpha\zeta^{2-\alpha} + \gamma^{-1}\overline{\upsilon}\zeta - \gamma^{-1}\overline{\xi} \tag{5.136}$$

and

$$\chi_- = \begin{cases} \left(\frac{\gamma\alpha}{\overline{\upsilon}}\right)^{\frac{1}{\alpha-1}} & \text{if } \overline{\upsilon} > 0 \\ +\infty & \text{otherwise} \end{cases} \tag{5.137}$$

$$\chi_+ = \begin{cases} \left(\frac{\gamma(1-\alpha)}{\overline{\xi}}\right)^{1/\alpha} & \text{if } \overline{\xi} < 0 \\ +\infty & \text{otherwise.} \end{cases} \tag{5.138}$$

□

Note that (5.132) becomes a polynomial equation when α is a rational number. In particular, when $\alpha = 2$, it reduces to the cubic equation:

$$\rho^3 + (2 + \gamma^{-1}\overline{\xi})\rho = \gamma^{-1}\overline{\upsilon} \tag{5.139}$$

with $\widehat{\zeta} = \rho^{-1}$.

5.4.6 I_α divergence

Let $\alpha \in]0,1[$ and let us now consider the function of $\Gamma_0(\mathbb{R}^2)$ given by

$$\Phi\colon (\upsilon,\xi) \mapsto \begin{cases} \alpha\upsilon + (1-\alpha)\xi - \upsilon^\alpha\xi^{1-\alpha} & \text{if } \upsilon \in [0,+\infty[\text{ and } \xi \in [0,+\infty[\\ +\infty & \text{otherwise} \end{cases} \tag{5.140}$$

which corresponds to the case when

$$(\forall \zeta \in]0,+\infty[) \qquad \varphi(\zeta) = 1 - \alpha + \alpha\zeta - \zeta^\alpha. \tag{5.141}$$

Proposition 5.4.8 *The proximity operator of $\gamma\Phi$ with $\gamma \in]0,+\infty[$ is* $(\forall (\overline{v},\overline{\xi}) \in \mathbb{R}^2)$

$$\operatorname{prox}_{\gamma\Phi}(\overline{v},\overline{\xi}) = \begin{cases} \big(\overline{v}+\gamma\alpha(\widehat{\zeta}^{1-\alpha}-1), \overline{\xi}+\gamma(1-\alpha)(\widehat{\zeta}^{-\alpha}-1)\big) & \text{if } \Big(\overline{v} < \gamma\alpha \text{ and } \\ & \quad \Big(1-\frac{\gamma^{-1}\overline{\xi}}{(1-\alpha)}\Big) < \Big(1-\frac{\overline{v}}{\gamma\alpha}\Big)^{\frac{\alpha}{\alpha-1}}\Big) \\ & \text{or } \overline{v} \geq \gamma\alpha \\ (0,0) & \text{otherwise} \end{cases} \tag{5.142}$$

where, in the first case, $\widehat{\zeta}$ is the unique solution on $\big]\big(\max\big(1-\frac{\overline{v}}{\gamma\alpha},0\big)\big)^{\frac{1}{1-\alpha}},+\infty[$ of the equation:

$$\alpha\widehat{\zeta}^2 + (\gamma^{-1}\overline{v}-\alpha)\widehat{\zeta}^{\alpha+1} + (1-\alpha-\gamma^{-1}\overline{\xi})\widehat{\zeta}^{\alpha} = 1-\alpha. \tag{5.143}$$

Proof. We have then $(\forall \zeta \in]0,+\infty[)$

$$\Theta(\zeta) = \alpha\Big(\frac{\zeta^2}{2} - \frac{\zeta^{3-\alpha}}{3-\alpha}\Big) \tag{5.144}$$

$$\vartheta_-(\zeta) = \alpha(1-\zeta^{1-\alpha}) \tag{5.145}$$

$$\vartheta_+(\zeta) = (1-\alpha)(1-\zeta^{-\alpha}) \tag{5.146}$$

$$\psi'(\zeta) = \alpha\zeta^{2-\alpha} + (\gamma^{-1}\overline{v}-\alpha)\zeta + (\alpha-1)\zeta^{-\alpha} + 1 - \alpha - \gamma^{-1}\overline{\xi} \tag{5.147}$$

and

$$\chi_- = \begin{cases} \Big(1-\frac{\overline{v}}{\gamma\alpha}\Big)^{\frac{1}{1-\alpha}} & \text{if } \overline{v} < \gamma\alpha \\ 0 & \text{otherwise} \end{cases} \tag{5.148}$$

$$\chi_+ = \begin{cases} \Big(1-\frac{\overline{\xi}}{\gamma(1-\alpha)}\Big)^{-1/\alpha} & \text{if } \overline{\xi} < \gamma(1-\alpha) \\ +\infty & \text{otherwise.} \end{cases} \tag{5.149}$$

The result follows by noticing that $\lim_{\substack{\zeta\to 0\\ \zeta>0}} \psi'(\zeta) = -\infty$ and $\lim_{\zeta\to+\infty} \psi'(\zeta) = +\infty$. □

As for the Renyi divergence, (5.143) becomes a polynomial equation when α is a rational number.

§ 5.5 Connection with epigraphical projections

A number of recent works Chierchia et al. [2013]; Harizanov et al. [2013]; Tofighi et al. [2014]; Ono and Yamada [2014] have shown that the projection

onto the epigraph of a convex function is a useful tool for solving constrained convex optimization with the help of proximal algorithms. Epigraphical projection may be of main interest when dealing with constraint sets for which a closed form expression of the projection does not exist. In this part, we briefly explain how the forms of the proximity operator of divergences which have been derived in the previous sections can be employed to deduce the expression of the projection onto the epigraph of some convex functions.

Proposition 5.5.1 *Let $\varphi\colon \mathbb{R} \to [0,+\infty]$ be a function in $\Gamma_0(\mathbb{R})$ which is twice differentiable on $]0,+\infty[$. Let Φ be the function defined by (5.9) and $\varphi^* \in \Gamma_0(\mathbb{R})$ the Fenchel-conjugate function of the restriction of φ on $[0,+\infty[$, defined as*

$$(\forall \zeta^* \in \mathbb{R}) \qquad \varphi^*(\zeta) = \sup_{\zeta \in [0,+\infty[} \big(\zeta\zeta^* - \varphi(\zeta)\big). \tag{5.150}$$

Let $\operatorname{epi}\varphi^* = \big\{(\upsilon^*,\xi^*) \in \mathbb{R}^2 \;\big|\; \varphi^*(\upsilon^*) \leq \xi^*\big\}$ *be the epigraph of φ^*. Then, the projection onto* $\operatorname{epi}\varphi^*$ *is given by*

$$\big(\forall (\upsilon^*,\xi^*) \in \mathbb{R}^2\big) \qquad \mathsf{P}_{\operatorname{epi}\varphi^*}(\upsilon^*,\xi^*) = (\upsilon^*,-\xi^*) - \operatorname{prox}_{\Phi}(\upsilon^*,-\xi^*). \tag{5.151}$$

Proof. The conjugate function of Φ is

$$(\forall (\upsilon,\xi) \in \mathbb{R}^2) \qquad \Phi^*(\upsilon^*,\xi^*) = \sup_{(\upsilon,\xi)\in\mathbb{R}^2} \big(\upsilon\upsilon^* + \xi\xi^* - \Phi(\upsilon,\xi)\big). \tag{5.152}$$

From the definition of Φ, we deduce that

$$\begin{aligned}
&(\forall (\upsilon,\xi) \in \mathbb{R}^2) && && (5.153)\\
&\Phi^*(\upsilon^*,\xi^*) = \sup\Bigg\{ && \sup_{(\upsilon,\xi)\in[0,+\infty[\times]0,+\infty[} \Big(\upsilon\upsilon^* + \xi\xi^* - \xi\varphi\Big(\frac{\upsilon}{\xi}\Big)\Big), && (5.154)\\
& && \sup_{\upsilon\in]0,+\infty[} \Big(\upsilon\upsilon^* - \lim_{\substack{\xi\to 0\\ \xi>0}} \xi\varphi\Big(\frac{\upsilon}{\xi}\Big)\Big), 0\Bigg\}\\
& = \sup\Bigg\{ && \sup_{(\upsilon,\xi)\in[0,+\infty[\times]0,+\infty[} \Big(\upsilon\upsilon^* + \xi\xi^* - \xi\varphi\Big(\frac{\upsilon}{\xi}\Big)\Big), 0\Bigg\}\\
& = \sup\{\iota_{\operatorname{epi}\varphi^*}(\upsilon^*,-\xi^*), 0\} && && (5.155)\\
& = \iota_{\operatorname{epi}\varphi^*}(\upsilon^*,-\xi^*), && && (5.156)
\end{aligned}$$

where the equality in (5.155) stems from [Bauschke and Combettes, 2011, Example 13.8]. Then, (5.151) follows from the conjugation property of the proximity operator (see Proposition 5.2.3 (v)). □

The above proposition is applicable to all the examples of φ-divergences wich have been presented in Section 5.3. In Table 5.1, we give the expressions of the corresponding functions φ^*.

Divergence	$\varphi(\zeta)$ $\zeta > 0$	$\varphi^*(\zeta^*)$ $\zeta^* \in \mathbb{R}$
Kullback-Leibler	$\zeta \ln \zeta - \zeta + 1$	$e^{\zeta^*} - 1$
Jeffreys-Kullback	$(\zeta - 1) \ln \zeta$	$W(e^{1-\zeta^*}) + \big(W(e^{1-\zeta^*})\big)^{-1} + \zeta^* - 2$
Chi square	$(\zeta - 1)^2$	$\begin{cases} \frac{\zeta^*(\zeta^*+4)}{4} & \text{if } \zeta^* \geq -2 \\ -1 & \text{otherwise} \end{cases}$
Hellinger	$1 + \zeta - 2\sqrt{\zeta}$	$\begin{cases} \frac{\zeta^*}{1-\zeta^*} & \text{if } \zeta^* > -1 \\ +\infty & \text{otherwise} \end{cases}$
Renyi, $\alpha \in]1, +\infty[$	ζ^α	$\begin{cases} (\alpha-1)\Big(\frac{\zeta^*}{\alpha}\Big)^{\frac{\alpha}{\alpha-1}} & \text{if } \zeta^* \geq 0 \\ 0 & \text{otherwise} \end{cases}$
I_α, $\alpha \in]0, 1[$	$1 - \alpha + \alpha\zeta - \zeta^\alpha$	$\begin{cases} (1-\alpha)\Big(\Big(1 - \frac{\zeta^*}{\alpha}\Big)^{\frac{\alpha}{\alpha-1}} - 1\Big) & \text{if } \zeta^* \leq \alpha \\ +\infty & \text{otherwise} \end{cases}$

Table 5.1: *Conjugate function φ^* of the restriction of φ to $[0, +\infty[$.*

§ 5.6 Conclusion

In this chapter, we have shown how to deal with convex optimization problems involving discrete information divergences. Our analysis has emphasized that the difficulty lies in computing the proximity operator of two-variable real functions. To this end, we have carried out a systematic study of the proximity properties of φ-divergences, which has eventually led us to derive new closed-form expressions of proximity operators. These novel results have shed light on a new methodology that may replace the approaches usually followed to deal with problems involving these divergences (i.e. minimizing by fixing one of the two arguments or by alternating between the two variables).

In addition, we have provided a closed-form expression of the projection onto the epigraph of convex functions related to the considered φ-divergences.

In the next chapter, we will consider applications of these divergences, which may be useful in some problems, in the context of image restoration and image registration.

§ 5.7 Convergence proof of algorithm 9

We aim at finding the zero of the function ψ' given below (along with its derivative ψ''):
$(\forall\, \zeta \in]W(e^{1-\gamma^{-1}\overline{v}}), +\infty[)$

$$\psi'(\zeta) = (\zeta+1)\log\zeta + \frac{\zeta}{2} + 1 - \frac{1}{\zeta} + \zeta^2 + \Big(\gamma^{-1}\overline{v} - \frac{3}{2}\Big)\zeta - \gamma^{-1}\overline{\xi},$$
$$\psi''(\zeta) = \log\zeta + \frac{1}{\zeta} + \frac{1}{\zeta^2} + 2\zeta + \gamma^{-1}\overline{v}.$$

To do so, we employ the Newton algorithm, whose global convergence is guaranteed for any initial value by the following condition Thorlund-Petersen [2004]:

$$(\forall a \in]0,+\infty[)(\forall b \in]a,+\infty[) \qquad \psi''(a) + \psi''(b) > \frac{\psi'(b) - \psi'(a)}{b-a},$$

which is equivalent to

$$(b-a)\psi''(a) + (b-a)\psi''(b) - \psi'(b) + \psi'(a) > 0. \tag{5.157}$$

Condition (5.157) can be rewritten as follows

$$\begin{aligned}
&(b-a)\Big(\log a + \frac{1}{a} + \frac{1}{a^2} + 2a + \gamma^{-1}\overline{v}\Big)\\
&+(b-a)\Big(\log b + \frac{1}{b} + \frac{1}{b^2} + 2b + \gamma^{-1}\overline{v}\Big)\\
&+\bigg((a+1)\log a + \frac{a}{2} + 1 - \frac{1}{a} + a^2 + \Big(\gamma^{-1}\overline{v} - \frac{3}{2}\Big)a - \gamma^{-1}\overline{\xi}\bigg)\\
&-\bigg((b+1)\log b + \frac{b}{2} + 1 - \frac{1}{b} + b^2 + \Big(\gamma^{-1}\overline{v} - \frac{3}{2}\Big)b - \gamma^{-1}\overline{\xi}\bigg) > 0,
\end{aligned}$$

which, after some simplification, boils down to

$$\begin{aligned}
&(b+1)\log a - (a+1)\log b + \frac{b}{a} + \frac{b}{a^2} - \frac{2}{a} - \frac{a}{b} - \frac{a}{b^2} + \frac{2}{b} - a^2 + b^2\\
&+\gamma^{-1}\overline{v}(b-a) + \frac{1}{2}(b-a) > 0.
\end{aligned} \tag{5.158}$$

We now show that Condition (5.158) holds true because $b > a$ and it is a sum of two terms

$$\underbrace{(b+1)\log a - (a+1)\log b + \frac{b}{a} + \frac{b}{a^2} - \frac{2}{a} - \frac{a}{b} - \frac{a}{b^2} + \frac{2}{b} - a^2 + b^2}_{>0} \tag{5.159}$$

and

$$\underbrace{\gamma^{-1}\overline{v}(b-a)+\frac{1}{2}(b-a)}_{>0}. \tag{5.160}$$

Proof. Equation (5.159) can be rewritten as

$$(\forall a \in]0,+\infty[)(\forall b \in]a,+\infty[) \qquad g(a,b)-g(b,a)>0$$

where

$$g(x,y)=-(x+1)\log y-\frac{x}{y}+\frac{y}{x^2}-\frac{2}{x}+y^2.$$

Therefore, we shall demonstrate that, for every $b>a>0$, g is decreasing w.r.t. the first argument and increasing w.r.t. to the second argument, i.e.

$$g(a,b)>g(b,b) \qquad \text{and} \qquad g(b,b)>g(b,a), \tag{5.161}$$

which implies that

$$g(a,b)>g(b,a).$$

To prove these two inequalities, we will study the derivative of g with respect to its arguments. The conditions in (5.161) are indeed equivalent to

$$(\forall y \in]0,+\infty[)(\forall x \in]0,y[) \qquad \frac{\partial g}{\partial x}(x,y)<0, \tag{5.162}$$

$$(\forall x \in]0,+\infty[)(\forall y \in]x,+\infty[) \qquad \frac{\partial g}{\partial y}(x,y)>0. \tag{5.163}$$

The first and second partial derivatives of g w.r.t. x read

$$(\forall y \in]0,+\infty[)(\forall x \in]0,y[) \qquad \begin{aligned}\frac{\partial g}{\partial x}(x,y)&=-\log y-\frac{1}{y}-\frac{2y}{x^3}+\frac{2}{x^2}\\ \frac{\partial^2 g}{\partial x^2}(x,y)&=\frac{6y}{x^4}-\frac{4}{x^3}=\frac{6y-4x}{a^4}>0.\end{aligned}$$

Since $\partial^2 g/\partial x^2$ is strictly positive, $\partial g/\partial x$ is strictly increasing w.r.t. x and

$$\lim_{x\to y}\frac{\partial g}{\partial x}(x,y)=-\log y-\frac{1}{y}=\log\frac{1}{y}-\frac{1}{y}<0.$$

Therefore, Condition (5.162) holds, and g is decreasing with respect to x. The first and second partial derivatives of g w.r.t. y read

$$(\forall x \in]0,+\infty[)(\forall y \in]x,+\infty[) \qquad \begin{aligned}\frac{\partial g}{\partial y}(x,y)&=\frac{-x}{y}-\frac{1}{y}+\frac{x}{y^2}+\frac{1}{x^2}+2y\\ \frac{\partial^2 g}{\partial y^2}(x,y)&=\frac{x}{y^2}+\frac{1}{y^2}-\frac{2x}{y^3}+2.\end{aligned}$$

For every $y \in [1, +\infty[$,

$$(\forall x \in]0, y[) \qquad \frac{\partial^2 g}{\partial y^2}(x, y) = \frac{x}{y^2} + \frac{1}{y^2} - \underbrace{\frac{2x}{y^3}}_{<1} + 2 > 0,$$

and $\partial g / \partial y$ is strictly increasing w.r.t. y (since $\partial^2 g / \partial y^2$ is strictly positive) and

$$\begin{aligned} (\forall x \in [1, +\infty[) \qquad & \lim_{y \to x} \frac{\partial g}{\partial y}(x, y) = -1 + \frac{1}{x^2} + 2x > 0, \\ (\forall x \in]0, 1]) \qquad & \frac{\partial g}{\partial y}(x, 1) = \frac{1}{x^2} + 1 > 0. \end{aligned}$$

For every $y \in]0, 1[$, we have

$$\begin{aligned} (\forall x \in]0, y[) \qquad \frac{\partial g}{\partial y}(x, y) &= \frac{-x}{y} + \frac{x}{y^2} - \frac{1}{y} + \frac{1}{x^2} + 2y, \\ &= \frac{x - xy}{y^2} - \frac{1}{y} + \frac{1}{x^2} + 2y, \end{aligned}$$

since $x < y < 1$, this implies that $xy < x$ and $\frac{1}{y} < \frac{1}{x} < \frac{1}{x^2}$ and

$$(\forall x \in]0, y[) \qquad \frac{\partial g}{\partial y}(x, y) = \underbrace{\frac{x - xy}{y^2}}_{>0} \underbrace{- \frac{1}{y} + \frac{1}{x^2}}_{>0} + 2y > 0.$$

As Condition (5.163) holds, g is increasing with respect to y. □

- Chapter 6 -

2D Applications of divergence proximity operators

"Knowledge without application is like a book that is never read."
`Christopher Crawford`

Measures based on distance or divergence play a pivotal role in evaluating the dissimilarity between two objects (such as numbers, vectors, matrices, functions, images, ...). In this chapter, we investigate the use of divergences as a data fidelity term or a regularization term in two different image processing applications: restoration and registration.

As we have presented in the previous chapter, the objective function will have the following form:

$$x \mapsto D(Ax, Bx) \tag{6.1}$$

where D is a function in $\Gamma_0(\mathbb{R}^P \times \mathbb{R}^P)$, and A and B are matrices in $\mathbb{R}^{P\times N}$. The most common assumption on the signal of interest $x \in \mathbb{R}^N$ is that, by making an appropriate choice of matrices A and B, Ax and Bx are close in the metric induced by D.

Divergences have been widely used in the literature within many different contexts. Some measures have been used for regression analysis, some have been applied to compare two probability density functions, and some are used to measure the dissimilarity between two random vectors of the same distribution. In this chapter, we use information measures in two applications. Firstly, we extend the NLTV framework by using some information divergences to build new sparsity measures for signal recovery. In denoising and deconvolution examples, our approach is compared with other existing approaches. Secondly, an image registration application constitutes another example illustrating the use of the proposed divergence proximity operators.

§ 6.1 Regularization based on φ-divergences

In this section, we consider divergences for building non-local smoothness measures. The Non-Local Total Variation Gilboa and Osher [2009]; Buades et al. [2006] has been used as a popular and effective image prior model in regularization-based imaging problems Werlberger et al. [2010]; Peyré [2011]; Chierchia et al. [2013]; Couprie et al. [2013]; Teuber et al. [2013]. As we have mentioned in Chapter 2, NLTV is a powerful regularization operator. It makes full use of spatial information distributed over the different image regions.

In Equation (3.48), we have defined NLTV as follows:

$$\mathrm{NLTV}(x) = \sum_{s\in\mathcal{A}} \sqrt{\sum_{n\in\mathcal{N}_s\subset\mathcal{W}_s} \omega_{s,n}|x^{(s)} - x^{(n)}|^2}, \tag{6.2}$$

where $\mathcal{W}_s$ is the set of positions $n \in \mathcal{A}\setminus\{s\}$ located into a $Q\times Q$ window centered at s, where $Q\in\mathbb{N}$ is odd. The idea behind NLTV is illustrated in Fig.6.1. For each pixel in the image, we select some neighbours within a limited region and we compute the weights $\omega_{s,n}$ according to a similarity measure between $\widetilde{Q}\times\widetilde{Q}$ patches built around pixels s and n. Large weights are desirable in flat areas in order to effectively suppress the noise, while small weights are desirable around object edges in order to preserve details and fine structures.

A key observation is that NLTV can be interpreted as a similarity measure. Indeed, the vector of weighted differences can be rewritten as:

$$(\forall s\in\mathcal{A}) \qquad [\omega_{s,n}(x^{(s)} - x^{(n)})]_{n\in\mathcal{N}_s} = A_s x - B_s x, \tag{6.3}$$

where

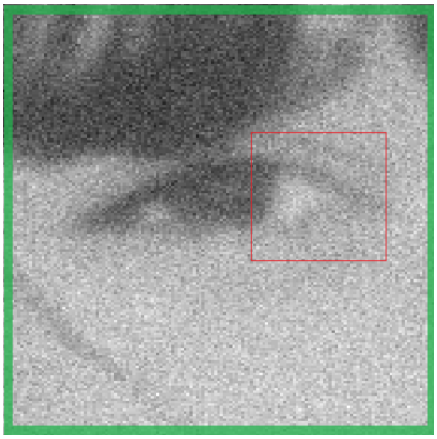
$$A_s : x \mapsto [\omega_{s,n}x^{(s)}]_{n\in\mathcal{N}_s}, \tag{6.4}$$
$$B_s : x \mapsto [\omega_{s,n}x^{(n)}]_{n\in\mathcal{N}_s}. \tag{6.5}$$

Consequently, the $\ell_{1,p}$-NLTV regularization term can be expressed as follows

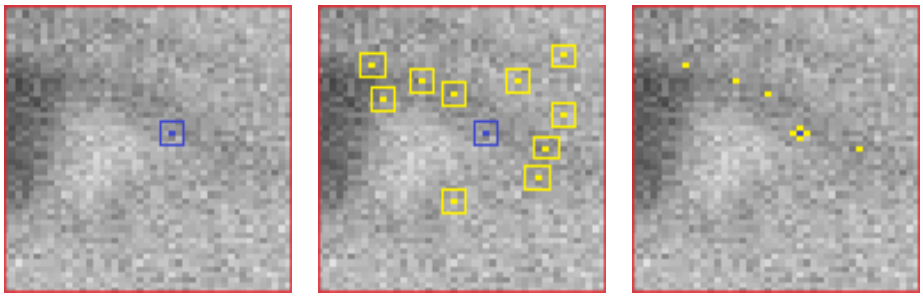
$$\ell_{1,p}-\mathrm{NLTV}(x) = \sum_{s\in\mathcal{A}} \|A_s x - B_s x\|_p. \tag{6.6}$$

Let, for all $s\in\mathcal{A}$, $a^{(s)} = A_s x \in [0,+\infty[^{|\mathcal{N}_s|}$ and $b^{(s)} = B_s x \in [0,+\infty[^{|\mathcal{N}_s|}$. Thus, Equation (6.6) is equivalent to

$$\ell_{1,p}-\mathrm{NLTV}(x) = \sum_{s\in\mathcal{A}} \|a^{(s)} - b^{(s)}\|_p. \tag{6.7}$$



(a) *Fragment of the noisy image (Chapter 2) Image part highlighted by red color is illustrated in (b), (c) and (d).*



(b) *Reference pixel and surrounding 5×5 bloc.* **(c)** *Candidates* **(d)** *Chosen neighboors*

Figure 6.1: *Illustration of the idea behind nonlocal total variation.*

By setting now $a^{(s)} = (a^{(s,m)})_{1\leq m\leq|\mathcal{N}_s|}$ and $b^{(s)} = (b^{(s,m)})_{1\leq m\leq|\mathcal{N}_s|}$, then the smoothed version of NLTV regularization, namely the squared ℓ_2-NLTV, can be expressed as

$$\ell_2\text{-NLTV}(x) = \sum_{s\in\mathcal{A}}\sum_{m=1}^{|\mathcal{N}_s|}(a^{(s,m)} - b^{(s,m)})^2. \tag{6.8}$$

By analogy, we propose to define the φ-divergence version of NLTV as

$$\text{D-NLTV}(a,b) = \sum_{s\in\mathcal{A}}\sum_{m=1}^{|\mathcal{N}_s|} b^{(s,m)}\varphi\left(\frac{a^{(s,m)}}{b^{(s,m)}}\right), \tag{6.9}$$

where $\varphi\colon \mathbb{R}\to[0,+\infty]$ belongs to $\Gamma_0(\mathbb{R})$ and is twice differentiable on $]0,+\infty[$.

The aim of this part is to investigate the influence of the choice of the similarity measure on the performance of optimization approaches in the context of image recovery.

6.1.1 Numerical experiments

In order to assess the relevance of the proposed regularization, we present numerical experiments relative to an image restoration problem. The considered degradation model is

$$z = H\overline{x} + w,$$

where $\overline{x} \in \mathbb{R}^N$ denotes the image to be recovered, $z \in \mathbb{R}^M$ are the observed data, $H \in \mathbb{R}^{M\times N}$ corresponds to a matrix associated with a blur operator, and $w \in \mathbb{R}^M$ is a realization of an additive zero-mean white Gaussian noise. The restoration process is based on a variational approach aiming at finding an image $x \in \mathbb{R}^N$ as close as possible to $\overline{x} \in \mathbb{R}^N$ from the observation z. Indeed, as inverse problems are usually ill-posed, one needs some prior information about the estimate, such as the dynamic range of the pixel values and the smoothness of the estimate of $\bar{x}$. Therefore, the restoration can be achieved by solving the following convex optimization problem:

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; \frac{1}{2\lambda}\|Hx - z\|^2 + D(Ax, Bx) + \iota_C(x) \tag{6.10}$$

where $\lambda \in]0, +\infty[$ is the regularization constant, and the operators A and B are chosen as explained before. The last term ι_C constrains x to belong to the convex set $C = [0, 255]^N$.

Alternatively, we seek to describe the problem within a set theoretic framework. In several works Combettes and Trussell [1991], it was observed that an upper bound on the data fidelity term allows us to efficiently restrict the solution to vectors x such that:

$$Hx \in C' = \left\{u \in \mathbb{R}^M \mid \|u - z\|^2 \leq \delta M\sigma^2\right\} \tag{6.11}$$

where δ is a positive constant (usually close to 1) and σ^2 is the noise variance. This leads to the following variant of Problem (6.10):

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; D(Ax, Bx) + \iota_{C'}(Hx) + \iota_C(x). \tag{6.12}$$

Note that, under technical assumptions (Chapter 2, Section 2.1.1), Problem (6.12) is equivalent to Problem (6.10). However the constrained formulation given above is usually considered to be more practical as the solution is less sensitive to the choice of δ than λ Chierchia et al. [2013].

The above problem can be solved using proximal optimization algorithms. Such methods require to compute the proximity operator of the divergence D, the projection onto the ℓ_2 ball, and the projection onto the hypercube $[0, 255]^N$. These two projections are quite standard. A possible algorithm for solving Problem (6.12) is thus the M+LFBF primal-dual algorithm Combettes and Pesquet [2012]. The associated iterations are recalled in Algorithm 1, where at each iteration k, $\gamma^{[k]}$ is a step-size and $e^{[k]} \in (\mathbb{R}^P)^2$

corresponds to a possible error in the computation of the proximity operators of the divergence term.

Algorithm 10 M+LFBF

`Initialization`

$v_1^{[0]} \in \mathbb{R}^P$, $v_2^{[0]} \in \mathbb{R}^P$, $v_3^{[0]} \in \mathbb{R}^M$, $x^{[0]} \in \mathbb{R}^N$

$\beta = \left(\|A\|^2 + \|B\|^2 + \|H\|^2 \right)^{1/2}$, $\varepsilon \in]0, 1/(\beta+1)[$

`For` $k = 0, 1, \ldots$

$\gamma^{[k]} \in [\varepsilon, (1-\varepsilon)/\beta]$

$y_1^{[k]} = x^{[k]} - \gamma^{[k]}(A^\top v_1^{[k]} + B^\top v_2^{[k]} + H^\top v_3^{[k]})$

$p_1^{[k]} = \mathsf{P}_C(y_1^{[k]})(y_{2,0}^{[k]}, y_{2,1}^{[k]}) = (v_1^{[k]}, v_2^{[k]}) + \gamma^{[k]}(Ax^{[k]}, Bx^{[k]})$

$(p_{2,0}^{[k]}, p_{2,1}^{[k]}) = (y_{2,0}^{[k]}, y_{2,1}^{[k]}) - \gamma^{[k]} \mathrm{prox}_{\frac{D}{\gamma^{[k]}}} \left(\frac{y_{2,0}^{[k]}}{\gamma^{[k]}}, \frac{y_{2,1}^{[k]}}{\gamma^{[k]}} \right) + e^{[k]}$

$(q_{2,0}^{[k]}, q_{2,1}^{[k]}) = (p_{2,0}^{[k]}, p_{2,1}^{[k]}) + \gamma^{[k]}(Ap_1^{[k]}, Bp_1^{[k]})$

$(v_1^{[k+1]}, v_2^{[k+1]}) = (v_1^{[k]}, v_2^{[k]}) - (y_{2,0}^{[k]}, y_{2,1}^{[k]}) + (q_{2,0}^{[k]}, q_{2,1}^{[k]})$

$y_3^{[k]} = v_3^{[k]} + \gamma^{[k]} H x^{[k]}$

$p_3^{[k]} = y_3^{[k]} - \gamma^{[k]} z - \gamma^{[k]} \mathsf{P}_{C'} \left(\frac{y_3^{[k]}}{\gamma^{[k]}} - z \right)$

$q_3^{[k]} = p_3^{[k]} + \gamma^{[k]} H p_1^{[k]}$

$v_3^{[k+1]} = v_3^{[k]} - y_3^{[k]} + q_3^{[k]}$

$q_1^{[k]} = p_1^{[k]} - \gamma^{[k]}(A^\top p_{2,0}^{[k]} + B^\top p_{2,1}^{[k]} + H^\top p_3^{[k]})$

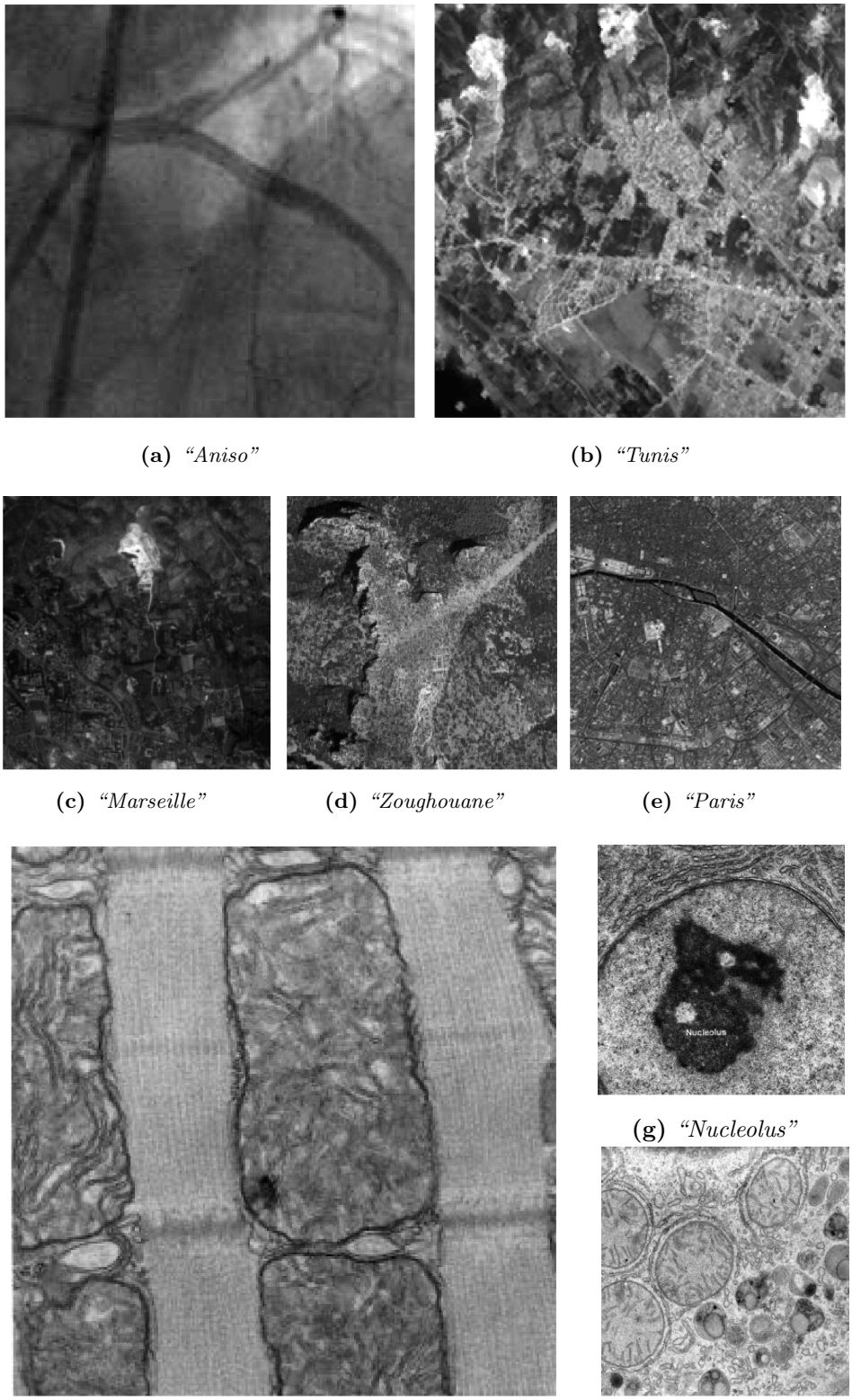
$x^{[k+1]} = x^{[k]} - y_1^{[k]} + q_1^{[k]}$.

6.1.2 Results

In this section, we present the performance of the proposed D-NLTV regularization method in restoration experiments. In particular, the following choices for D are considered: squared ℓ_2-norm, $\ell_{1,2}$-norm, Kullback-Leibler divergence, Jeffreys-Kullback, Hellinger divergence, Chi square divergence, and I_α divergence (α=0.2). Note also that Rényi divergence is not suitable for the considered application, since its use is limited to probability distributions.

In our experiments, we have considered three types of images: angiographic,[1] medical, [2] and satellite images (Figure 6.2).

[1] www.pcronline.com
[2] www.cellimagelibrary.org/



(a) *"Aniso"* **(b)** *"Tunis"*

(c) *"Marseille"* **(d)** *"Zoughouane"* **(e)** *"Paris"*

(g) *"Nucleolus"*

(f) *"Cardiac muscle"* **(h)** *"Lysosomes"*

Figure 6.2: *Example of angiographic (a), satellite (b,c,d,e), and medical (f,g,h) images.*

Table 6.1: (SNR, SSIM) *denoising results with* σ^2*=400.*

$\ell_{1,2}$-TV	$\ell_{1,2}$-NLTV	KL-NLTV	JK-NLTV	Hel-NLTV	Chi-NLTV	$I_{0.2}$-NLTV
Tunis						
21.32-0.795	21.56-0.808	21.73-0.823	21.71-0.822	21.69-0.824	21.70-0.822	21.49-0.808
Marseille						
18.66-0.759	19.12-0.784	19.62-0.807	19.68-0.807	19.16-0.804	19.68-0.807	19.11-0.793
Zoughouane						
21.94-0.702	22.19-0.715	22.63-0.749	22.61-0.748	22.54-0.755	22.61-0.748	22.43-0.734
Paris						
18.07-0.726	18.38-0.738	18.64-0.767	18.64-0.767	18.67-0.767	18.63-0.766	18.59-0.765
Aniso						
24.56-0.830	25.11-0.837	26.33-0.876	26.28-0.874	26.51-0.881	26.32-0.876	26.36-0.875
Nucleolus						
18.58-0.803	18.85-0.817	18.95-0.825	18.95-0.826	18.96-0.825	18.93-0.825	18.84-0.814
Cardiac muscle						
22.21-0.709	22.56-0.708	22.97-0.759	22.96-0.758	22.90-0.762	22.95-0.758	22.78-0.736
Lysosomes						
22.14-0.788	22.48-0.800	22.54-0.808	22.53-0.808	22.56-0.810	22.52-0.808	22.45-0.810

The observed image is generated by degrading the original one with a convolution operator H, which is equal to identity for the denoising problem and corresponds to a truncated Gaussian point spread function with standard deviation 1.6 and kernel size 3×3, and 7×7 for the deconvolution problem. The noise variance is equal to 400 and 64 for the image denoising and restoration problems, respectively. The linear operators A and B associated with NLTV are computed from the TV image result obtained using the code in Foi and Boracchi [2012]. For the regularization parameters, we set $Q = 11$ and $\widetilde{Q} = 5$.

The balance between the smoothness of the estimate and the data fidelity is controlled by the parameter δ tuned so as to maximize the Signal-to-Noise Ratio (SNR). The quality of the results, presented in Tables 6.1,6.2, and 6.3, is evaluated in terms of the SNR and the Structural Similarity index (SSIM) Malpica and Bovik [2009]. One can observe that the results obtained with D-NLTV improve upon the standard $\ell_{1,2}$-NLTV, for the different divergences with a gain up to **0.44** dB. Fig. 6.3 and 6.4 show that our method using Div-NLTV leads to significant visual improvements compared with $\ell_{1,2}$-TV and $\ell_{1,2}$-NLTV. Note that the considered divergences lead to similar results. We believe that this behavior can be explained by the fact that the φ-divergences present little difference when the two arguments are similar, i.e. the ratio is close to one. Indeed, the Taylor expansion around one is the same and equal to the quadratic difference as shown next.

Table 6.2: (SNR, SSIM) *restoration results with* σ^2*=64 and a blur operator of size 3×3.*

$\ell_{1,2}$-TV	$\ell_{1,2}$-NLTV	KL-NLTV	JK-NLTV	Hel-NLTV	Chi-NLTV	$I_{0.2}$-NLTV
Tunis						
22.37-0.843	22.70-0.851	23.09-0.871	23.12-0.871	22.77-0.863	23.05-0.868	22.70-0.864
Marseille						
20.84-0.837	21.19-0.852	21.93-0.873	21.94-0.873	21.90-0.873	21.70-0.866	21.62-0.865
Zoughouane						
22.60-0.733	22.78-0.731	23.18-0.770	23.19-0.770	22.96-0.760	23.09-0.758	22.59-0.744
Paris						
18.38-0.740	18.70-0.747	19.12-0.782	19.12-0.782	19.11-0.782	19.02-0.776	18.91-0.771
Aniso						
27.27-0.885	28.17-0.899	28.41-0.908	28.42-0.908	28.23-0.889	28.41-0.908	28.21-0.888
Nucleolus						
17.13-0.724	17.43-0.741	17.62-0.756	17.62-0.756	17.62-0.755	17.61-0.755	17.57-0.751
Cardiac muscle						
22.55-0.724	22.99-0.724	23.35-0.763	23.36-0.763	23.22-0.760	23.36-0.763	23.17-0.731
Lysosomes						
22.06-0.773	22.33-0.783	22.54-0.797	22.54-0.797	22.39-0.791	22.54-0.797	22.38-0.789

Table 6.3: (SNR, SSIM) *restoration results with* σ^2*=64 and a blur operator of size 7×7.*

$\ell_{1,2}$-TV	$\ell_{1,2}$-NLTV	KL-NLTV	JK-NLTV	Hel-NLTV	Chi-NLTV	$I_{0.2}$-NLTV
Tunis						
18.54-0.623	19.05-0.659	19.49-0.698	19.53-0.700	19.30-0.690	19.52-0.700	19.27-0.682
Marseille						
17.98-0.716	18.56-0.745	19.00-0.769	19.03-0.770	18.99-0.769	19.02-0.770	18.96-0.768
Zoughouane						
19.95-0.502	20.36-0.534	20.71-0.579	20.74-0.581	20.61-0.565	20.73-0.581	20.64-0.572
Paris						
15.29-0.467	15.70-0.504	15.99-0.546	16.01-0.548	15.98-0.544	16.00-0.548	15.92-0.541
Aniso						
24.49-0.826	25.52-0.850	25.79-0.860	25.90-0.863	25.73-0.859	25.89-0.863	25.80-0.868
Nucleolus						
14.51-0.451	14.88-0.494	15.14-0.527	15.20-0.534	15.10-0.522	15.14-0.528	15.15-0.528
Cardiac muscle						
19.76-0.529	20.45-0.562	20.81-0.604	20.84-0.605	20.64-0.595	20.83-0.605	20.78-0.599
Lysosomes						
18.83-0.498	19.47-0.558	19.84-0.599	19.86-0.600	19.69-0.582	19.85-0.601	19.71-0.585

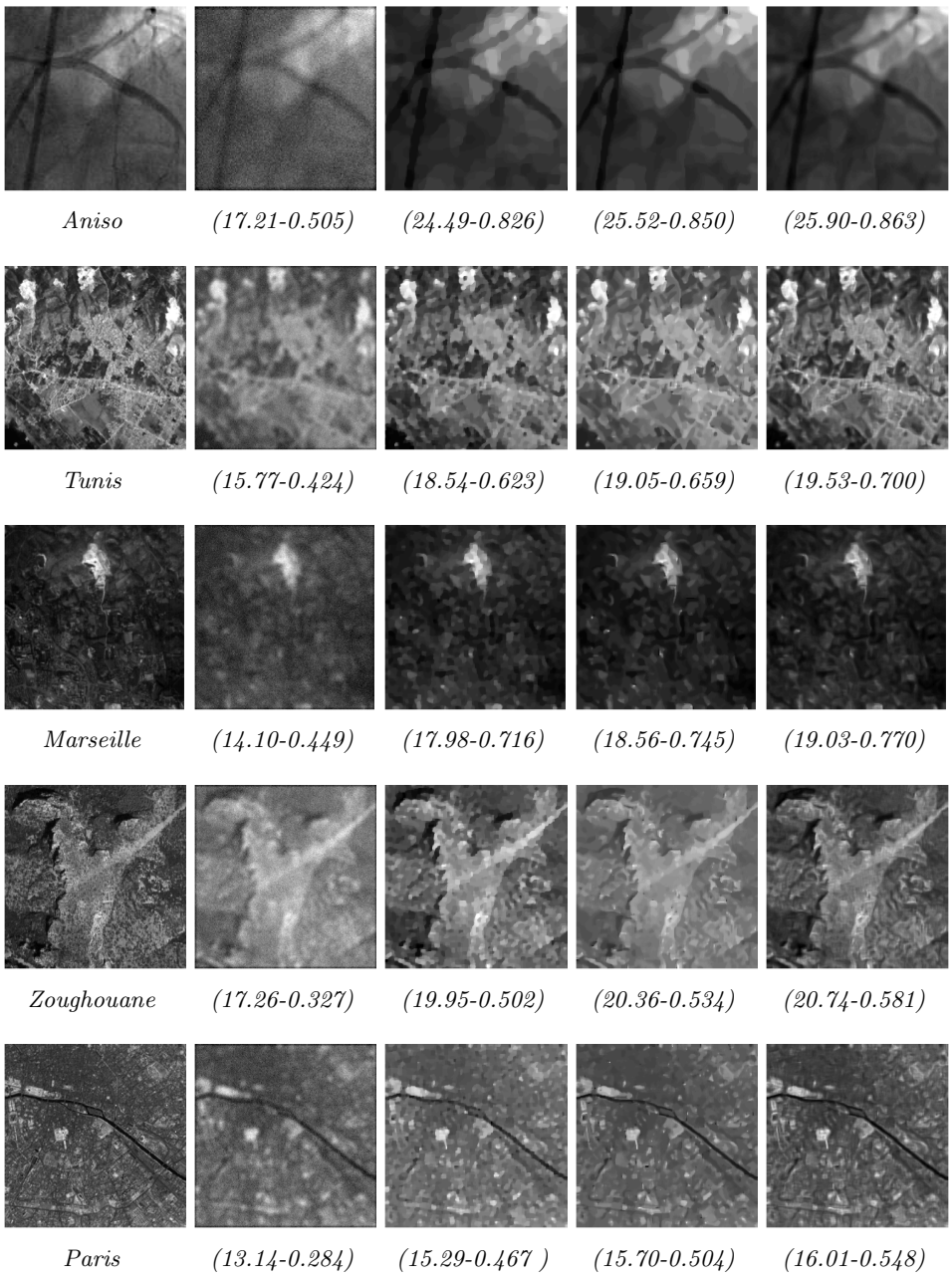


Figure 6.3: *Angiographic + Satellite images: comparison of restoration results (SNR-SSIM) with σ^2=64 and a blur operator of size 7×7. From the Left to the Right: Original image, degraded image, ℓ_{12}-TV result, ℓ_{12}-NLTV result and JK-NLTV result.*

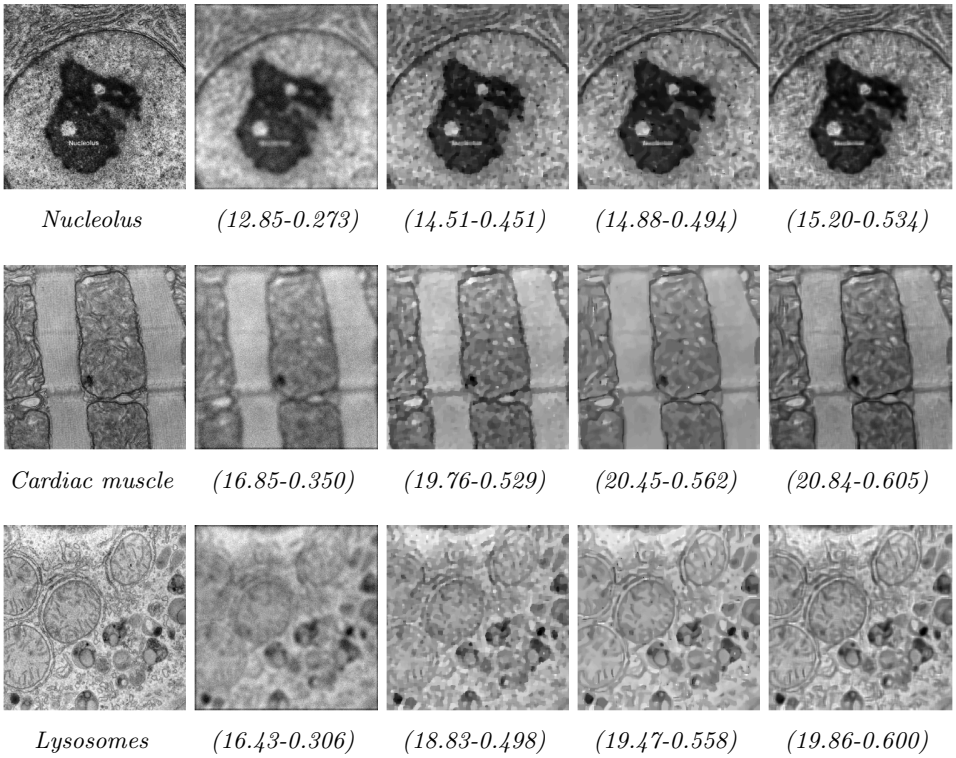


Figure 6.4: *Medical images: comparison of restoration results (SNR-SSIM) with σ^2=64 and a blur operator of size 7×7. From the Left to the Right: Original image, degraded image, ℓ_{12}-TV result, ℓ_{12}-NLTV result and JK-NLTV result.*

Recall that the considered φ-divergences are additive information measures of the form given by Equation (5.8).

Let us now perform the second-order Taylor expansion of the function Φ around an initial estimate $(\bar{v}, \bar{\xi}) \in \mathbb{R}^2$ as follows:

$$\begin{aligned}\Phi(v,\xi) \simeq\ & \Phi(\bar{v},\bar{\xi}) + (v-\bar{v})\frac{\partial\Phi}{\partial v}(\bar{v},\bar{\xi}) + (\xi-\bar{\xi})\frac{\partial\Phi}{\partial \xi}(\bar{v},\bar{\xi}) \\ &+ \frac{1}{2}(v-\bar{v})^2\frac{\partial^2\Phi}{\partial v^2}(\bar{v},\bar{\xi}) + \frac{1}{2}(\xi-\bar{\xi})^2\frac{\partial^2\Phi}{\partial \xi^2}(\bar{v},\bar{\xi}) \\ &+ (v-\bar{v})(\xi-\bar{\xi})\frac{\partial^2\Phi}{\partial v\partial \xi}(\bar{v},\bar{\xi}), \end{aligned} \tag{6.13}$$

where

$$\begin{aligned}\frac{\partial\Phi}{\partial v}(v,\xi) &= \varphi'\big(\frac{v}{\xi}\big) \\ \frac{\partial\Phi}{\partial \xi}(v,\xi) &= \varphi\big(\frac{v}{\xi}\big) - \frac{v}{\xi}\varphi'\big(\frac{v}{\xi}\big)\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2\Phi}{\partial\upsilon^2}(\upsilon,\xi) &= \frac{1}{\xi}\varphi''\big(\frac{\upsilon}{\xi}\big)\\
\frac{\partial^2\Phi}{\partial\xi^2}(\upsilon,\xi) &= \frac{\upsilon^2}{\xi^3}\varphi''\big(\frac{\upsilon}{\xi}\big)\\
\frac{\partial^2\Phi}{\partial\upsilon\partial\xi}(\upsilon,\xi) &= -\frac{\upsilon}{\xi^2}\varphi''\big(\frac{\upsilon}{\xi}\big).
\end{aligned} \tag{6.14}$$

Since φ is such that

$$\varphi(1) = \varphi'(1) = 0, \tag{6.15}$$

we can deduce that when $\bar{\xi} \to \bar{\upsilon}$, the equations in (6.14) boil down to

$$\begin{aligned}
&\frac{\partial\Phi}{\partial\upsilon}(\bar{\upsilon},\bar{\xi}) = 0, && \frac{\partial\Phi}{\partial\xi}(\bar{\upsilon},\bar{\xi}) = 0,\\
&\frac{\partial^2\Phi}{\partial\upsilon^2}(\bar{\upsilon},\bar{\xi}) = \frac{1}{\bar{\upsilon}}\varphi''(1), && \frac{\partial^2\Phi}{\partial\xi^2}(\bar{\upsilon},\bar{\xi}) = \frac{1}{\bar{\upsilon}}\varphi''(1),\\
&\frac{\partial^2\Phi}{\partial\upsilon\partial\xi}(\bar{\upsilon},\bar{\xi}) = \frac{-1}{\bar{\upsilon}}\varphi''(1).
\end{aligned} \tag{6.16}$$

Thus (6.13) reads

$$\begin{aligned}
\Phi(\upsilon,\xi) &\simeq \frac{1}{2\bar{\upsilon}}\varphi''(1)\big((\upsilon-\bar{\upsilon})^2 + (\xi-\bar{\xi})^2 - 2(\upsilon-\bar{\upsilon})(\xi-\bar{\xi})\big)\\
&\simeq \frac{1}{2\bar{\upsilon}}\varphi''(1)(\upsilon-\xi)^2,
\end{aligned} \tag{6.17}$$

where $\varphi''(1) > 0$ since φ is strictly convex.

So, as we can notice, divergences perform similarly to the squared ℓ_2-norm when the two arguments are close. The latter computation is validated in the same context of image restoration using the NLTV regularization. Tables 6.4, 6.5, and 6.6 show that ℓ_2-NLTV leads to close performance to the φ-divergences ("Best of Div." designates the best result obtained in restoration using φ-divergences, i.e. Tables 6.1, 6.2, and 6.3).

Table 6.4: (SNR, SSIM) *denoising results with* σ^2*=400.*

	$\ell_{1,2}$-NLTV	l2-NLTV	Best Of Div
Tunis	21.56-0.808	21.72-0.824	21.73-0.823
Marseille	19.12-0.784	19.67-0.808	19.68-0.807
Zoughouane	22.19-0.715	22.61-0.745	22.63-0.749
Paris	18.38-0.738	18.54-0.763	18.67-0.767
Aniso	25.11-0.837	26.35-0.873	26.51-0.881
Nucleolus	18.85-0.817	18.89-0.817	18.96-0.825
Cardiac	22.56-0.708	22.87-0.755	22.97-0.759
Lysosomes	22.48-0.800	22.51-0.810	22.56-0.810

Table 6.5: (SNR, SSIM) *restoration results with* σ^2*=64 and a blur operator of size 3×3.*

	$\ell_{1,2}$-NLTV	l2-NLTV	Best Of Div
Tunis	22.70-0.851	22.97-0.867	23.12-0.871
Marseille	21.19-0.852	21.67-0.866	21.94-0.873
Zoughouane	22.78-0.731	23.09-0.762	23.19-0.770
Paris	18.70-0.747	18.89-0.772	19.12-0.782
Aniso	28.17-0.899	28.33-0.907	28.42-0.908
Nucleolus	17.43-0.741	17.61-0.754	17.62-0.756
Cardiac	22.99-0.724	23.35-0.755	23.36-0.763
Lysosomes	22.33-0.783	22.54-0.797	22.54-0.797

Table 6.6: (SNR, SSIM) *restoration results with* σ^2*=64 and a blur operator of size* 7×7*.*

	$\ell_{1,2}$-NLTV	l2-NLTV	Best Of Div
Tunis	19.05-0.659	19.47-0.698	19.53-0.700
Marseille	18.56-0.745	19.00-0.770	19.03-0.770
Zoughouane	20.36-0.534	20.72-0.578	20.74-0.581
Paris	15.70-0.504	15.93-0.543	16.01-0.548
Aniso	25.52-0.850	25.87-0.864	25.90-0.863
Nucleolus	14.88-0.494	15.12-0.527	15.15-0.528
Cardiac	20.45-0.562	20.84-0.599	20.84-0.605
Lysosomes	19.47-0.558	19.83-0.599	19.86-0.600

§ 6.2 Data fidelity based on φ-divergences

6.2.1 Image registration

The objective of image registration is to determine spatial transformations that maximize a similarity metric between two images resulting from two different acquisitions. Let the two original images be represented by data vectors $I_1 \in \mathbb{R}^P$ and $I_2 \in \mathbb{R}^P$. For every $j \in \{1, 2\}$, let F_j be an image mapping operator such that

$$\begin{aligned} F_j\colon \quad & \mathbb{R}^P \times \mathbb{R}^{N_j} \to \mathbb{R}^P \\ & (I_j, z_j) \quad \mapsto F_j(I_j, z_j) \end{aligned} \tag{6.18}$$

where $F_j(I_j, z_j)$ is the mapped image, and $z_j \in \mathbb{R}^{N_j}$ is a vector of mapping parameters He et al. [2003]; Hamza and Krim [2003]; Zitovà and Flusser [2003]; Malviya and Bhirud [2009].
When a Kullback metric is adopted, the image registration criterion to be minimized w.r.t. $x = \left[z_1^\top,\ z_2^\top\right]^\top$ reads

$$D\big(F_1(I_1, z_1), F_2(I_2, z_2)\big).$$

To determine the optimal parameter vector $x \in \mathbb{R}^N$ with $N = N_1 + N_2$, one can proceed by supposing that F_1(resp. F_2) is differentiable and by performing a first-order Taylor expansion around an initial estimate $\bar{z}_1$(resp. $\bar{z}_2$) as follows: $(\forall j \in \{1, 2\})$

$$F_j(I_j, z_j) \simeq F_j(I_j, \bar{z}_j) + \frac{\partial F_j}{\partial z_j}(I_j, \bar{z}_j)(z_j - \bar{z}_j) \tag{6.19}$$

where $\frac{\partial F_j}{\partial z_j}(I_j, \bar{z}_j)$ is a Jacobian matrix. With the linearization expressed in (6.19), the image registration problem can be reformulated under the form of Problem 5.1 (Chapter 5), where

$$A = \left[\frac{\partial F_1}{\partial z_1}(I_1, \bar{z}_1) \;\; \mathbf{0}\right], \tag{6.20}$$

$$B = \left[\mathbf{0} \quad \frac{\partial F_2}{\partial z_2}(I_2, \bar{z}_2)\right], \tag{6.21}$$

and

$$u = F_1(I_1, \bar{z}_1) - \bar{z}_1 \frac{\partial F_1}{\partial z_1}(I_1, \bar{z}_1), \tag{6.22}$$

$$v = F_2(I_2, \bar{z}_2) - \bar{z}_2 \frac{\partial F_2}{\partial z_2}(I_2, \bar{z}_2). \tag{6.23}$$

The additional regularization terms $(R_s \circ T_s)_{1 \leq s \leq S}$ allow the incorporation of prior knowledge about the sought parameter vector x.

6.2.2 Disparity estimation under illumination variation

We conducted experiments for disparity estimation under illumination variation, using a stereoscopic pair of grayscale images I_1 and I_2 with size $P_1 \times P_2$ ($P = P_1 P_2$). The related image mapping operators are given by

$$F_1(I_1, z_1) = \mathrm{vec}\left[(z_1^{(i_1,i_2)} I_1^{(i_1,i_2)})_{\substack{1 \leq i_1 \leq P_1 \\ 1 \leq i_2 \leq P_2}}\right] \tag{6.24}$$

$$F_2(I_2, z_2) = \mathrm{vec}\left[(I_2^{(i_1 - z_2^{(i_1,i_2)}, i_2)})_{\substack{1 \leq i_1 \leq P_1 \\ 1 \leq i_2 \leq P_2}}\right] \tag{6.25}$$

where $I_1^{(i_1,i_2)}$ (resp. $I_2^{(i_1,i_2)}$) is the intensity value in the left (resp. right) view at pixel (i_1, i_2), $z_1^{(i_1,i_2)}$ the illumination variation coefficient at (i_1, i_2), and $z_2^{(i_1,i_2)}$ the associated disparity. The parameter vectors are here

$$z_1 = \mathrm{vec}\left[(z_1^{(i_1,i_2)})_{1 \leq i_1 \leq P_1, 1 \leq i_2 \leq P_2}\right]$$

and

$$z_2 = \mathrm{vec}\left[(z_2^{(i_1,i_2)})_{1 \leq i_1 \leq P_1, 1 \leq i_2 \leq P_2}\right].$$

It can be observed that

$$\frac{\partial F_1}{\partial z_1}(I_1, \bar{z}_1) = \mathrm{Diag}\left[(I_1^{(i_1,i_2)})_{\substack{1 \leq i_1 \leq P_1 \\ 1 \leq i_2 \leq P_2}}\right] \tag{6.26}$$

$$\frac{\partial F_2}{\partial z_2}(I_2, \bar{z}_2) = -\,\mathrm{Diag}\left[(\nabla^{(1)} I_2^{(i_1 - z_2^{(i_1,i_2)}, i_2)})_{\substack{1 \leq i_1 \leq P_1 \\ 1 \leq i_2 \leq P_2}}\right] \tag{6.27}$$

where $\nabla^{(1)}$ is the gradient operator w.r.t. the first spatial coordinate.
It is useful to incorporate prior information about the solution so as to

convert the original matching problem to a well-posed one. Similarly to what we have presented in Chapter 3, a first constraint set is introduced that takes into account the range of possible values for the disparity and the illumination variation:

$$C_1 = \big\{[z_1^\top, z_2^\top]^\top \in \mathbb{R}^{2P} \mid (z_1)_{\min} \leq z_1 \leq (z_1)_{\max}, \\ (z_2)_{\min} \leq z_2 \leq (z_2)_{\max}\big\}. \tag{6.28}$$

A second constraint is employed to enforce the smoothness of the estimated fields:

$$C_2 = \{[z_1^\top, z_2^\top]^\top \in \mathbb{R}^{2P} \mid \|\nabla z_1\|_{\ell_2}^2 \leq \kappa_1,\ \mathrm{TV}(z_2) \leq \kappa_2\} \tag{6.29}$$

where $\|\cdot\|_{\ell_2}$ is the ℓ_2 norm, ∇ is the spatial gradient operator and TV is the discrete total variation semi-norm. Consequently, we have now $S = 2$ in Problem 5.1 (previous chapter), where R_1 and R_2 reduce to indicator functions of convex sets, $T_1 = \mathbf{I}_{2P}$, and T_2 is a block diagonal matrix with diagonal terms set to gradient operators.

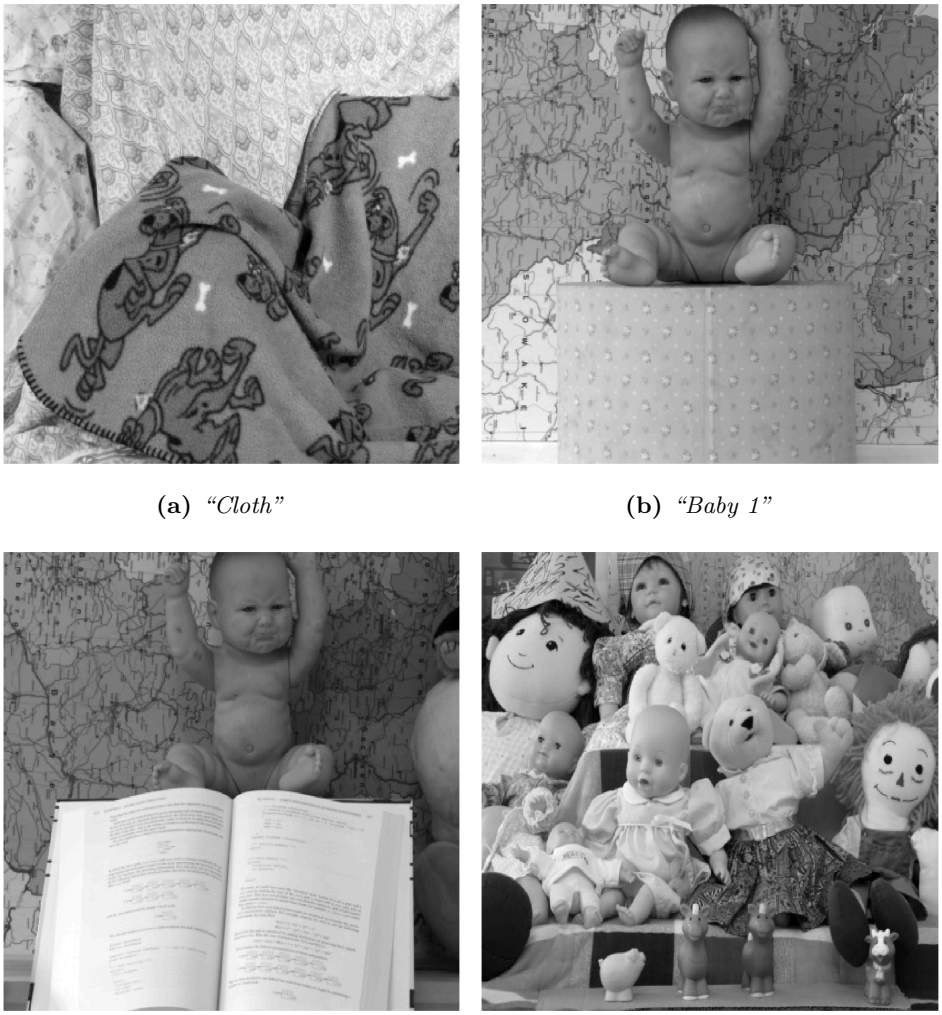
6.2.3 Experiments

We now illustrate the practical performance of our method on the "Cloth", "Baby 1", "Baby 2", and "Dolls" stereo pairs (Figure 6.5) downloaded from the MiddleBury website.[3]

The results are provided in Table 6.7 and Figure 6.6. The quality of the results was evaluated using three different metrics based on a ground truth: the MAE (Mean Absolute Error) evaluated between the computed maps $(z_j)_{j\in\{1,2\}}$ and the ground truth $(\widetilde{z}_j)_{j\in\{1,2\}}$, the percentage of bad matching pixels in the disparity map

$$\mathrm{Err}_T = \frac{1}{P}\sum_{i_1=1}^{P_1}\sum_{i_2=1}^{P_2} \mathbf{1}(|z_2^{(i_1,i_2)} - \widetilde{z}_2^{(i_1,i_2)})| > T),$$

[3] http://vision.middlebury.edu/stereo/



(a) *"Cloth"* **(b)** *"Baby 1"*

(c) *"Baby 2"* **(d)** *"Dolls"*

Figure 6.5: *Left images of the stereo pairs used in our simulations.*

and the MS-SSIM (multi-scale structure similarity index) Malpica and Bovik [2009]. According to the percentage of bad pixels and the Mean Absolute Error, we can see that the JK divergence performs better than the other error measures.

Finally, in Figure 6.6, we illustrate the recovered disparity maps using JK and KL divergences for the gray stereo pair "Cloth". These visual results show the efficiency of the JK measure in recovering the depth estimation in the presence of the illumination variations.

Table 6.7: *Disparity estimation results* $(\mathrm{MAE}, \mathrm{Err}_2)$ *obtained for gray level images.*

	Cloth	Baby 1	Baby 2	Dolls
ℓ_2	(0.8385, 3.62)	(0.9706, 5.80)	(1.8336, 12.18)	(1.2591, 14.41)
KL	(0.8287, 3.36)	(0.9301, 5.40)	(1.6539, 11.56)	(1.2107, 13.86)
JK	(0.8348, 3.44)	(0.9298, 5.32)	(1.6503, 11.48)	(1.2157, 13.36)
Hel	(0.9057, 3.46)	(0.9333, 5.57)	(1.7631, 11.71)	(1.2107, 13.87)
Chi	(0.8335, 3.43)	(0.9328, 5.56)	(1.7636, 11.71)	(1.2127, 13.81)
$\mathrm{I}_{0.2}$	(0.9058, 3.39)	(0.9335, 5.55)	(1.7741, 11.64)	(1.2167, 13.83)

§ 6.3 Conclusion

In this chapter, we have applied φ-divergences as a measures of dissimilarity in two different contexts:

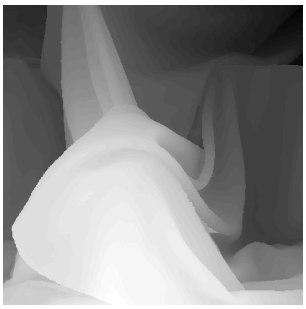
- We have extended the NLTV framework by using information divergences to build new sparsity regularization measures for signal recovery.
- We have used the divergence as a data fidelity term for image registration, more precisely for disparity estimation under illumination variation.

However, as we have shown before, the performance achieved by different types of φ-divergence is similar. This behaviour may be explained by the fact that the values of these two-variable functions have the same second-order Taylor expansion when the two arguments are close to each others.

The practical results obtained by implementing the proposed φ-divergence measures in an image processing context do not yield the expected performance gains with respect to standard measures such as the squared ℓ_2-norm. However, it would be interesting to investigate the effectiveness of divergences in problems involving probability distribution. Indeed, it is our belief that these measures could constitute a successful element for other applications and that they are undoubtedly better adapted to some other problems.

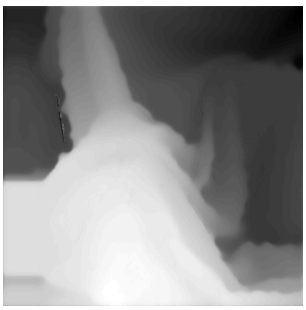
a) Left image

b) Right image



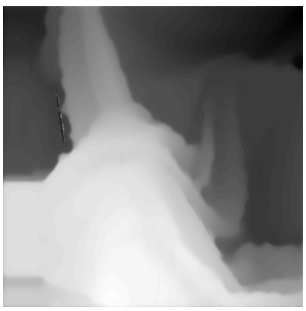
c) True disparity

d) True illumination



e) MAE= 0.8348, Err_2=3.44
MS − SSIM= 0.9856

f) MAE=0.0332
MS − SSIM=0.0275



g) MAE= 0.8287, Err_2= 3.36
MS − SSIM=0.9857

h) MAE=0.0330
MS − SSIM=0.0271

Figure 6.6: *Results for "Cloth" stereo pair: a)-b) stereo images, c)-d) ground truths, e)-f) KL divergence and g)-h) JK divergence. Parameters:* $((z_1)_{\min}, (z_1)_{\max}) = (20, 75)$, $((z_2)_{\min}, (z_2)_{\max}) = (0.1, 1.1)$, $(\kappa_1, \kappa_2) = (74000, 400)$.

- Chapter 7 -

Conclusion

"You may not end up where you thought you'd be, but you'll end up right where you're meant to be."

In the first part of this thesis, we have investigated the problem of disparity estimation within the framework of convex optimization and extended it to the case of multiple stereo views. In the second part, we have provided novel tools to deal with convex divergence measures in optimization problems and we have applied them to image recovery problems.

§ 7.1 Contributions

Disparity estimation under illumination variation We have proposed a convex optimization approach to deal with disparity estimation under illumination variation. Our solution consists of modeling the spatially-varying illumination variation by a multiplicative term. Since the global formulation is nonconvex, the problem is addressed by solving a sequence of convex relaxations.

- We have employed a relaxation based on a first-order Taylor approximation around an initial estimate to derive a convex energy function. This energy is then minimized subject to various convex constraints aiming at imposing some regularity on the solution. The originality in our approach lies in the ability to deal with non-strictly convex energies, which allows us to employ data fidelity terms other than quadratic ones.
- The problem of jointly estimating the disparity and the illumination variation is tackled in a set theoretic framework and solved using proximal tools. The original contributions of our work consist of the ability to consider multicomponent images in the

presence of illumination variation, where each component corresponds to one of the channels of a specific color system (RGB, YUV, ...).

The proposed approach is also able to exploit the potentials offered by multicore/GPU parallel system architectures. The experiments indicate that this approach can effectively deal with the local illumination variation and yields better results compared with other approaches.

Disparity map sequence We have extended the previous approach to multi-view problems. Rather than estimating a single depth map, we have associated a depth map with each input image (or a subset of them). We have proposed two different approaches to estimate the depth map for each chosen reference frame, which differ in the type of convex relaxation employed. The first approach employs a relaxation based on multi-labeling, with the possibility of using nonconvex similarity measures. So doing, there is no more dependence with respect to the initial estimate, at the cost of an increased complexity. The second approach employs a relaxation based on an improved first-order Taylor approximation around an initial estimate. The experimental results illustrate, through a complex specular scene example, the effectiveness of our proposals.

φ-divergence proximity operator We have shown how to deal with convex optimization problems involving φ-divergences, when the minimization is jointly performed with respect to both divergence arguments. More specifically, we have derived new expressions for the proximity operator of several φ-divergences (e.g. Kullback-Leibler, Jeffreys-Kullback, Hellinger, Chi-Square, I_α, and Renyi divergences), enriching the list of functions for which such operators can be easily computed. This has allowed us to address a generic class of convex optimization problems by means of efficient proximal algorithms, overcoming the limitations of current state-of-the-art approaches for such problems.

Applications for φ-divergence proximity operator We have investigated the use of divergences in two different image processing applications: restoration and registration. Firstly, we have extended the NLTV framework by using some information divergences to build new regularization terms for signal recovery. This regularity measure allows us to better exploit image features by making use of the relevant information along an edge or a regular texture pattern. Secondly, we have presented an image registration application to illustrate the use of divergences as data fidelity terms, where our objective is to determine spatial transformations that maximize a similarity metric between two images resulting from two different acquisitions.

Epigraphical projection Epigraphical projections have a great interest when dealing with constraint sets for which a closed form expression of the projection does not exist. Within this context, we have provided theoretical guidelines to show how the forms of the proximity operator of divergences can be employed to compute the expression of the projection onto the epigraph of a number of convex functions.

§ 7.2 Perspectives

Extension to View synthesis application: The problem of estimating the disparity between two stereo images (see Chapter 3) can be extended to view synthesis. Because only a few cameras are available, additional views have to be synthesized from them. The main idea is to simultaneously estimate a new view and the corresponding disparity map. To do so, we need to synthesize a virtual depth map from the current view by applying a warping technique. We do not need any other side information to synthesize the virtual depth maps except camera parameters. In particular, it could be a good idea to investigate new ways of estimating the illumination variation for the new view synthesis by studying how the light could change from one image to another in multi-view datasets. This could be applied later in 3D tracking, stereo video coding, 3D scene interpretation, and 3D television.

Exploiting the dependence among the disparity sequence maps: The multi-view methods presented in Chapter 4 can be improved in several ways. One of the most appealing ideas is to combine the disparity sequence estimation in a single optimization problem in order to exploit the huge correlation between the disparities of different views. In this case, an additional piece of prior information may be included into the minimization problem: the 3D point projected in different views should have the same disparity value (since cameras are rectified and equidistant). The main challenges in this context are the occlusion areas and the non-convexity of the considered constraint, which needs to be relaxed using one of the techniques presented in Chapter 6.

Combining the discrete and continuous methods: There is also room for improving the potential of the multi-label approach for the disparity estimation sequence. The images of a scene under varying illuminations and from different viewpoints are highly interrelated, which make it possible to predict the object depth and the illumination variation. Illumination due to imperfectly calibrated cameras, different perspective projection directions, and different reflection effects varies from an object to the other. We could consider the local illumination variation that arises due to differences in camera positioning. Hence, it would

be interesting to extend the multi-label approach in order to jointly estimate the disparity map and the illumination variation field. This would require combining the multilabel-based relaxation of Chapter 4 with the continuous approach of Chapter 3.

Disparity and Motion from a Multi-view video sequence: Stereo video sequences show the temporal evolution of a scene from two different viewpoints. In order to allow for an accurate stereo sequence analysis, it is essential to consider the spatio-temporal relationships that exist between the different images of the sequences. The stereo-motion consistency constraint can be exploited by jointly estimating the motion field of the left and right sequences, along with the disparity field of the current stereo pair. This approach can be efficiently extended to a multi-view sequence by using the techniques presented in Chapter 4.

φ-divergence in blind deconvolution φ-divergences can be applied in the context of blind deconvolution. Here, the degradation model is the same as in Equation (2.1):

$$z = T\overline{x} + w,$$

but one aims at estimating both the target image $\overline{x} \in \mathbb{R}^N$ and the linear operator $T \in \mathbb{R}^{K\times N}$. The common approach is to resort to an alternating minimization scheme, which however has no guaranteed convergence. An alternative approach consists of assuming that there exists an operator G such that $GT = I$, which allows one to reformulate the degradation model as

$$Gz = \overline{x} + \omega,$$

where $\omega = Gw$ denotes the transformed (and possibly amplified) noise. This leads to a convex optimization problem which aims at minimizing the divergence between Gz and x, while enforcing some kind of regularization on G and x. The presented approach constitutes a good example of minimization problems jointly performed on two variables.

φ-divergence for channel estimation It is worth mentioning that several applications in telecommunication systems could benefit from the proposed optimization approaches, e.g. for jointly estimating the transmitted data and the channel transfer function in the presence of noise. In this context, φ-divergences play a central role in the formulation of the recovery problem. In addition, some constraints can be added to the problem in order to model the temporal evolution of the channel (due to Doppler effect). The resulting optimization problem is nonconvex and the existing methods generally follow an alternating minimization scheme (see e.g. Hu et al. [2008]). An alternative approach would

consist of adopting the algorithms developed in Chapter 5 to address this nonconvex scenario, or by resorting to the convex reformulation provided above for blind deconvolution.

Epigraphical projection in video coding As we have seen in Section 5.5, the proximity operator of the considered φ-divergences allows us to compute the orthogonal projection onto the epigraph of functions listed in Table 5.1. Recently, epigraphical projections have been employed to address convex constraints for which the associated projection operator is not available in closed form Chierchia et al. [2013]. Our results thus substantially enrich the list of functions that can be handled by such an epigraphical approach. For example, the exponential function can be used in the context of video coding, where one needs to solve a rate allocation problem that aims at minimizing the rate budget while keeping the global distortion (expressed as a sum of exponentials composed by a linear operator) below a given bound.

List of Figures

List of Tables

Bibliography

M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.*, 20(3):681–695, Mar. 2011. 13, 92

O. D. Akyildiz and I. Bayram. An analysis prior based decomposition method for audio signals. In *Proc. Eur. Sig. and Image Proc. Conference*, pages 2437–2441, Bucharest, Roumain, Aug. 2012. 10

A. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society. Series B (Methodological)*, 28(1):131–142, 1966. 84, 86, 89

S.-I. Amari. Alpha-divergence is unique, belonging to both f-divergence and bregman divergence classes. *IEEE Trans. on inf. theory*, 55(11):4925–4931, 2009. 84, 86

S. Anthoine, J.F. Aujol, C. Mélot, and Y. Boursier. Some proximal methods for CBCT and PET tomography. *Inverse Problems in Imaging*, 6(4): 565–598, Nov. 2012. 85

S. Arimoto. Information-theoretical considerations on estimation problems. *Info. Contr.*, 19(3):181–194, Oct. 1971. 84

S. Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. on inf. theory*, 18(1):14–20, Jan. 1972. 84, 102

H. Attouch and M. Soueycatt. Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces - applications to games, PDE's and control. *Pac. J. Optim.*, 5(1):17–37, Jan. 2009. 20, 92

G. Aubert and R. Tahraoui. Sur la minimisation d'une fonctionnelle non convexe, non différentiable en dimension 1. *Bolletino UMI*, 5(17B), 1980. 16

J.-F. Aujol. Some first-order algorithms for total variation based image restoration. *J. Math. Imag. Vis.*, 34(3):307–327, Jul. 2009. 14

J. F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *J. Math. Imaging Vision*, 22(1):71–88, Jan. 2005. 38

F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning*, 4(1):1–106, Jan. 2012. 10

A. R. Barron, L. Gyrfi, and E. C. van der Meulen. Distribution estimates consistent in total variation and two types of information divergence. *IEEE Trans. Inf. Theory*, 38(5):1437–1454, Sep. 1990. 84

M. Basseville. Distance measures for signal processing and pattern recognition. *European J. Signal Process.*, 18(4):349–369, 1989. 84, 89

M. Basseville. Divergence measures for statistical data processing. *Signal Process.*, 93(4):621–633, Apr. 2013. 84

H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, Jan. 2011. 25, 88, 89, 110

H. H. Bauschke, P. L. Combettes, and D. Noll. Joint minimization with alternating Bregman proximity operators. *Pac. J. Optim*, 2(3):401–424, Sep. 2006. 83, 85

I. Bayram and M. Kamasak. Directional total variation. *Signal Process. Lett.*, 19(12):781–784, Dec. 2012. 14

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, Mar. 2009. 18

A. Ben-Tal and M. Teboulle. A smoothing technique for nondifferentiable optimization problems. *Lecture Notes in Mathematics*, 1405:1–11, 1989. 16

J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numer. Math.*, 84(3): 375–393, 2000. 86, 90

J.-D. Benamou, Y. Brenier, and K. Guittet. The monge-katonrovich mass transfer and its computational fluid mechanics formulation. *Int. J. Numer. Meth. Fluids*, 40:21–30, 2002. 86, 90

R. Beran. Minimum hellinger distance estimates for parametric models. *Ann. Statist.*, 5(3):445–463, 1977. 85

M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging*. Bristol, U.K.: IOP Publishing, 1998. 10

F. Besse, C. Rother, A. Fitzgibbon, and J. Kautz. PMBP: Patchmatch belief propagation for correspondence field estimation. *Int. J. Comp. Vis.*, pages 132.1–132.11, Aug. 2012. 32, 33

R. Bhotika, D. Fleet, and K. Kutulakos. A probabilistic theory of occupancy and emptiness. In *Computer Vision – ECCV*, volume 2352 of *Lecture Notes in Computer Science*, pages 112–130. Springer Berlin Heidelberg, 2002. 60

R.E. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Trans. on inf. theory*, 18(4):460–473, Jul. 1972. 84, 102

P. Bouboulis, K. Slavakis, and S. Theodoridis. Adaptive learning in complex reproducing kernel Hilbert spaces employing Wirtinger's subgradients. *IEEE Trans Neural Networks*, 23(3):425–438, Mar. 2012. 23

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, Nov. 2011. 32, 33

L. M. Bregman. The method of successive projection for finding a common point of convex sets. *Soviet Math. Dokl.*, 6:688–692, 1965. 17

L. M. Briceño-Arias and P. L. Combettes. A monotone + skew splitting model for composite monotone inclusions in duality. *SIAM J. Opt.*, 21(4): 1230–1250, Oct. 2011. 18, 34, 67, 93

A. Broadhurst, T. W. Drummond, and R. Cipolla. A probabilistic framework for space carving. In *Proc. IEEE Int. Conf. Comput. Vis.*, volume 1, pages 388–393, Vancouver, BC, Jul. 2001. 60

A. Buades, B. Coll, and J.-M. Morel. The staircasing effect in neighborhood filters and its solution. *IEEE Trans. Image Process.*, 15(6):1499–1505, Jun. 2006. 15, 116

A. Buades, B. Coll, and J.-M. Morel. Nonlocal image and movie denoising. *Int. J. Comput. Vision*, 76(2):123–139, Feb. 2008. 53

G. Buchsbaum. A spatial processor model for object colour perception. *J. of the Franklin Institute*, 310(1):1–26, Jul. 1980. 33

J. Burbea and C. Rao. On the convexity of some divergence measures based on entropy functions. *IEEE Trans. on inf. theory*, 28(3):489–495, 1982. 84

C. L. Byrne. Iterative image reconstruction algorithms based on cross-entropy minimization. *IEEE Trans. Image Process.*, 2(1):96 –103, Jan. 1993. 84

J.-F. Cai, B. Dong, S. Osher, and Z. Shen. Image restoration: Total variation, wavelet frames and beyond. *J. Amer. Math. Soc.*, 25(4):1033–1089, May 2012. 14

M. Carlavan and L. Blanc-Féraud. Sparse poisson noisy image deblurring. *IEEE Trans. Image Process.*, 21(4):1834–1846, Apr. 2012. 85

V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, pages 694–699, Cambridge, MA, Jun. 1995. 60

V. Caselles, R. Kimmel, G. Sapiro, and C. Sbert. Minimal surfaces: A three dimensional segmentation approach. *Numerische Mathematik*, 77(4): 423–451, Oct. 1997. 60

Y. Censor, W. Chen, P. L. Combettes, R. Davidi, and G. T. Herman. On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints. *Comput. Optim. Appl.*, 51(3):1065–1088, Apr. 2012. 17

A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vis.*, 40(1):120–145, May 2011. 18, 34, 67, 93

A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.*, 7:319–335, Mar. 1998. 38

S. Chan, T. Esedoglu and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.*, 66(5):1632–1648, 2006. 13

L. Chang-Hwan. A new measure of rule importance using hellinger divergence. In *Int. Conf. on Data Analytics*, pages 103–106, Barcelona, Spain, Sep. 2012. 85

C. Chaux, L. Duval, and J.-C. Pesquet. Image analysis using a dual-tree m-band wavelet transform. *IEEE Trans. Image Process.*, 15(8):2397–2412, Aug. 2006. 38

C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame-based inverse problems. *Inverse Problems*, 23(4): 1495–1518, Jun. 2007. 17, 18, 86

G. Chen and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Math. Program.*, 64(1–3):81–101, Mar. 1994. 18, 93

H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *Ann. Statist.*, 23:493–507, 1952. 86

M. Chiang and S. Boyd. Geometric programming duals of channel capacity and rate distortion. *IEEE Trans. on inf. theory*, 50(2):245 – 258, Feb. 2004. 84

G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu. Epigraphical splitting for solving constrained convex optimization problems with proximal tools. 2013. http://arxiv.org/abs/1210.5844. 13, 14, 15, 24, 53, 54, 86, 90, 109, 116, 118, 137

E. Chouzenoux, J. Idier, and S. Moussaoui. A Majorize-Minimize strategy for subspace optimization applied to image restoration. *IEEE Trans. Image Process.*, 20(6):1517–1528, Jun. 2011. 16

S. Chrétien and A. O. Hero. On EM algorithms and their proximal generalizations. *Control Optim. Calc. Var.*, 12:308–326, Jan. 2008. 85

R. Ciak, B. Shafei, and G. Steidl. Homogeneous penalizers and constraints in convex image restoration. *J. Math. Imag. Vis.*, 47(3):210–230, Nov. 2012. 13

A. Cichocki, L. Lee, Y. Kim, and S. Choi. Non-negative matrix factorization with α-divergence. *Pattern Recogn. Lett.*, 29(9):1433–1440, 2008. 86

D. Cieslak, T. Hoens, N. Chawla, and W. Kegelmeyer. Hellinger distance decision trees are robust and skew-insensitive. *Data Mining and Knowledge Discovery*, 24(1):136–158, 2012. 85

R.T. Collins. A space-sweep approach to true multi-image matching. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, pages 358–363, San Francisco, CA, Jun. 1996. 60

P. L. Combettes. The foundations of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208, Feb. 1993. 17

P. L. Combettes. Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections. *IEEE Trans. Image Process.*, 6(4):492–506, Apr. 1997. 17

P. L. Combettes. A block-iterative surrogate constraint splitting method for quadratic signal recovery. *IEEE Trans. Signal Process.*, 51(7):1771–1782, Jul. 2003. 23

P. L. Combettes and J.-C. Pesquet. Image restoration subject to a total variation constraint. *IEEE Trans. Image Process.*, 13(9):1213–1222, Sep. 2004. 14, 38

P. L. Combettes and J.-C. Pesquet. Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM J. Optim.*, 18(4):1351–1376, Nov. 2007a. 100

P. L. Combettes and J.-C. Pesquet. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE J. Selected Topics Signal Process.*, 1(4):564–574, Dec. 2007b. 18, 19

P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6):x+27, Dec. 2008. 14, 19, 38, 86, 90, 92

P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, New York, 2011. 17, 18, 41, 67, 85, 86, 90, 98

P. L. Combettes and J.-C. Pesquet. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.*, 20(2):307–330, Jun. 2012. 18, 22, 34, 67, 68, 93, 94, 118

P. L. Combettes and H. J. Trussell. The use of noise properties in set theoretic estimation. *IEEE Trans. Signal Process.*, 39(7):1630–1641, Jul. 1991. 13, 118

P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. and Simul.*, 4(4):1168–1200, Nov. 2005. 18

L. Condat. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *Journal of Opt. Theory and Applications*, 158(2):460–479, 2013. 18, 93

R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. *Adv. Comput. Math.*, 5(1):329–359, 1996. 100

C. Couprie, L. Grady, L. Najman, J.-C. Pesquet, and H. Talbot. Dual constrained TV-based regularization on graphs. *SIAM J. on Imaging Sciences*, 6(3):1246–1273, Oct. 2013. 14, 18, 53, 116

T. M. Cover and J. A. Thomas. *Elements of information theory*. New York, USA: Wiley-Interscience, 1991. 84, 85

I. J. Cox, S. Roy, and S. L. Hingorani. Dynamic histogram warping of image pairs for constant image brightness. In *Proc. Int. Conf. Image Process.*, volume 2, pages 366–369, Washington, DC, Oct. 1995. 34

D. Cremers, T. Pock, K. Kolev, and A. Chambolle. Convex relaxation techniques for segmentation, stereo and multiview reconstruction. In *Markov Rand. Fields for Vision and Image Processing*. MIT Press, 2011. 32, 56, 64

I. Csiszár. Eine informations theoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten. *Magyar Tud. Akad. Mat. Kutató Int. Köozl.*, 8:85–108, 1963. 84, 89

I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967. 89

I. Csiszár. Information measures: A critical survey. *IEEE Trans. on inf. theory*, A:73–86, 1974. 86

I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Applied Math.*, 57(11):1413–1457, Nov. 2004. 18

J. E. Davis, R. Yang, and L. Wang. BRDF invariant stereo using light transport constancy. In *Proc. IEEE Int. Conf. Comput. Vis.*, volume 1, pages 436–443, Beijing, China, Oct. 2005. 34

R. Deriche, P. Kornprobst, and G. Aubert. Optical-flow estimation while preserving its discontinuities: A variational approach. In S. Li, D. Mital, E. Teoh, and H. Wang, editors, *Recent Developments in Computer Vision*, volume 1035 of *Lecture Notes in Computer Science*, pages 69–80. Springer Berlin / Heidelberg, 1996. 10, 32, 33

P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice-Hall International, 1982. 85

Y. Duan, L. Yang, H. Qin, and D. Samaras. Shape reconstruction from 3D and 2D data using PDE-based deformable surfaces. In *Computer Vision – ECCV*, volume 3023 of *Lecture Notes in Computer Science*, pages 238–251. Springer Berlin Heidelberg, 2004. 60

J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Effcient projections onto the l1-ball for learning in high dimensions. In *Proc. of Intl. Conf. on Machine Learning*, pages 272–279, Helsinki, Finland, Jul. 2008. 25

F.-X. Dupé, M. J. Fadili, and J.-L. Starck. A proximal iteration for deconvolving Poisson noisy images using sparse representations. *IEEE Trans. Image Process.*, 18(2):310–321, Feb. 2009. 10, 84, 85, 102

J. Eckstein. Parallel alternating direction multiplier decomposition of convex programs. *J. Optim. Theory Appl.*, 80(1):39–62, Jan. 1994. 20

I. Ekeland and R. Téman. *Convex analysis and variational problems*. SIAM, Philadelphia, PA, USA, 1999. 65

Y. C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. on inf. theory*, 55(11):5302–5316, Nov. 2009. 41

E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.*, 3(4):1015–1046, Dec. 2010. 18, 93

G. Facciolo, A. Almansa, J.-F. Aujol, and V. Caselles. Irregular to regular sampling, denoising, and deconvolution. *Multiscale Model. and Simul.*, 7 (4):1574–1608, Avr. 2009. 10

J. M. Fadili and G. Peyré. Total variation projection with first order schemes. *IEEE Trans. Image Process.*, 20(3):657–669, Mar. 2011. 14, 25

O. Faugeras and R. Keriven. Variational principles, surface evolution, pdes, level set methods, and the stereo problem. *IEEE Trans. Image Process.*, 7 (3):336–344, 1998. 60

J.A. Fessler. Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans. *IEEE Trans. Image Process.*, 4(10):1439 –1450, Oct. 1995. 84

M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J. Selected Topics Signal Process.: Special Issue on Convex Optimization Methods for Signal Processing*, 5981(4):586–598, Dec. 2007. 18

M. A. T. Figueiredo and R. D. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.*, 12(8):906–916, Aug. 2003. 34

G. D. Finlayson, B. Schiele, and J. L. Crowley. Comprehensive colour image normalization. In *Proc. European Conference on Computer Vision*, volume 1406, pages 475–490, Island of Rhodes, Greece, Jun. 1998. 33

A. Foi and G. Boracchi. Foveated self-similarity in nonlocal image filtering. In *Proc. SPIE Electronic Imaging 2012, Human Vision and Electronic Imaging XVII*, volume 8291, Burlingame (CA), USA, Jan. 2012. 54, 121

M. Fornasier and C.-B. Schönlieb. Subspace correction methods for total variation and ℓ_1-minimization. *IMA J. Numer. Anal.*, 47(8):3397–3428, 2009. 18

M. Fortin and R. Glowinski, editors. *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*. Elsevier Science Ltd, Amsterdam: North-Holland, 1983. 20

Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Int.*, 32(8):1362–1376, 2010. 60

A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Mach. Vis. Appl.*, 12(1):16–22, Jul. 2000. 35

R. Gaetano, G. Chierchia, and B. Pesquet-Popescu. Parallel implementations of a disparity estimation algorithm based on a proximal splitting method. In *Visual Communications and Image Process.*, pages 1–6, San Diego, CA, Nov. 2012. 47

D. Gallup, J.-M. Frahm, P. Mordohai, and M. Pollefeys. Variable baseline/resolution stereo. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, pages 1–8, Anchorage, AK, Jun. 2008. 61

P. Gargallo and P. Sturm. Bayesian 3D modeling from images using multiple depth maps. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, volume 2, pages 885–891, San Diego, Jun. 2005. 60, 61

P. Gargallo, P. Sturm, and S. Pujades. An occupancy-depth generative model of multi-view images. *Comp. Vision ACCV*, 4844:373–383, 2007. 60

C. F. Gauss. *Theory of the Combination of Observations Least Subject to Errors*. Philadelphia, SIAM Publishing, 1820. 10

M.A. Gennert. Brightness-based stereo matching. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 139–143, Tampa, FL, Dec. 1988. 34, 35

A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002. 85

G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *Multiscale Model. and Simul.*, 6(2):595–630, 2007. 54

G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Model. and Simul.*, 7(3):1005, Nov. 2009. 14, 15, 53, 116

T. Goldstein and S. Osher. The split bregman method for ℓ_1-regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–343, Apr. 2009. 20

L. G. Gurin, B. T. Polyak, and E. V. Raik. Projection methods for finding a common point of convex sets. *Zh. Vychisl. Mat. Mat. Fiz.*, 7(6):1211–1228, 1967. 17

J. Hadamard. Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902. 11

N. Hajlaoui, C. Chaux, G. Perrin, F. Falzon, and A. Benazza-Benyahia. Satellite image restoration in the context of a spatially varying point spread function. *J. Opt. Soc. Am.*, 27(6):1473–1481, Jun. 2010. 10

A. B. Hamza and H. Krim. Image registration and segmentation by maximizing the Jensen-Rnyi divergence. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 2683 of *Lecture Notes in Computer Science*, pages 147–163. Springer Berlin Heidelberg, 2003. 127

D. Han and D. R. Larson. Frames, bases, and group representations. In *Mem. Amer. Math. Soc.*, volume 147, pages x+94. AMS, 2000. 38

S. Harizanov, J.-C. Pesquet, and G. Steidl. Epigraphical projection for solving least squares anscombe transformed constrained optimization problems. In A. Kuijper et al., editor, *Scale-Space and Variational Methods in Computer Vision*, volume 7893 of *Lecture Notes in Computer Science*, pages 125–136. SSVM 2013, LNCS 7893 Springer-Verlag, 2013. 109

P. Harremoës. Interpretations of rnyi entropies and divergences. *Physica A: Statistical Mechanics and its Applications*, 365(1):57–62, 2006. 86

R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. 62

Yun He, A. B. Hamza, and H. Krim. A generalized divergence measure for robust image registration. 51(5):1211 – 1220, May 2003. 127

E. Hellinger. Neue begrndung der theorie quadratischer formen von unendlichvielen vernderlichen. *J. fr die reine und angewandte Mathematik*, 136:210–271, 1909. 84, 85

Yong Seok Heo, Kyoung-Mu Lee, and Sang-Uk Lee. Robust stereo matching using adaptive normalized cross-correlation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(4):807–822, Jul. 2011. 33

Yong Seok Heo, Kyoung Mu Lee, and Sang Uk Lee. Joint depth map and color consistency estimation for stereo images with different illuminations and cameras. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(5):1094–1106, Aug. 2013. 33

C. Hernandez and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004. 60

A. O. Hero, B. Ma, O. Michel, and J. Gorman. Applications of entropic spanning graphs. *IEEE Signal Process. Mag.*, 19(5):85–95, 2002. 86

S. Hiltunen, J. Pesquet, and B. Pesquet-Popescu. Comparison of two proximal splitting algorithms for solving multilabel disparity estimation problems. In *Proc. IEEE Eur. Signal Process. Conf.*, pages 1134 –1138, Bucharest, Romania, Aug. 2012. 63

J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms, Part I : Fundamentals*, volume 305 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, Heidelberg, N.Y., 2nd edition, 1996. 16, 88

H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, pages 1–8, Minneapolis, USA, Jun. 2007. 33

B. Hu, I. Land, L. Rasmussen, R. Piton, and B. H. Fleury. A divergence minimization approach to joint multiuser decoding for coded CDMA. *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, 26(3): 432–445, Apr. 2008. 136

H. Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(10):1333–1336, Sep. 2003. 32, 33

H. Jeffreys. An invariant form for the prior probability in estimation problems. In *Proceedings of the Royal Society of London*, volume 186, pages 453–461, 1946. 85

A. Jezierska, E. Chouzenoux, J.-C. Pesquet, and H. Talbot. A primal-dual proximal splitting approach for restoring data corrupted with Poisson-Gaussian noise. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, pages 1085–1088, Kyoto, Japan, Mar. 2012. 10

Hao Jiang, M.S. Drew, and Ze-Nian Li. Matching by linear programming and successive convexification. *IEEE Trans. Pattern Anal. Mach. Int.*, 29 (6):959–975, Jun. 2007. 13

Il-Lyong Jung and Chang-Su Kim. Robust view synthesis under varying illumination conditions using segment-based disparity estimation. In *Proc. in APSIPA*, pages 1–4, Hollywood, Canada, Dec. 2012. 33

T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: theory and experiment. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(9):920–932, Sep. 1994. 33

S.B. Kang and R. Szeliski. Extracting view-dependent depth maps from a collection of images. *Int. J. Comp. Vis.*, 58(2):139–163, Jul. 2004. 60

Sing Bing Kang, R. Szeliski, and Jinxiang Chai. Handling occlusions in dense multi-view stereo. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, volume 1, pages 103–110, Kauai, USA, Dec. 2001. 33

M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. Comput. Vision*, 1(4):321–331, Jan. 1988. 60

S. M. Kay. *Fundamentals of statistical signal processing: estimation theory.* Prentice-Hall PTR, April 1998. 10

A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *Proc. Int. Conf. Pattern Recogn.*, volume 3, pages 15–18, Hong Kong, Aug. 2006. 32, 33

K. Koh, S. J. Kim, and S. Boyd. Solver for l1-regularized least squares problems. tech. report, Stanford University, 2007. 25

V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *Proc. IEEE Int. Conf. Comput. Vis.*, volume 2, pages 508–515, Vancouver, BC, Canada, Jul. 2001. 32, 33, 46, 47

N. Komodakis, N. Paragios, and G. Tziritas. MRF energy minimization and beyond via dual decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(3):531–552, May 2011. 32

M. Kowalski, E. Vincent, and R. Gribonval. Beyond the narrowband approximation: Wideband convex methods for under-determined reverberant audio source separation. *IEEE Trans. Audio, Speech Language Process.*, 18(7):1818–1829, Sep. 2010. 10

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951. 84

K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. In *Proc. IEEE Int. Conf. Comput. Vis.*, volume 1, pages 307–314, Kerkyra, Sep. 1999. 60

A. T. Kyrillidis, S. Becker, and V. Cevher. Sparse projections onto the simplex. In *International Conference on Machine Learning*, volume 28, page 235243, Atlanta, Georgia, USA, 2013. 25

E.S. Larsen, P. Mordohai, M. Pollefeys, and H. Fuchs. Temporally consistent reconstruction from multiple video streams using enhanced belief propagation. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 1–8, Rio de Janeiro, Oct. 2007. 61

L. LeCam. Convergence of estimates under dimensionality restrictions. *Ann. Statist.*, 1(1):38–53, 1973. 84, 85

S. Lefkimmiatis, A. Bourquard, and M. Unser. Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Trans. Image Process.*, 21(3):983–995, Mar. 2012. 39

F. Liese and I. Vajda. *Convex Statistical Distances*. Treubner, 1987. 84, 86

F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *IEEE Trans. on inf. theory*, 52(10):4394–4412, 2006. 84, 85, 86

V. Lieven and B. Stephen. Semidefinite programming. *SIAM Review*, 38(1): 49–95, Apr. 1995. 13

Shubao Liu and D.B. Cooper. A complete statistical inverse ray tracing approach to multi-view stereo. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, pages 913–920, Providence, RI, Jun. 2011. 60

F. Malgouyres. Minimizing the total variation under a general convex constraint for image restoration. *IEEE Trans. Image Process.*, 11(12): 1450–1456, Dec. 2002. 14

E. Malis. Improving vision-based control using efficient second-order minimization techniques. In *Proc. IEEE Int. Conf. Robot. Automat.*, volume 2, pages 1843 –1848, Barcelona, Spain, Apr. 2004. 69

S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, USA, 1997. 14, 38

W. S. Malpica and A. C. Bovik. Range image quality assessment by structural similarity. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, pages 1149–1152, Washington, DC, USA, Apr. 2009. 121, 130

A. Malviya and S. G. Bhirud. Wavelet based image registration using mutual information. In *Proc. IEEE ELECTRO Conference*, pages 241–244, Varanasi, Dec. 2009. 127

L. Meizhu, B.C. Vemuri, S.-I. Amari, and F. Nielsen. Total bregman divergence and its applications to shape retrieval. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, pages 3463–3468, San Francisco, CA, Jun. 2010. 85

Y. Meyer. *Ondelettes et opérateurs: Ondelettes*. Actualités mathématiques. Hermann, 1990. ISBN 9782705661250. URL http://books.google.fr/books?id=ZzbvAAAAMAAJ. 38

W. Miled, J.-C. Pesquet, and M. Parent. Disparity map estimation using a total variation bound. In *The 3rd Canadian Conf. on Computer and Robot Vision*, pages 48–55, Quebec, Canada, Jun. 2006. 32, 33

W. Miled, J.-C. Pesquet, and M. Parent. A convex optimisation approach for depth estimation under illumination variation. *IEEE Trans. Image Process.*, 18(4):813–830, Apr. 2009a. 10, 14, 32, 33, 34, 47, 48

W. Miled, J.-C. Pesquet, and M. Parent. A convex optimisation approach for depth estimation under illumination variation. *IEEE Trans. Image Process.*, 18(4):813–830, Apr. 2009b. 13

D. Min, J. Lu, and M. Do. A revisit to cost aggregation in stereo matching: how far can we reduce its computational redundancy? In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 1567–1574, Barcelona, Spain, Nov. 2011. 47

T. Minka. Divergence measures and message passing. Technical report, Microsoft Research Technical Report (MSR-TR-2005), 2005. 86

J. J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sci.*, 255:2897–2899, 1962. 99, 101, 104

J. J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965. 17, 89

D. Mukherjee, G. Wang, and J. Wu. Stereo matching algorithm based on curvelet decomposition and modified support weights. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, pages 758–761, Dallas, Texas, USA, Mar. 2010. 47

Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex*. Society for Industrial and Applied Mathematics, 1994. Philadelphia. 17

F. Nielsen. On the symmetrical Kullback-Leibler Jeffreys centroids. 2013. http://arxiv.org/abs/1303.7286v2. 85

F. Nielsen and R. Nock. Sided and symmetrized bregman centroids. *IEEE Trans. on inf. theory*, 55(6):2882–2904, 2009. 85

F. Nielsen and R. Nock. On the chi square and higher-order chi distances for approximating f-divergences. 2013. http://arxiv.org/abs/1309.3029. 84

Vincent Nozick. Multiple view image rectification. In *Proc. IEEE Int. Symp. on Access Spaces*, pages 277–282, Yokohama, Japan, Jun. 2011. ISBN 978-1-4577-0716-2. 62

M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Trans. Pattern Anal. Mach. Int.*, 15(4):353–363, Apr. 1993. 60

S. Ono and I. Yamada. Second-order Total Generalized Variation constraint. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, 2014. preprint. 109

F. Österreicher. Distances based on the perimeter of the risk set of a testing problem. *Austrian Journal of Statistics*, 42(1):3–19, 2013. 84

N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):123–231, 2013. To appear. 25, 67, 85, 90

Il Park, S. Seth, M. Rao, and J.C. Principe. Estimation of symmetric chi-square divergence for point processes. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, pages 2016–2019, Prague, May 2011. 86

K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonable be supposed to have arisen from random sampling. *Phil. Mag.*, 50:157–175, 1900. 84, 85

J.-C. Pesquet and N. Pustelnik. A parallel inertial proximal optimization method. *Pac. J. Optim.*, 8(2):273–305, Apr. 2012. 18, 20, 34, 43, 92, 93

G. Peyré. A review of adaptive image representations. *IEEE J. Selected Topics Signal Process.*, 5(5):896–911, Feb. 2011. 14, 53, 116

G. Peyré and J. Fadili. Group sparsity with overlapping partition functions. In *Proc. Eur. Sig. and Image Proc. Conference*, pages 303–307, Barcelona, Spain, Aug. 2011. 14, 54

P. Piro, S. Anthoine, E. Debreuve, and M. Barlaud. Image retrieval via kullback-leibler divergence of patches of multiscale coefficients in the knn framework. In *Content-Based Multimedia Indexing*, pages 230–235, London, Jun. 2008. 84

P. Piro, S. Anthoine, E. Debreuve, and M. Barlaud. Combining spatial and temporal patches for scalable video indexing. *Multimedia Tools and Applications*, 48(1):89–104, May 2010. ISSN 1380-7501. 85

T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multilabel problems. In *Proc. Eur. Conf. Comput. Vis.*, pages 792–805, Marseille, France, Oct. 2008. 46, 47, 63

T. Pock, D. Cremers, H. Bischof, and A. Chambolle. Global solutions of variational models with convex regularization. *SIAM J. Imaging Sci.*, 3 (4):1122–1145, Dec. 2010. 10, 13, 32

B. T. Polyak. Minimization of unsmooth functionals. *USSR Computational Mathematics and Mathematical Physics*, 9(3):14–29, 1969. 23

J.-P. Pons, R. Keriven, and O. Faugeras. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *Int. J. Comp. Vis.*, 72(2):179–193, Apr. 2007. 60

N. Pustelnik, C. Chaux, and J.-C. Pesquet. Parallel ProXimal algorithm for image restoration using hybrid regularization. *IEEE Trans. Image Process.*, 20(9):2450 –2462, Nov. 2011. 14, 38, 85, 101, 102

N. Pustelnik, J.-C. Pesquet, and C. Chaux. Relaxing tight frame condition in parallel proximal methods for signal restoration. *IEEE Trans. Signal Process.*, 60(2):968–973, Feb. 2012. 18, 43

A. Quattoni, X. Carreras, M. Collins, and T. Darrell. An efficient projection for $\ell_{1,\infty}$ regularization. In *International Conference on Machine Learning*, Montreal, Quebec, Jun. 2009. 25

H. Raguet, J. Fadili, and G. Peyré. Generalized forward-backward splitting. *SIAM J. Imaging Sci.*, 6(3):1199–1226, Jul. 2013. 34

T.W. Rauber, T. Braun, and K. Berns. Probabilistic distance measures of the dirichlet and beta distributions. *Pattern Recogn.*, 41(2):637–645, 2008. 85

A. Rényi. On measures of entropy and information. In *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.,*, volume 1, pages 547–561, California, Berkeley, Jun. 1961. 86

W. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am. A*, 62(1):55–59, Jan. 1972. 84

L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, Nov. 1992. 14, 38

D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, volume 1, pages 195–202, Madison, WI, Jun. 2003. 72

M. Schmidt, E. van den Berg, M. Friedlander, and K. Murphy. Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm. In *Proc. of Conf. on Artifcial Intelligence and Statistics*, volume 5, pages 456–463, Clearwater Beach, Florida, Apr. 2009. 25

S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, volume 1, pages 519–528, Washington, DC, USA, Jun. 2006. 60

S. Setzer, G. Steidl, and T. Teuber. Deblurring Poissonian images by split Bregman techniques. *J. Visual Communication and Image Representation*, 21(3):193–199, Apr. 2010. 20, 67, 92

C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 623–656, Jul. 1948. 84

N. Z. Shor. *Minimization Methods for Non-differentiable Functions*. Springer-Verlag, New York, 1985. 23

K. Slavakis, I. Yamada, and N. Ogura. The adaptive projected subgradient method over the fixed point set of strongly attracting nonexpansive mappings. *Numerical Functional Analysis and Optimization*, 27(7-8):905–930, Nov. 2006. 23

G. Steidl and T. Teuber. Removing multiplicative noise by Douglas-Rachford splitting methods. *J. Math. Imag. Vis.*, 36(3):168–184, Feb. 2010. 18

T. Teuber, G. Steidl, and R. H. Chan. Minimization and parameter estimation for seminorm regularization models with I-divergence constraints. *Preprint University of Kaiserslautern*, 2012. URL https://kluedo.ub.uni-kl.de/frontdoor/index/index/docId/3218. 85

T. Teuber, G. Steidl, and R. H. Chan. Minimization and parameter estimation for seminorm regularization models with I-divergence constraints. *Inverse Problems*, 29:1–28, 2013. 13, 116

S. Theodoridis, K. Slavakis, and I. Yamada. Adaptive learning in a world of projections. *IEEE Signal Process. Mag.*, 28(1):97–123, Jan. 2011. 10

L. Thorlund-Petersen. Global convergence of Newtons method on an interval. *Mathematical Methods of Operations Research*, 59(1):91–110, Jan. 2004. 112

A. Tikhonov. Tikhonov regularization of incorrectly posed problems. *Soviet Mathematics Doklady*, 4:1624–1627, 1963. 12, 16

M Tofighi, K. Kose, and A. E. Cetin. Signal reconstruction framework based on Projections onto Epigraph Set of a Convex cost function (PESC). 2014. preprint. 109

D. Tsai, C. Lin, and J. Chen. The evaluation of normalized cross correlations for defect detection. *Pattern Recogn. Lett.*, 24(15):2525–2535, Nov. 2003. 33, 36, 44

P. Tseng. Alternating projection-proximal methods for convex programming and variational inequalities. *SIAM J. Optim.*, 7(4):951–965, 1997. 83

P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, 38(2):431–446, 1998. 21

P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.*, 109(3):475–494, Jun. 2001. 17

B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667–681, Apr. 2013. 18, 93

I. Vajda. *Theory of Statistical Inference and Information*. Kluwer, Dordrecht, 1989. 84, 86

Sara van de Geer. Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.*, 21(1):14–44, 1993. 85

J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Trans. Image Process.*, 16(9):2207–2214, Sep. 2007. 33

E. Van Den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31(2):890–912, Nov. 2008. 14, 25, 41, 67

B. Vandeghinste, B. Goossens, J. De Beenhouwer, A. Pizurica, W. Philips, S. Vandenberghe, and S. Staelens. Split-Bregman-based sparse view CT reconstruction. In *International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 431–434, Potsdam, Germany, Jul. 2011. 10

B.C. Vemuri, L. Meizhu, S.-I. Amari, and F. Nielsen. Total bregman divergence and its applications to DTI analysis. *IEEE Trans. Med. Imag.*, 30 (2):475–483, Feb. 2011. 85

G. Vogiatzis, C. Hernandez, P. H S Torr, and R. Cipolla. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(12):2241–2246, 2007. 60

L. Wang and R. Yang. Global stereo matching leveraged by sparse ground control points. In *Proc. IEEE Conf. Comput. Vis. and Patt. Rec.*, pages 3033–3040, Colorado Springs, USA, Jun. 2011. 47

S. Wanner and B. Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013. to appear. 32, 33

P. Weiss, L. Blanc-Féraud, and G. Aubert. Efficient schemes for total variation minimization under constraints in image processing. *SIAM J. Sci. Comput.*, 31(3):2047–2080, Apr. 2009. 14, 25

M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn. (CVPR)*, pages 2464–2471, San Francisco, CA, Jun. 2010. 14, 53, 116

O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon. Global stereo reconstruction under second-order smoothness priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(12):2115–2128, Dec. 2009. 75, 76, 77

S. J. Wright. *Primal-dual interior-point methods*. SIAM J. Control Optim., Philadelphia, PA, 1997. 17

S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.*, 57(7):2479–2493, Mar. 2009. 25

J. Xiao, H. Cheng, H. Sawhney, C. Rao, and M. Isnardi. Bilateral filtering-based optical flow estimation with occlusion detection. In *Computer Vision ECCV*, volume 3951 of *Lecture Notes in Computer Science*, pages 211–224. Springer Berlin Heidelberg, 2006. 53

Q. Yang, L. Wang, R. Yang, S. Wang, M. Liao, and D. Nistr. Realtime global stereo matching using hierarchical belief propagation. In *Proc. British Machine Vision Conference*, pages 989–998, Edinburgh, U.K, Sep. 2006. 32, 33

K.-J. Yoon and I.-S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Anal. Mach. Int.*, 28(4):650–656, Apr. 2006. 33, 44, 53

D. C. Youla and H. Webb. Image restoration by the method of convex projections. Part I - theory. *IEEE Trans. Med. Imag.*, 1(2):81–94, Oct. 1982. 17

A. L. Yuille and T. Poggio. A generalized ordering constraint for stereo correspondence. Technical Report AI-M-777, Massachussetts Institute of Technology (MIT). Cambridge (MA US), 1984. 45

R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In Jan-Olof Eklundh, editor, *Proc. Eur. Conf. Comput. Vis.*, volume 801, pages 151–158. Springer Berlin Heidelberg, May 1994. 33

R. Zanella, P. Boccacci, L. Zanni, and M. Bertero. Efficient gradient projection methods for edge-preserving removal of poisson noise. *Inverse Problems*, 25(4), 2009. 84

J. Zhang. Divergence function, duality, and convex analysis. *Neural Comput.*, 16:159–195, 2004. 86

B. Zitovà and J. Flusser. Image registration methods: a survey. *Image Vision Comput.*, 21(11):977–1000, 2003. 127