Chapitre 1

Théorie semi-quadratique, régularisation et estimation robuste

Pierre Charbonnier,
Jean-Philippe Tarel
Sio-Song Teng

1.1 Introduction

Les images numériques ont de plus en plus de domaines d'application de nos jours. Il convient d'ailleurs de prendre le mot "image" au sens large de cartographie d'une propriété physique, éventuellement complexe ou vectorielle, en tout cas non limitée à l'intensité photométrique. Ces images doivent souvent être obtenues par le calcul, à partir de données issues de capteurs, voire à partir d'autres images. Les images optiques elle-mêmes sont issues de systèmes imageurs qui peuvent introduire du flou ou des perturbations aléatoires – le bruit – dont il convient de les débarrasser pour les rendre exploitables. On peut parler, de manière générique, de *reconstruction* d'image.

Le terme *vision par ordinateur*, quant à lui, regroupe les techniques visant à analyser automatiquement les images afin d'en ressortir des informations utilisables par des systèmes de détection, de commande ou de reconnaissance. L'approche classique consiste à extraire dans les images des primitives géométriques qui sont ensuite abstraites, grâce à l'utilisation de modèles mathématiques. Il s'agit alors d'estimer les paramètres permettant d'ajuster au mieux un modèle aux données observées. Plus récemment, on a vu apparaître des méthodes, rendues accessibles par la montée en puissance des ordinateurs, qui utilisent l'ensemble des valeurs de l'image. Elles capturent l'apparence globale des objets à travers des modèles statistiques. L'analyse d'une image implique, là encore, un ajustement de paramètres.

Ces deux types d'applications, reconstruction et analyse d'images, peuvent sembler très différents. Dans les deux cas, cependant, on peut considérer que les données p auxquelles nous avons accès sont une réalisation bruitée d'une relation fonctionnelle $\mathcal{R}(f)$ dont on souhaite estimer les paramètres f. Pour résoudre ce problème inverse, il est nécessaire d'introduire des connaissances sur le bruit. De plus, il est souvent souhaitable – et parfois nécessaire – d'introduire une connaissance a priori sur la nature de la solution recherchée, on parle alors de régularisation. Le cadre bayesien est le plus générique pour spécifier ces informations de nature statistique.

Le paragraphe 1.2 est consacré à la description du modèle et des principaux estimateurs associés. Nous verrons également que les hypothèses gaussiennes classiquement utilisées sont inadaptées. Ainsi, dans le cas de l'estimation de paramètres, elles ne permettent pas de gérer correctement la présence de données erronées. En reconstruction d'image, elles entraînent un lissage excessif des contours. Nous proposons au paragraphe 1.3 une présentation unifiée de modèles basés sur des densités de probabilité non gaussiennes (ou d'énergies non quadratiques). Ces modèles permettent de relâcher le comportement gaussien lorsque cela devient nécessaire. La théorie semi-quadratique précise les conditions à imposer aux fonctions employées pour autoriser un tel comportement. On montre que cela revient à introduire une variable auxiliaire dont le rôle est à la fois de "marquer" les données et de linéariser le problème, simplifiant ainsi l'algorithmique associée. Nous illustrons au paragraphe 1.4 deux applications d'estimation robuste de paramètres dans le cadre de la détection automatique de la signalisation routière. Nous montrons également le gain de qualité obtenu en reconstruction d'image régularisée.

1.2 Le modèle classique et ses limites

1.2.1 Modèle génératif linéaire

Dans un premier temps, on se donne un modèle déterministe de formation des données, ou modèle génératif. Dans la pratique, un modèle linéaire s'avère souvent suffisant et on écrira cette relation (en variables discrètes) :

$$p = Rf. \tag{1.2.1}$$

La forme de la matrice R, connue, dépend de l'application traitée. Dans le cas de la reconstruction d'image, les données p et l'image inconnue, f, sont arrangées sous forme de vecteurs par ordonnancement lexicographique. Le vecteur p peut être vu comme une somme des colonnes de la matrice R, chacune étant pondérée par la valeur du pixel de f correspondant. On comprend ainsi que chaque colonne de R représente la fonction d'étalement d'un point de l'espace objet (en anglais *Point Spread Function* ou PSF). Ce modèle s'applique par exemple au cas de la tomographie, où la PSF est différente pour chaque point. Cela peut également être le cas en déconvolution mais, la plupart du temps, la PSF est invariante spatialement. Dans le cas du débruitage, la matrice R est simplement égale à l'identité.

En ce qui concerne l'analyse d'images, prenons l'exemple de la régression quadratique. On dispose d'une liste de coordonnées $\{x_i, y_i\}_{i=1\cdots N}$ de points d'intérêt, extraits d'une image par un détecteur, et que l'on suppose alignés selon une courbe. Une relation sous-jacente de la forme $y = a + bx + cx^2$ est vérifiée pour chaque couple de coordonnées. La relation (1.2.1) s'écrit donc en collectant les ordonnées $\{y_i\}$ dans le vecteur p. Les inconnues, a, b et c forment le vecteur f et la matrice R est formée d'un vecteur de 1, du vecteur des abscisses $\{x_i\}$, et du vecteur des monômes de degré deux $\{x_i^2\}$. Ce modèle peut aisément être étendu à des combinaisons linéaires d'autres fonctions que les monômes. On peut ainsi modéliser des relations non linéaires entre les coordonnées, mais le modèle génératif (1.2.1) demeure linéaire par rapport à f.

Le modèle (1.2.1) résume la partie déterministe de la connaissance que l'on peut avoir du processus de formation des données. Il est cependant incomplet, car il ne tient pas compte de la nature des perturbations aléatoires sur l'observation, ou bruit. Très

souvent, celui-ci est considéré additif et indépendant de f et (1.2.1) devient alors :

$$p = Rf + n. \tag{1.2.2}$$

Dans le cadre de l'estimation bayesienne, on adjoint à ce modèle des informations de nature statistique, sur les différentes variables qui interviennent.

1.2.2 Structure probabiliste et estimation bayesienne

Une première densité de probabilité intéressante traduit le caractère aléatoire des mesures : $\mathcal{P}(p|f)$. Elle quantifie la vraisemblance d'une observation connaissant les valeurs des paramètres. Il s'agit en fait de la densité de probabilité du bruit, $n = p - Rf$. Une hypothèse très classique considère que le bruit est additif, gaussien, centré, isotrope, indépendant et identiquement distribué (iid). Dans ce cas :

$$\mathcal{P}(p|f) \propto exp\left(-\frac{1}{2\sigma^2}\|p - Rf\|^2\right). \tag{1.2.3}$$

On peut rechercher les paramètres rendant la plus plausible l'observation en maximisant la vraisemblance. Dans le cas gaussien, on retrouve l'estimateur classique des moindres carrés :

$$\hat{f} = \underset{f}{arg\,min}\,\frac{1}{2\sigma^2}\|p - Rf\|^2. \tag{1.2.4}$$

L'estimateur du maximum de vraisemblance (MV) a de bonnes propriétés statistiques [Kay93]. Cependant, il considère la variable f comme un paramètre déterministe. Au contraire, l'approche bayesienne considère f comme une variable aléatoire dont la distribution $\mathcal{P}(f)$ traduit la connaissance a priori sur la solution recherchée. Dans ce cadre, la probabilité associée à p est obtenue en intégrant par rapport à f la vraisemblance conjointe :

$$\mathcal{P}(p) = \int \mathcal{P}(p, f)\, df = \int \mathcal{P}(p|f)\, \mathcal{P}(f)\, df. \tag{1.2.5}$$

Cependant, cette intégrale est souvent inexploitable analytiquement et on a recours à une estimation ponctuelle qui correspond implicitement à une approximation classique en inférence bayesienne. Elle considère que la distribution conjointe est suffisamment "piquée" pour qu'on puisse approcher son intégrale par la hauteur du pic multipliée par sa "largeur", σ_{pic} : $\mathcal{P}(p) \simeq \sigma_{pic}\, \max_f \mathcal{P}(p, f) \propto \mathcal{P}(p|\hat{f})\, \mathcal{P}(\hat{f})$, où l'estimée $\hat{f}$ est donnée par :

$$\hat{f} = \underset{f}{arg\,max}\, \mathcal{P}(p|f)\, \mathcal{P}(f). \tag{1.2.6}$$

La distribution maximisée étant proportionnelle à la loi a posteriori $\mathcal{P}(f|p)$, cet estimateur est appelé maximum a posteriori (MAP). Les connaissances a priori sur f prises en compte par le MAP peuvent s'avérer très utiles dans des problèmes d'estimation comme la régression, par exemple. On peut ainsi limiter la sensibilité aux bruits, l'influence des forts degrés ou bien contraindre les coefficients de la courbe autour d'une valeur connue ou prédite, notamment dans le cas d'un algorithme de suivi. Pour cela, on impose à f une densité de probabilité, par exemple gaussienne, $\mathcal{N}(\bar{f}, \Sigma)$:

$$\mathcal{P}(f) \propto exp\left(-\frac{1}{2}(f - \bar{f})^T \Sigma^{-1} (f - \bar{f})\right). \tag{1.2.7}$$

En reconstruction d'image, introduire une information a priori est non seulement utile, mais absolument nécessaire. En effet, les problèmes inverses rencontrés sont en général mal posés et la résolution au sens des moindres carrés est instable. Une technique classique, la régularisation de Tikhonov, pénalise la norme de f ou, plus souvent, celle de son gradient. Cela revient également à introduire un a priori gaussien :

$$\mathcal{P}(f) \propto exp\left(-\frac{1}{2\sigma^2}\|\nabla f\|^2\right). \tag{1.2.8}$$

Dans le cas de distributions gaussiennes, en passant au logarithme dans (1.2.6), on retrouve l'estimateur des moindres carrés pénalisés. Soit, dans l'exemple ci-dessus, en collectant les constantes :

$$\hat{f} = \underset{f}{arg\,min}\ |p - Rf\|^2 + \lambda^2\|\nabla f\|^2. \tag{1.2.9}$$

Lorsque la contrainte porte sur f et non sur son gradient, on retrouve le filtre classique de Wiener. Notons que l'estimation au sens du MV peut être vue comme le cas limite du MAP où $\mathcal{P}(f)$ est uniforme.

1.2.3 Limites des estimateurs gaussiens

Les densités de probabilité mises en jeu dans le cadre de l'estimation bayesienne sont souvent de forme gaussienne :

$$\mathcal{P}(u) \propto exp\left(-\frac{1}{2}\sum_{i=1}^{N} u_i^2\right), \tag{1.2.10}$$

où la variable muette u_i peut représenter soit un résidu $n_i = (p - Rf)_i$ (cas de l'estimation) soit la valeur d'un pixel f_i, d'une composante ou de la norme du gradient : $(\nabla_x f)_i$, $(\nabla_y f)_i$ ou $(|\nabla f|)_i$ (cas de la régularisation, cf. §1.4.3). Maximiser $\mathcal{P}(u)$ est équivalent à minimiser une forme quadratique :

$$J(u) = \frac{1}{2}\sum_{i=1}^{N} u_i^2. \tag{1.2.11}$$

Malheureusement, ce type de modèle est notoirement sensible aux fortes déviations u_i. Prenons l'exemple de la régression linéaire (cf. figure 1.1). La présence d'une seule donnée erronée suffit à créer une valeur forte de résidu. L'algorithme chargé de minimiser $J(u)$ a tendance à limiter en priorité ce résidu de la seule façon possible, c'est-à-dire en agissant sur les coefficients f du modèle. On dit que le point de rupture de cet estimateur est de 0 : une seule donnée erronée suffit à modifier le résultat de l'estimation.

Dans le cas de la régularisation, la fonction quadratique s'oppose à la création de forts gradients dans la solution. Celle-ci aura donc un aspect trop lisse (cf. fig. 1.4.3(e) ou 1.8(c)). On souhaiterait, au contraire, obtenir une solution formée de régions lisses, séparées par des bords francs.

Le problème inhérent au modèle gaussien est qu'il ne considère qu'un seul type de données alors qu'il y en a deux, en réalité. Dans le cas de la régression, on distingue les données conformes au modèle (ou "inlier") qui correspondent aux résidus faibles tandis que les données erronées (ou "outlier") entraînent de fortes déviations. Il s'agit

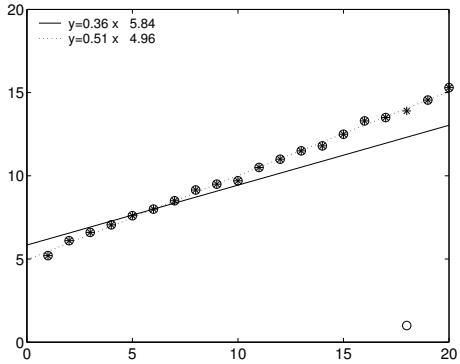

FIGURE 1.1 – Influence d'une donnée erronée sur l'estimation de paramètres au sens des moindres carrés : la droite de régression en pointillés correspond à la série marquées par des points, celle en trait plein correspond à la série marquée par des cercles.

de trouver un moyen de limiter l'influence des données erronées sur l'estimation. Dans le cas de la régularisation par le gradient, les valeurs faibles du gradient correspondent au bruit, qu'il convient de lisser, tandis que les grandes valeurs correspondent aux discontinuités ou contours présents dans les images, que l'on souhaite en fait préserver.

1.3 Théorie semi-quadratique

Intuitivement, on peut envisager deux façons de sortir du comportement quadratique. La première consiste à utiliser une fonctionnelle de la forme :

$$J(f) = \frac{1}{2}\sum_{i=1}^{N} \varphi(u_i). \tag{1.3.1}$$

Cette formulation est connue sous le nom de M-estimateur [Hub81] en estimation robuste (φ est souvent notée ρ dans ce cadre) et de modèle à phi-fonction en régularisation. Toute la question réside alors dans le choix de la fonction φ. Prenons l'exemple de la quadratique tronquée $min(u^2, 1)$. Pour une faible valeur de u_i, cette fonction garde le comportement quadratique habituel. Par contre, à partir de $u_i = 1$, la valeur de la fonction ne change plus et ce, quelle que soit la valeur de u_i. L'algorithme de minimisation n'a donc pas particulièrement intérêt à modifier cette valeur en agissant sur f. La forte déviation perd ainsi toute influence sur le résultat. L'inconvénient de cette fonction est qu'elle présente un point anguleux en 1, ce qui est une source potentielle d'instabilité. Une fonction continue serait préférable. Mais l'asymptote horizontale est-elle nécessaire ? Du point de vue de l'optimisation, une fonction convexe serait préférable. De fait, de nombreuses fonctions ont été proposées dans la littérature (voir un état de l'art dans [Cha94] pour la régularisation). Quelques exemples sont donnés dans le tableau 1.1.

La seconde approche consiste à utiliser une variable auxiliaire pour "marquer" explicitement les données [GG84]. Considérant une variable binaire $b_i \in \{0, 1\}$, on peut définir le critère *augmenté* suivant :

$$J^*(f, b) = \frac{1}{2}\sum_{i=1}^{N} b_i(u_i)^2 + (1 - b_i). \tag{1.3.2}$$

La variable b est appelée “processus de ligne” en reconstruction régularisée. Le second terme de cette fonctionnelle est un coût associé à la détection d'une forte déviation permettant d'éviter le minimum dégénéré où tous les b_i seraient nuls. L'intérêt de cette formulation est que le critère est quadratique lorsqu'on fixe b (d'où l'appellation “semi-quadratique”) : c'est celui des moindres carrés pondérés.

Nom	$\varphi(u)$	$\varphi'(u)/2u$	(1.3.4-1.3.6)	Convexité
Quadratique Q	u^2	1	non	oui
Hyper-surfaces HS	$2\sqrt{1+u^2}-2$	$1/\sqrt{1+u^2}$	oui	oui
Cauchy ou Hebert & Leahy HL	$\log(1+u^2)$	$1/(1+u^2)$	oui	non
Geman & McClure GM	$u^2/1+u^2$	$1/(1+u^2)^2$	oui	non

TABLE 1.1 – Exemples de fonctions φ proposées dans la littérature (voir fig. 1.2).

Naturellement se pose la question de l'estimation des valeurs de b. En l'absence d'information supplémentaire, cela doit se faire à partir des données, donc de u. Considérons l'optimisation par rapport à b de $J^*(u,b)$ [BZ87]. La fonction quadratique étant inférieure à 1 sur $[0,1]$, le minimum est réalisé pour $b_i = 1$ si $u_i < 1$ et $b_i = 0$ sinon. On remarque alors l'équivalence entre ce modèle et le précédent. De même que précédemment, un comportement moins “binaire” serait cependant souhaitable. Peut-on généraliser cette technique à une variable auxiliaire à valeurs réelles ?

La théorie semi-quadratique répond à ces questions en définissant une famille de fonctions φ qui assure le comportement souhaité. Elle montre l'équivalence entre ce modèle et un modèle où la prise en compte de la nature des données est explicite, grâce à une variable auxiliaire réelle. En cela, elle étend à la fois les travaux de Huber [Hub81] en estimation robuste et de Geman et al. [GR92, GY95] en régularisation.

1.3.1 Propriétés des fonctions φ

Les fonctions de potentiel φ proposées dans la littérature semblent d'allures très différentes (voir figure 1.2). Leur étude a été proposée dans le cas de la régularisation dans [Cha94] et dans [BFCAB95]. Elle se base sur un examen des équations normales liées à l'optimisation du critère. En estimation robuste, les fonctions considérées étant paires, on peut écrire le critère (1.3.1) pour $u_i = |n_i|$, où $n = p - Rf$ est le résidu. Le minimum de J est atteint lorsque sa dérivée est nulle, soit (puisque $sign(x) = \frac{x}{|x|}$) lorsque :

$$\frac{1}{2}\sum_{i=1}^{N}\varphi'(|n_i|)sign(n_i)\frac{\partial n_i}{\partial f_j} = \frac{1}{2}\sum_{i=1}^{N}\underbrace{\frac{\varphi'(|n_i|)}{2|n_i|}}_{Pondération.} \cdot \underbrace{2n_i\frac{\partial n_i}{\partial f_j}}_{Cas\ quadratique} = 0, \quad \forall j = 1\dots D. \tag{1.3.3}$$

On voit que si $\frac{\varphi'(|n_i|)}{2|n_i|}$ tend vers 0, la donnée correspondante n'entre pas en jeu dans la somme, ce qui est souhaitable pour les forts résidus. Par contre, cette pondération doit tendre vers 1 pour que la donnée soit totalement prise en compte, comme dans le cas de faible résidus. En résumé, les conditions à vérifier pour obtenir un comportement

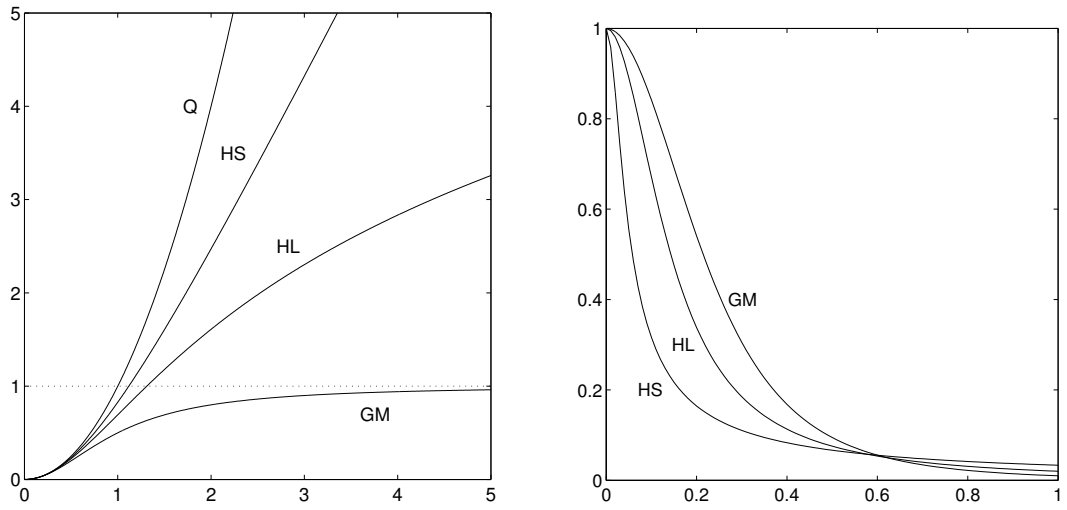

FIGURE 1.2 – A gauche : exemples de fonctions $\varphi(u)$; à droite : fonctions de pondération $\frac{\varphi'(u)}{2u}$ (avec un cœfficient d'échelle arbitraire sur u [Cha94]).

différencié selon les valeurs de u sont les suivantes :

$$\lim_{u\to 0} \frac{\varphi'(u)}{2u} = 1 \tag{1.3.4}$$

$$\lim_{u\to \infty} \frac{\varphi'(u)}{2u} = 0 \tag{1.3.5}$$

$$\frac{\varphi'(u)}{2u} \quad \text{strictement décroissante.} \tag{1.3.6}$$

Ces conditions comparent les comportements de la fonction φ et de la quadratique, en termes de dérivées. Les deux fonctions évoluent de la même façon près de 0 (1.3.4), mais la fonction φ est sous-quadratique à l'infini (1.3.5). La condition (1.3.6) assure la cohérence du modèle. De fait, on constate sur les exemples de la figure 1.2 que les fonctions φ ont des fonctions de pondération au comportement similaire.

1.3.2 Extensions semi-quadratiques

On peut montrer que les fonctions réelles φ paires, croissantes, vérifiant les conditions (1.3.4-1.3.6) et quelques autres hypothèses "techniques", sont telles que :

$$\varphi(u) = \min_{b\in[0,1]} \left(bu^2 + \Psi(b)\right), \forall u \qquad \text{et} \qquad b_{inf} = \frac{\varphi'(u)}{2u}, \tag{1.3.7}$$

où Ψ est une fonction strictement convexe définie à partir de φ et b est la variable auxiliaire. On parle d'extension multiplicative. Ce résultat généralise un théorème introduit dans [GR92], initialement limité aux fonctions à asymptote horizontale. On a également, sous les mêmes conditions, une extension additive :

$$\varphi(u) = \min_{b\in\mathbb{R}^+} \left((b-u)^2 + \xi(b)\right), \forall u \quad \text{et} \quad b_{inf} = \left(1 - \frac{\varphi'(u)}{2u}\right) u, \tag{1.3.8}$$

où ξ est une fonction strictement convexe définie à partir de φ. Ce résultat est basé sur les travaux de [GY95].

1.3.3 Algorithmique

En reportant les équations (1.3.7) et (1.3.8) dans (1.3.1), on obtient respectivement les critères augmentés suivants :

$$J^*(f,b) = \frac{1}{2}\sum_{i=1}^{N} b_i(u_i)^2 + \Psi(b_i) = \frac{1}{2}\|u\|_B^2 + \frac{1}{2}\sum_{i=1}^{N}\Psi(b_i), \tag{1.3.9}$$

avec $B = diag(b_i)$ et :

$$J^{\#}(f,b) = \frac{1}{2}\sum_{i=1}^{N}(b_i - u_i)^2 + \xi(b_i) = \frac{1}{2}\|u-b\|^2 + \frac{1}{2}\sum_{i=1}^{N}\xi(b_i). \tag{1.3.10}$$

Ces critères augmentés d'une variable sont en fait plus simples à minimiser que le critère original. En effet, la minimisation de J conduit à des équations normales non linéaires. Par contre, à b fixée, les critères augmentés sont quadratiques et conduisent donc à la résolution d'équations normales linéaires, de la forme $Af = c$. Notons que dans cette minimisation, les fonctions Ψ et ξ n'interviennent pas (il n'existe d'ailleurs pas d'expression explicite de ξ). D'autre part, lorsque la variable f est fixée le critère augmenté est convexe en b et la valeur du minimum est donnée dans (1.3.7) et (1.3.8). Cette valeur dépend de la fonction de pondération et d'après (1.3.4) et (1.3.5), il est clair que la variable auxiliaire b joue le rôle de "marqueur" de données ("inlier/outlier", "bruit/discontinuité" dans nos applications).

Bien que des algorithmes stochastiques existent [GR92, Jal01], les propriétés semi-quadratiques sont, la plupart du temps, exploitées sous forme d'algorithmes déterministes opérant par minimisations alternées par rapport à chacune des variables. En général, on positionne initialement tous les b_i à 1, ce qui revient à commencer par un algorithme des moindres carrés classique. Dans le cadre de l'estimation robuste, on retrouve pour la forme multiplicative l'algorithme des moindres carrés (re-)pondérés itératifs (Iterative Reweighted Least Squares ou IRLS) et pour la forme additive l'algorithme des moindres carrés avec remise à jour des résidus, proposé dans [Hub81]. Dans le cas de la régularisation, les algorithmes sont nommés ARTUR [CBFAB97] pour la forme multiplicative et LEGEND pour la forme additive [Cha94]. Ces algorithmes convergent dans le cas des fonctions φ convexes. Dans le cas non-convexe, ils convergent vers un minimum local à condition que les points critiques de la fonctionnelle soient isolés. Notons que la convergence est généralement plus lente dans le cas de la seconde forme semi-quadratique, mais que celle-ci permet d'obtenir des algorithmes moins calculatoires (seul le second membre des équations normales est modifié à chaque étape). Le compromis peut donc être intéressant dans le cas d'algorithmes de reconstruction d'image, par exemple. Pour une synthèse et des résultats récents concernant la vitesse de convergence des algorithmes semi-quadratique, le lecteur pourra se référer à [All02].

Enfin, pour améliorer la convergence des algorithmes, on peut adopter une stratégie de continuation, dans la philosophie du GNC (Graduated Non Convexity) de [BZ87]. Cela consiste à utiliser d'abord une fonction convexe (ex. φ_{HS}), puis à repartir du résultat obtenu en employant une fonction non convexe "peu sévère" (ex. φ_{HL}), etc.

1.3.4 Approche Lagrangienne

Il existe plusieurs présentations de la théorie semi-quadratique et des algorithmes associés, chacune apportant un éclairage particulier. Il est cependant important de remarquer qu'elles reposent toutes sur la convexité des fonctions : $-\varphi(\sqrt{u}) = -\phi(u)$ pour la forme multiplicative et $\frac{1}{2}\left(u^2 - \varphi(u)\right)$ pour la forme additive. Ces conditions sont acquises dès lors que φ respecte les conditions (1.3.4-1.3.6).

Nous avons récemment proposé une relecture de la théorie semi-quadratique dans le cadre de l'optimisation sous contrainte [TIC02]. En introduisant la fonction $\phi(u_i^2)$ en place de $\varphi(u_i)$ dans (1.3.1), la pondération s'écrit $\phi'(u_i^2)$ et les conditions d'applications traduites sur ϕ, définie sur $\mathbb{R}^+$, sont qu'elle doit être strictement croissante et concave. Ces conditions sont suffisantes pour pouvoir poser et résoudre le problème d'optimisation en passant par un développement lagrangien qui permet d'interpréter (1.3.9) comme un problème dual. Cette relecture est également possible pour la version additive. On aboutit dans les deux cas aux mêmes algorithmes. Une telle méthodologie a l'avantage de faciliter la preuve de convergence, en la replacant dans le cadre classique du théorème de Khun et Tucker. Ainsi, elle s'étend assez naturellement au cas de l'estimation simultanée des paramètres de plusieurs instances du modèle [TIC05].

1.3.5 Hyper-paramètres

Dans ce qui précède nous avons, par souci de simplicité, éludé le paramètre d'échelle. Dans (1.3.1), la variable u_i doit, en réalité, être divisée par un facteur δ qui ajuste le comportement de la fonction à la dynamique des données. D'autre part, lorsque le critère est régularisé, un coefficient λ^2 permet de pondérer les termes de vraisemblance et d'a priori. Ces *hyper*-paramètres sont en général laissés au choix de l'utilisateur. Il est cependant possible de les fixer de manière plus automatique (voir [Ien04] pour l'estimation robuste et [Jal01] dans le cas de la reconstruction régularisée).

1.4 Applications

Nous illustrons maintenant quelques applications de la théorie semi-quadratique en analyse d'images, puis en reconstruction régularisée.

1.4.1 Régression robuste pour la détection et le suivi des marquages routiers

Nous nous intéressons ici à la détection et au suivi des marquages routiers à partir d'images numériques. Dans des applications telles que l'aide à la conduite, la robustesse des traitements aux perturbations est d'une importance cruciale.

Le principe du système développé [Ien04] (cf. fig. 1.3) consiste à effectuer dans un premier temps une détection des centres supposés des marquages, puis à les approcher par ajustement de courbes. Le fonctionnement du détecteur est détaillé dans [Ien04]. Il se base sur des techniques simples de seuillage et exploite des informations géométriques a priori sur la largeur des marquages, ce qui rend l'algorithme relativement moins sensible à certaines variations d'intensité que les détecteurs de contours classiques. Cependant, les cartes de détection demeurent relativement bruitées et la régression semi-quadratique trouve ici un champ d'application naturel.

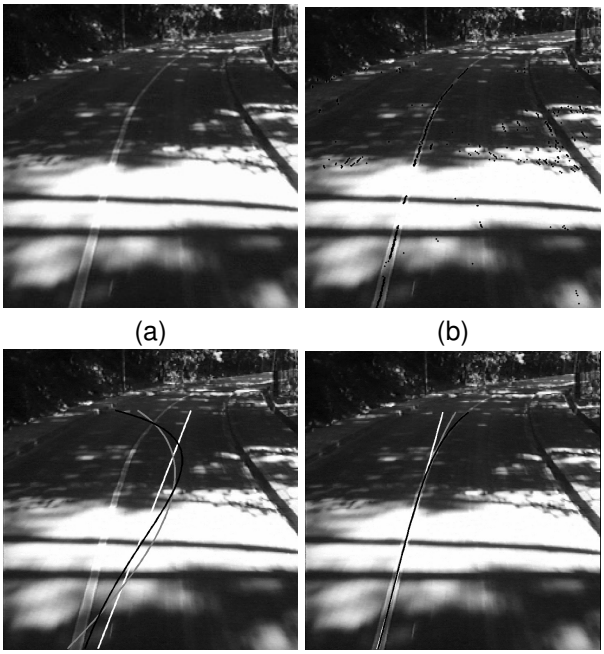

FIGURE 1.3 – (a) Image de route. (b) Centres de marquages extraits (·) : on note la présence de nombreuses fausses alarmes. Ajustement (c) gaussien, (d) robuste : en blanc, degré 1 ; en gris, degré 2 ; en noir, degré 3.

Dans cette application, la relation sous-jacente entre les coordonnées images (x, y) des points de marquage est choisie de type polynomial :

$$y = \sum_{j=0}^{D} X_j(x) f_j, \tag{1.4.1}$$

où $X_j(x)$ représente la valeur en x d'une fonction de base. Selon l'angle de prise de vues, on peut utiliser différentes formes de polynômes (cf. tableau 1.2). En collectant les équations correspondant aux N points de mesure, on obtient un système de la forme (1.2.2). Un modèle original de bruit, appelé famille exponentielle lissée, a été proposé dans [TIC02] et se base sur la fonction de potentiel suivante :

$$\varphi_\alpha(u) = \frac{1}{\alpha}\left((1+u^2)^\alpha - 1\right). \tag{1.4.2}$$

Cette famille paramétrée par α s'avère bien adaptée à la modélisation des résidus dans notre application. D'autre part, en faisant varier α, elle permet une transition continue entre les différents modèles de bruit classiques (de Gauss pour $\alpha = 1$ à Geman & McClure pour $\alpha = -1$, en passant par Cauchy pour $\alpha = 0.5$). Un algorithme semi-quadratique est obtenu à partir de l'approche lagrangienne décrite au paragraphe 1.3.4. La figure 1.3 montre un exemple d'application de cet algorithme pour la détection d'une ligne de marquage dans des conditions d'illumination difficiles.

Il est possible d'intégrer à la fonctionnelle un terme de régularisation quadratique permettant de contrôler le degré de la courbe obtenue, voire de tenir compte des corrélations existant entre les fonctions de base. Il est également possible de tenir compte dans l'a priori d'un résultat d'ajustement précédent. L'estimateur s'intègre donc

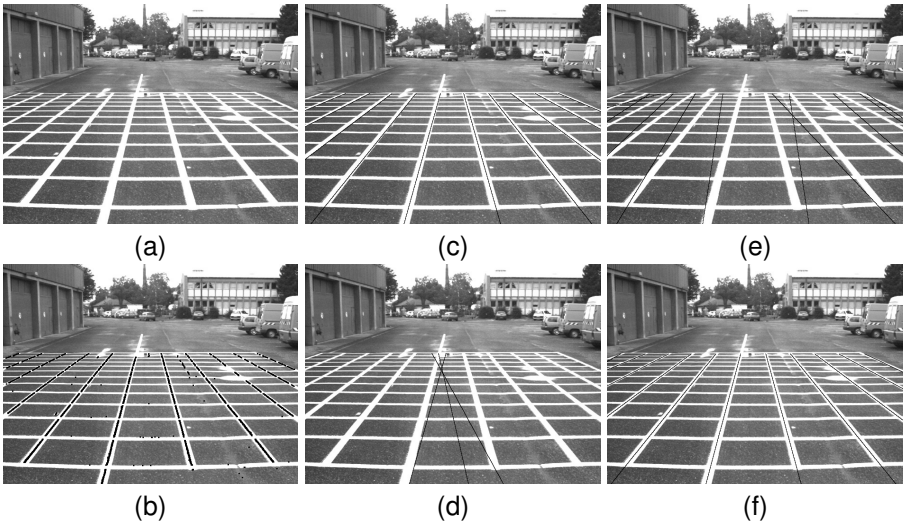

FIGURE 1.4 – (a) Image originale de la grille d'étalonnage. (b) Centres de marquages extraits (·). (c) Initialisation de 10 lignes pour l'ajustement des marquages verticaux. (d) Lignes ajustées dans l'hypothèse de distribution gaussienne pour le bruit. (e) Initialisation de 12 lignes. (f) ajustement robuste simultané des marquages, potentiel φ_α, $\alpha = 0.1$.

naturellement dans un algorithme de suivi par filtrage de Kalman [TIC02]. Cela pose le problème de la définition de la matrice de covariance de l'estimée, laquelle ne peut être qu'approchée dans le cas de l'estimation robuste. Plusieurs formes de matrice ont été comparées [Ien04]. L'algorithme de suivi a été testé avec succès à bord de véhicules équipés pour l'aide à la conduite au LIVIC [Ien04].

Une extension à l'ajustement simultané de plusieurs courbes a été proposée récemment [TIC05] à partir du cadre lagrangien. La figure 1.4 en donne un exemple d'application. Il s'agit de détecter simultanément toutes les lignes longitudinales d'une grille de marquage utilisée pour l'étalonnage de caméras. On peut constater qu'en utilisant un modèle gaussien, l'algorithme produit un résultat erroné malgré une initialisation manuelle très proche de la solution. Au contraire, l'algorithme produit de très bon résultats, à partir d'une initialisation pourtant plus éloignée, lorsqu'on utilise un modèle robuste de vraisemblance.

Polynôme	Prise de vue	Expression
Usuel	Verticale	$y = \sum_{j=0}^{D} f_j x^j$
Hyperbolique	Perspective	$y = f_0 x + f_1 + \sum_{j=2}^{D} \frac{f_j}{(x-x_h)^j}$ (x_h=ligne d'horizon)
Fractionnaire	Perspective	$y = \sum_{j=0}^{D} f_j x^{(\frac{j}{D})}$

TABLE 1.2 – Types de polynôme en fonction de l'angle de prise de vue. Les polynômes fractionnaires ne nécessitent pas la connaissance de la position de la ligne d'horizon.

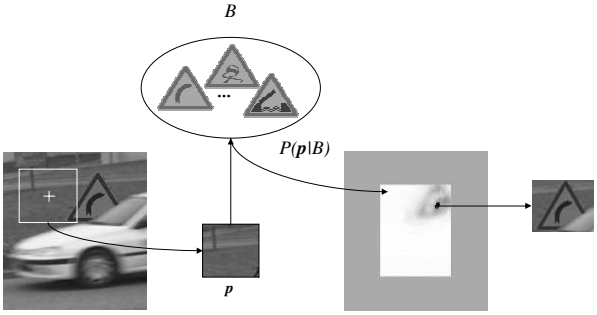

FIGURE 1.5 – Principe de l'algorithme de détection.

1.4.2 Modèles d'apparence pour la détection et la reconnaissance de la signalisation verticale

La détection d'objets dans des images est un sujet difficile qui a fait l'objet de nombreux travaux par le passé. Les approches par *modèles d'apparence*, notamment, rencontrent un succès certain depuis le début des années 90 [TP91]. Elle permettent la représentation de classes d'objets et l'apprentissage de variations de forme, d'orientation ou d'illumination. Nous en présentons ici une application à la détection et à la reconnaissance de la signalisation verticale dans les scènes routières [Dah01].

Dans cette approche, toute image est représentée par un vecteur dont chaque composante est la valeur de niveau de gris d'un pixel. Une telle représentation, globale, de l'apparence est évidemment peu parcimonieuse : c'est pourquoi on a en général recours à des techniques de réduction de dimension. Nous utilisons pour notre part l'analyse en composantes principales (ACP). Elle permet de modéliser une base $\mathcal{B}$ d'images d'apprentissage à l'aide d'une image moyenne μ et d'un nombre réduit de vecteurs propres u_j, représentant les axes principaux de variation du nuage de points. Dans notre modèle, toute image peut se décomposer comme une combinaison linéaire de ces vecteurs de base, à une erreur w près :

$$p = \mu + \sum_{j=1}^{J} f_j u_j + w. \tag{1.4.3}$$

Nous retrouvons un modèle génératif linéaire similaire à (1.2.2). L'intérêt de ce type de modèle est que J est relativement faible (typiquement, quelques dizaines de coefficients, à comparer aux centaines de milliers de pixels des images traitées). Les comparaisons nécessaires à la détection et à la reconnaissance s'en trouvent donc fortement simplifiées.

Le principe de l'algorithme de détection d'objets d'intérêt dans une scène est illustré figure 1.5. En chaque position de l'image, on extrait un vecteur d'observation p dont on évalue la vraisemblance $\mathcal{P}(p|\mathcal{B})$, par rapport au modèle appris sur la base $\mathcal{B}$. Cette valeur est affectée à la position de l'extrait. Lorsque toute la scène a été parcourue, on dispose d'une carte de vraisemblance. La position des objets d'intérêt éventuellement présents dans la scène est alors obtenue par seuillage. La vraisemblance de l'observation, $\mathcal{P}(p|\mathcal{B})$, est donnée par :

$$\mathcal{P}(p|\mathcal{B}) \propto \mathcal{P}(p|\hat{f}, \mathcal{B})\, \mathcal{P}(\hat{f}|\mathcal{B}), \tag{1.4.4}$$

où $\hat{f} = arg\,max_f\, \mathcal{P}(p, f|\mathcal{B})$ est l'estimée au sens du MAP des paramètres du modèle.

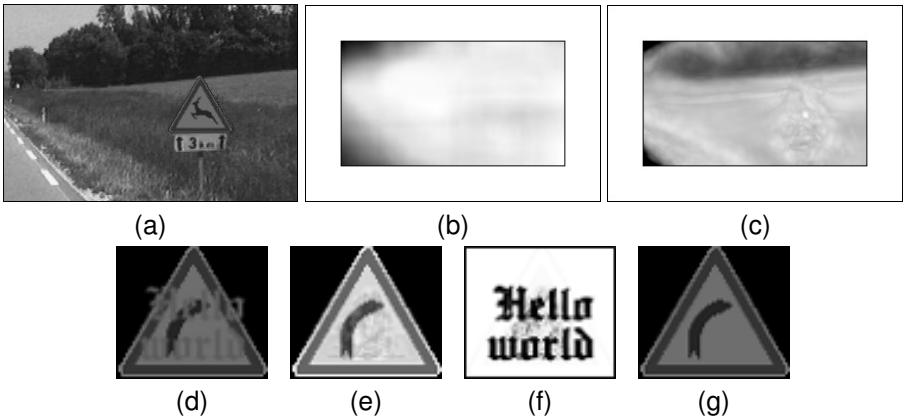

FIGURE 1.6 – Ligne du haut : expérience de détection. (a) Image analysée. Cartes de vraisemblance pour des hypothèses de bruit et d'a priori (respectivement) : (b) gaussien-gaussien et (c) robuste-non gaussien. Les plus fortes vraisemblances apparaissent en clair. Ligne du bas : expérience de reconnaissance. (d) image analysée. (e) images reconstruite. (f) carte des données erronées, b. (g) panneau reconnu.

Nous avons montré [DCH04] que l'association d'un modèle robuste de distribution du bruit, $\mathcal{P}(p|f,\mathcal{B})$, et d'un modèle a priori, $\mathcal{P}(f|\mathcal{B})$, bien adapté permet d'améliorer de manière significative les performances des systèmes de détection existant. La figure 1.6 illustre cela sur un exemple synthétique. La base d'images d'apprentissage contient 43 signaux de danger, sous 36 angles de rotation chacun. On ne retient que 30 vecteurs propres. La méthode faisant référence jusqu'à présent, utilisant des hypothèses gaussiennes [MP97], est mise en défaut sur cette image. Une meilleure discrimination est obtenue en associant l'estimation robuste à une modélisation a priori adaptée aux données. Dans cet exemple, on connaît la forme analytique de cette dernière distribution, mais nous avons également proposé une méthode, basée sur l'algorithme *mean shift* [Che95], permettant de gérer des formes arbitraires de distributions [VHC03].

La reconnaissance, quant à elle, est effectuée en comparant les coordonnées dans l'espace propre estimées, $\hat{f}$, à celles des panneaux d'apprentissage. Ici encore, l'utilisation d'un estimateur robuste améliore les performances du système [Dah01], comme le montre la figure 1.6. Dans cette expérience, le panneau est en partie caché par une inscription jaune. Le panneau reconnu à partir de l'estimée au sens du MV gaussien est le panneau temporaire à fond jaune, alors que le panneau reconnu dans le cas robuste est correct.

1.4.3 Reconstruction régularisée

Nous nous intéressons ici aux problèmes de reconstruction sous le modèle génératif linéaire. Il s'agit d'un problème inverse mal posé. Existence et unicité de la solution peuvent être obtenues respectivement en calculant la solution au sens des moindres carrés et en prenant la solution de norme minimale : on parle alors d'inversion généralisée. En décomposant R en valeurs singulières, $R = UDV^T$, on obtient :

$$\hat{f}^+ = R^+ p \qquad \text{où} \quad R^+ = VD^+U^T, \tag{1.4.5}$$

D^+ étant la matrice diagonale contenant l'inverse des valeurs singulières non nulles de R, et des 0 aux positions des valeurs singulières nulles. On résout ainsi le problème

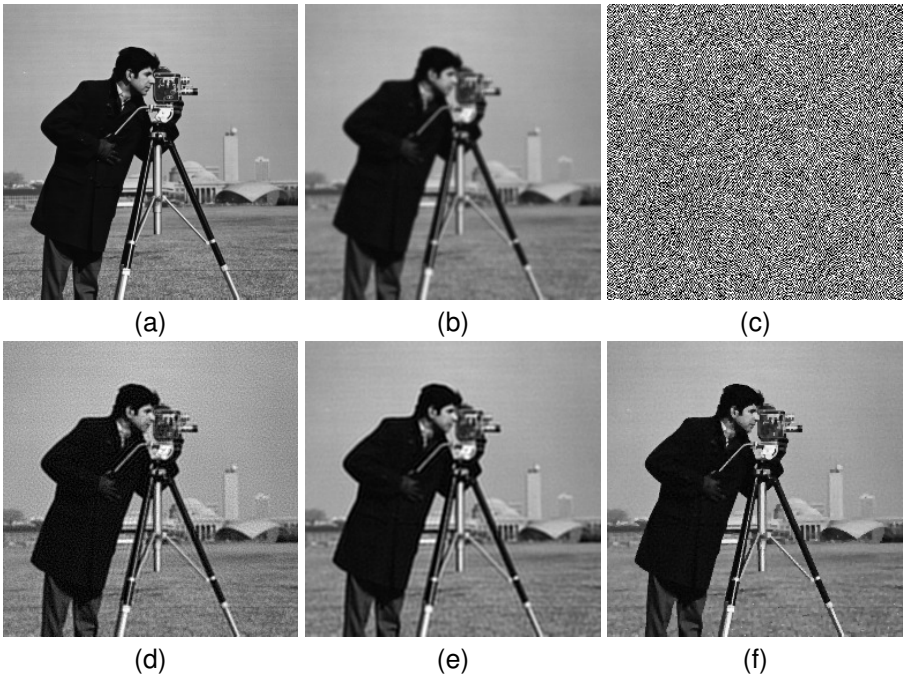

FIGURE 1.7 – (a) Image originale “cameraman”. (b) Image floue (défocalisation de rayon 2 pixels) et bruitée. (c) Inverse généralisée (1.4.5). (d) Algorithme de Landweber, 100 itérations (1.4.6), $SNR_{var} = 15,7dB$. (e) Régularisation de Tikhonov (1.4.7), $\lambda = 2$, $SNR_{var} = 14,9dB$. (f) Régularisation semi-quadratique ICM-DCT, $\lambda = 2, \delta = 15$, 100 itérations, $SNR_{var} = 18,4dB$.

de l’inversion de (1.2.1) mais pas celui de la stabilité en présence de bruit (1.2.2). En effet la matrice R est toujours mal conditionnée et l’inversion de ses valeurs singulières faibles conduit à une amplification des hautes fréquences du bruit, qui peuvent même dominer la solution (cf. figure 1.4.3). Il est possible d’y remédier en éliminant les valeurs singulières faibles ou bien en calculant l’inverse généralisée par une descente de gradient sur le critère des moindres carrés (algorithme de Landweber) :

$$\hat{f}_{LW}^{(k+1)} = \hat{f}_{LW}^{(k)} + \gamma R^T(p - R\hat{f}_{LW}^{(k)}), \tag{1.4.6}$$

et en arrêtant la solution après un nombre fini d’itérations. La régularisation de Tikhonov, évoquée au paragraphe 1.2.2, consiste à ajouter une contrainte quadratique sur la solution ou ses dérivées. Dans le cas d’une régularisation sur le gradient, la solution vérifie les équations normales :

$$(R^T R - \lambda^2 \Delta)\hat{f}_T = R^T p, \tag{1.4.7}$$

où Δ est l’opérateur Laplacien. On peut montrer que ces 3 solutions correspondent chacune à une façon de filtrer les valeurs singulières faibles. On parle de régularisation linéaire. Son défaut est d’agir globalement sur les images et de ne pas respecter les fortes variations locales, correspondant aux contours (cf. figure 1.4.3).

Afin de préserver les discontinuités, on remplace le terme quadratique de régularisation par un terme de la forme (1.3.1). En variables continues, le critère régularisé s’écrit alors :

$$J(f) = \|p - Rf\|^2 + \lambda^2 \int_\Omega \varphi(|\nabla f|) \tag{1.4.8}$$

où Ω représente le domaine image. Le lecteur pourra se référer à [BF00] pour une

étude en variables continues en lien avec la théorie des équations aux dérivées partielles (EDP). Nous utiliserons ici, comme dans [Cha94], une formulation discrète séparable :

$$J(f) = \|p - Rf\|^2 + \lambda^2 \sum_i \varphi\left((D_x f)_j\right) + \lambda^2 \sum_i \varphi\left((D_y f)_j\right). \tag{1.4.9}$$

Par rapport à la formulation continue, ce critère a le défaut de ne pas être isotrope. Par contre, l'aspect séparable peut être intéressant pour l'implantation. Le critère J est étendu à deux variables selon l'un ou l'autre des développements semi-quadratiques et une stratégie déterministe de minimisations alternées par rapport à chaque variable est mise en œuvre. Chaque étape de l'algorithme comprend donc le calcul des variables auxiliaires $b_x^{(k+1)}$ et $b_y^{(k+1)}$ selon les expressions (1.3.7) et (1.3.8) et la résolution des équations normales. Dans le cas multiplicatif (algorithme ARTUR), elles s'écrivent :

$$(R^T R - \lambda^2 \Delta_A^{(k+1)}) \hat{f}_A^{(k+1)} = R^T p, \tag{1.4.10}$$

où :

$$\Delta_A^{(k+1)} = -D_x^T B_x^{(k+1)} D_x - D_y^T B_y^{(k+1)} D_y,$$

avec $B_x^{(k+1)} = diag(b_x^{(k+1)})$ et $B_y^{(k+1)} = diag(b_y^{(k+1)})$. Dans le cas additif (algorithme LEGEND), on résout :

$$(R^T R - \lambda^2 \Delta) \hat{f}_L^{(k+1)} = R^T p + \lambda^2 D_x^T b_x^{(k+1)} + \lambda^2 D_y^T b_y^{(k+1)} \tag{1.4.11}$$

La première forme converge en général plus rapidement que la seconde, en nombre d'itérations. Par contre, cette dernière est intéressante car le premier terme ne varie pas au cours des itérations k. On peut, par exemple, le factoriser par l'algorithme de Cholesky pour faciliter ensuite les calculs [Cha94]. Dans le cas de la déconvolution, avec conditions aux bords circulaires, la transformée de Fourier permet de diagonaliser ce terme. On peut cependant lui préférer la transformée en cosinus discret lorsque la PSF est symétrique. Cela sous-entend des conditions aux bords symétriques, correspondant aux conditions de Neumann associées aux équations normales. On obtient ainsi l'algorithme ICM-DCT [Jal01].

Les figures 1.4.3 et 1.8 présentent les résultats obtenus en déconvolution à l'aide de l'algorithme ICM-DCT, en comparaison d'estimateurs classiques. Le gain en qualité, évident visuellement, est confirmé en termes de rapport signal-sur-bruit (SNR). Celui-ci est mesuré à partir des variances de l'image de référence, connue, f et de l'image reconstruite $\hat{f}$: $SNR_{VAR} = 10 \log_{10}(\sigma_f^2 / \sigma_{f-\hat{f}}^2)$. Les filtres utilisés ici modélisent un flou de défocalisation dans le premier cas et un bougé dans le second.

La régularisation semi-quadratique a été appliquée avec succès à de nombreux problèmes de reconstruction comme, entre autres, la stéréovision, l'estimation de mouvement, la tomographie ou encore l'imagerie micro-ondes (diffraction inverse) [BF00].

1.5 Conclusion

L'hypothèse classique de distribution gaussienne (correspondant à une énergie quadratique) s'avère inadaptée dans de nombreux problèmes réels où les données se composent en réalité de deux catégories ("inlier/outlier"), nécessitant chacune un traitement adapté. On fait alors appel à l'estimation robuste combinée à la régularisation lorsque le problème est mal posé. Nous avons présenté un cadre général, la théorie

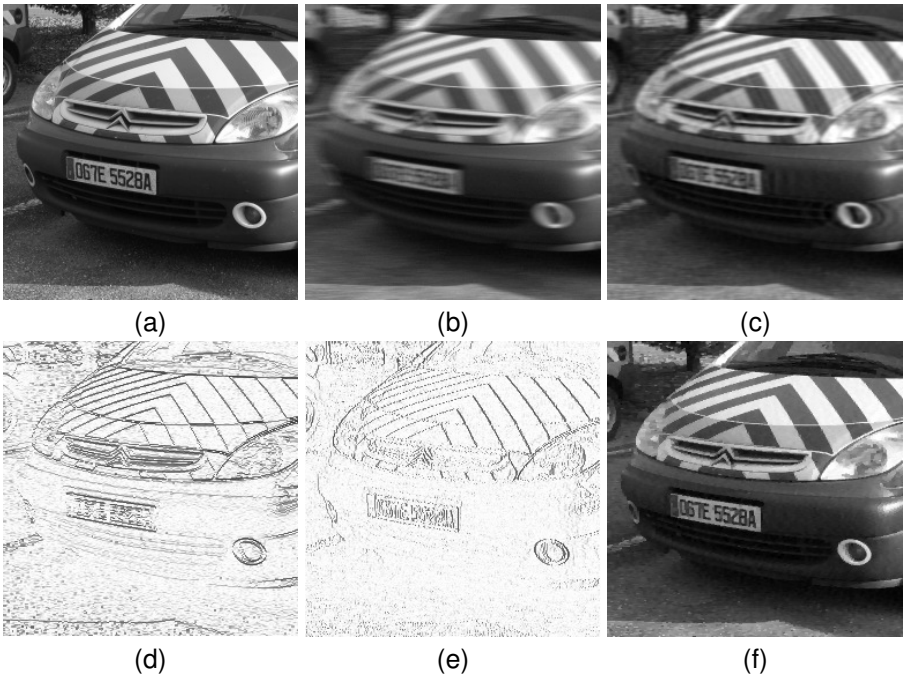

FIGURE 1.8 – (a) Image originale “ecodyn”. (b) Image floue (bougé horizontal de 12 pixels) et bruitée. (c) Régularisation de Tikhonov, $\lambda = 1$, $SNR_{var} = 14,5dB$. (d-e) Discontinuités entre lignes, b_x et entre colonnes b_y (visualisation en log). (f) Reconstruction régularisée semi-quadratique par ICM-DCT, $\lambda = 1, \delta = 7$, 100 itérations, $SNR_{var} = 18,1dB$, où la plaque minéralogique est de nouveau lisible.

semi-quadratique, permettant de sortir du comportement purement quadratique, soit par l’introduction d’une fonction bien choisie qui atténue l’influence des fortes déviations, soit par l’introduction directe de variables auxiliaires et d’un terme de pénalité énergétique associé. La théorie semi-quadratique permet d’établir le lien entre ces deux approches et précise les conditions de choix possibles, en termes simples. De plus, ceci conduit le plus souvent à des algorithmes simples. Cette théorie se situe à la confluence des statistiques robustes [Hub81], des champs markoviens [GG84], des approches variationnelles [AK02], et de l’analyse convexe. C’est ce qui en fait son succès, même si elle n’est pas toujours identifiée en tant que telle. Elle permet en effet de poser et de résoudre de multiples problèmes pratiques, comme nous l’avons illustré par des exemples : la détection et le suivi des marquages routiers, la reconnaissance des panneaux, et l’amélioration de la qualité d’image.

Bibliographie

[AK02] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing : Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2002.

[All02] M. Allain. *Approche pénalisée en tomographie hélicoïdale en vue de l'application à la conception d'une prothèse personnalisée du genou*. PhD thesis, Université Paris Sud - Paris XI, France, 2002.

[BF00] L. Blanc-Féraud. *Sur quelques Problèmes Inverses en Traitement d'Image*. Habilitation à diriger des recherches, Université de Nice-Sophia Antipolis, France, July 2000.

[BFCAB95] L. Blanc-Féraud, P. Charbonnier, G. Aubert, and M. Barlaud. Nonlinear image processing : Modeling and fast algorithm for regularization with edge detection. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume I, pages 474–477, Washington D.C., USA, September 1995.

[BZ87] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.

[CBFAB97] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2) :298–311, February 1997. Prix IEEE 1998 du meilleur article de jeune auteur, section "Image and Multidimensional Signal Processing".

[Cha94] P. Charbonnier. *Reconstruction d'image : régularisation avec prise en compte des discontinuités*. PhD thesis, Université de Nice-Sophia Antipolis, France, 1994.

[Che95] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(8) :790–799, August 1995.

[Dah01] R. Dahyot. *Analyse d'images séquentielles de scènes routières par modèles d'apparence pour la gestion du réseau routier.* PhD thesis, Université Paris VI, France, 2001.

[DCH04] R. Dahyot, P. Charbonnier, and F. Heitz. Robust Bayesian detection using appearance-based models. *Pattern Analysis and Applications*, 7 :317–332, 2004. On-line, http ://www.springerlink.com.

[GG84] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(6) :721–741, November 1984.

[GR92] S. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(3) :367–383, March 1992.

[GY95] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization and FFT's. *IEEE Transactions on Image Processing*, 4(7) :932–946, July 1995.

[Hub81] P.J. Huber. *Robust statistics*. John Wiley and Sons, New York, 1981.

[Ien04] S. S. Ieng. *Méthodes robustes pour la détection et le suivi des marquages.* PhD thesis, Université Paris VI, France, 2004.

[Jal01] A. Jalobeanu. *Modèles, estimation bayesienne et algorithmes pour la déconvolution d'images aériennes et satellitaires*. PhD thesis, Université de Nice-Sophia Antipolis, France, 2001.

[Kay93] S. M. Kay. *Fundamentals of Statistical signal processing - Estimation Theory*. Prentice Hall International Editions, 1993.

[MP97] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(7) :696–710, July 1997.

[TIC02] J.P. Tarel, S.S. Ieng, and P. Charbonnier. Using robust estimation algorithms for tracking explicit curves. In Springer, editor, *6th European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science*, volume 2350, pages 492–407, Copenhague, Danemark, May 2002.

[TIC05] J.P. Tarel, S.S. Ieng, and P. Charbonnier. 2005. soumis à 9th European Conference on Computer Vision (ECCV), *blind review*.

[TP91] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1) :71–86, 1991.

[VHC03] T. Vik, F. Heitz, and P. Charbonnier. Mean shift-based Bayesian image reconstruction into visual subspace. In *Proceedings of the 2003 IEEE International Conference on Image Processing (ICIP 2003)*, Barcelona, Spain, September 2003.