INSA Toulouse
GMM 5ème année, Image

Cours 1 : Formulation variationnelle pour le traitement d'images

Introduction

Pour résoudre des problèmes de traitement d'images, on utilise souvent des algorithmes qui sont des suites d'opérations élémentaires. Certains algorithmes peuvent traiter une large gamme de problèmes. Dans ce chapitre, nous présentons un cadre permettant de traiter une grande variété de problèmes de traitement d'images. On définit une fonction F sur l'ensemble des images et à valeurs dans $\mathbb{R} \cup \{+\infty\}$ de telle sorte que le ou les minimiseurs de F sont des solutions ou en tout cas apportent une solution acceptable au problème que l'on s'est posé. Les algorithmes utilisés pour résoudre le problème auront alors pour but de minimiser la fonction F. Cette approche variationnelle en traitement d'images n'est évidemment pas la seule et certainement pas toujours la meilleure mais elle peut donner d'excellents résultats et se prête souvent à une analyse mathématique rigoureuse. De plus la fonction F permet facilement de comparer deux *solutions* au problème. Certaines propriétés de la fonction F telles que la convexité et la coercivité permettent d'assurer l'existence d'au moins une solution au problème tel qu'on se le pose. Dans une première partie, nous allons voir que l'approche bayésienne fournit des éléments de réflexion pour justifier une approche variationnelle. Un tel point de vue permet de comprendre comment on peut construire une fonction F adaptée à un problème donné. Dans la partie suivante nous verrons comment on peut choisir une régularisation. Puis nous verrons le lien entre ce que nous appellerons la régularisation et la parcimonie.

1 Le cadre Bayésien

On observe un vecteur y et on considère y comme la réalisation d'une variable aléatoire (VA) Y dont la loi dépend d'un objet d'intérêt x :

$$y = x + b \quad \text{ou} \quad y = h \star x + b \quad \text{ou plus généralement} \quad y = Ax + b \tag{1}$$

où b est un bruit, h un flou et A un opérateur de mesure (Tomographie, IRM) ou un opérateur de dégradation, flou ou masquage.
Si $b \sim \mathcal{N}(0, \sigma_1^2 I)$ alors $Y \sim \mathcal{N}(x, \sigma_1^2 I)$.

On note $f_x = f(\cdot|x)$ la densité de Y qui dépend donc de x. Une manière d'estimer x est de considérer que x suit une loi *a priori* Π, $x \sim \Pi$ par exemple $x \sim \mathcal{N}(0, \sigma_2^2 I)$.
Le fait de considérer l'objet d'intérêt comme la réalisation d'une variable aléatoire dont on connait la loi est le paradigme bayésien.
Un tel paradigme permet d'établir à partir des observations y, une loi *a posteriori* pour x en utilisant la loi de Bayes :

$$\Pi(x|Y = y) = \frac{f(Y = y|x)\Pi(x)}{\int_{u \in X} f(Y = y|u)\Pi(u)du} \propto f(Y = y|x)\Pi(x). \tag{2}$$

On peut alors estimer x par le maximum de cette loi *a posteriori*. (Exemples de loi *a posteriori* et de maximiseurs $\hat{x}$ dans le cas gaussien avec ou sans opérateurs). Ce n'est pas le seul choix, on peut aussi estimer x en utilisant une moyenne *a posteriori* ou d'autres paramètres de la loi, mais dans ce chapitre nous nous concentrons sur le choix de l'estimateur appelé Maximum A Posteriori (MAP).

Ce maximum $\hat{x}$ peut être vu comme le minimiseur d'une fonction qui est la somme de deux termes :

— un terme d'attache aux données,
— un terme de régularisation

Le terme d'attache aux données dépend essentiellement du modèle de bruit et du lien entre x et les observations y. Le terme de régularisation dépend de l'*a priori* que l'on fait sur x.

En traitement d'images on pose des problèmes d'optimisation en fonctions des trois éléments précédents en mettant souvent de côté l'aspect bayésien, qui peut servir ou non d'alibi pour le choix de la fonctionnelle à optimiser.

2 Choix de la régularisation

L'*a priori* gaussien sur le bruit est souvent acceptable mais l'*a priori* sur x est souvent hasardeux. le choix de la fonctionnelle à optimiser se fait essentiellement selon deux critères :

1. L'adéquation au modèle. (Exemples)
2. Le caractère résoluble du problème.

En effet il peut s'avérer inutile de poser un problème d'optimisation insoluble, on préfèrera souvent l'approcher par un problème proche que l'on sait résoudre.

Un problème peut être résoluble en plusieurs sens :

— Il existe une solution explicite

1. $$F(x) = \frac{1}{2}\|y - x\|_2^2 + \lambda\|x\|_2^2, \tag{3}$$
2. $$F(x) = \frac{1}{2}\|y - h \star x\|_2^2 + \lambda\|x\|_2^2, \tag{4}$$
3. $$F(x) = \frac{1}{2}\|y - x\|_2^2 + \lambda\|x\|_1, \tag{5}$$
4. $$F(x) = \frac{1}{2}\|y - x\|_2^2 + \lambda\|x\|_0, \text{ où } \|x\|_0 = \text{Cardinal}\{i \text{ tels que } x(i) \neq 0\} \tag{6}$$
5. $$F(x) = \frac{1}{2}\|y - x\|_2^2 + \lambda\|Tx\|_1, \text{ où } T \text{ est orthogonal.} \tag{7}$$

— Il existe un algorithme d'optimisation

1. $$F(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_2^2, \tag{8}$$
2. $$F(x) = \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1 \tag{9}$$
3. $$F(x) = \frac{1}{2}\|y - Ax\|_1 + \lambda\|x\|_1, \tag{10}$$

d'une manière plus générale, si le problème est convexe, il existe souvent un algorithme pour le résoudre. Si le problème est séparable aussi.

Il existe des problèmes d'optimisation qu'on se sait pas résoudre parfaitement sans faire de recherche exhaustive comme

$$F(x) = \frac{1}{2}\|y - Ax\|^2 + \lambda \|x\|_0, \tag{11}$$

Quand on est confronté à ce genre de problème, on a essentiellement deux manières de procéder : soit on utilise des algorithmes qui vont trouver de bonnes solutions à ce problème, sans garanties qu'ils trouveront la meilleure solution. C'est le cas de la plupart des algorithmes dits *gloutons ou greedy*. Soit on modifie la fonction F de manière à ce que les minimiseurs de la nouvelle fonction soient proches des minimiseurs de F. Dans le cas de la fonction précédentes les algorithmes Matching Pursuit (MP) et Orthogonal Matching Pursuit (OMP) peuvent fournir des vecteurs x tels que $F(x)$ soit assez bas (des minima locaux) et le remplacement de la pseudo norme ℓ_0 par la norme ℓ_1 permet de rendre le problème convexe. Et dans la pratique, les solutions de ce nouveau problème convexe sont très souvent parcimonieuses et permettent d'atteindre des minima locaux de F (couplé à du debiasing).

3 Régularisation et parcimonie

Un des intérêts de la régularisation est de stabiliser un problème qui peut être mal posé.

$$\min \|y - Ax\|_2^2 \Longrightarrow \min \|y - Ax\|_2^2 + \lambda \|x\|_1 \tag{12}$$

Le premier problème peut admettre une infinité de solutions si A n'est pas injectif alors que le second peut n'admettre qu'une seule solution même si A n'est pas injectif et même être relativement stable aux perturbations.

Elle permet aussi de combler une perte d'information due à un opérateur de dégradation (convolution, inpainting, Radon partiel). Le fait de minimiser une fonction de la forme

$$F(x) = \frac{1}{2}\|y - h \star x\|^2 + \lambda R(x), \tag{13}$$

permet de trouver parmi les x tels que $h \star x$ est proche de y au sens de la norme ℓ^2 un vecteur qui a une faible valeur $R(x)$. Le paramètre λ permet de définir le poids que l'on alloue à chacune des quantités. Plus λ est grand plus on pénalise R, plus $R(\hat{x})$ est petit et plus $\|y - h \star x\|_2^2$ sera grand, et réciproquement.

La régularisation oriente l'inversion vers les vecteurs qui ont une faible valeur $R(x)$. Si on choisit correctement la régularisation, on peut reconstruire parfaitement un vecteur x à partir de $y = Ax$ dans le cas où A n'est pas injectif.

Certaines hypothèses sur x, telles la régularité par morceaux ou l'adéquation à un modèle, peuvent se traduire par une parcimonie dans une base : i.e $\|Tx\|_0$ est faible.

On peut vouloir essayer de minimiser

$$F(x) = \frac{1}{2}\|y - Ax\|^2 + \lambda \|Tx\|_0, \tag{14}$$

pour résoudre un problème de déconvolution, de séparation de sources quand on sait que Tx est parcimonieux.

Malheureusement, ce problème fait partie des problèmes non convexes pour lesquels on ne dispose pas de méthodes de résolution exacte. Comme nous l'avons déjà indiqué, certains algorithmes greedy donnent des solutions raisonnables mais rarement des garanties de trouver un minimiseur.

Une vaste littérature montre que sous des hypothèses raisonnables, la minimisation d'une fonctionnelle où la régularisation ℓ^0, ou pénalité ℓ^0 a été remplacée par une régularisation ℓ^1 permet d'obtenir une solution parcimonieuse. Ainsi si on sait que Tx où x est l'objet d'intérêt, a peu de coefficients non nuls, alors le minimiseur $\hat{x}$ de la fonctionnelle suivante

$$F(x) = \frac{1}{2} \|y - Ax\|^2 + \lambda \|Tx\|_1 , \tag{15}$$

peut être proche de x.

D'une manière générale la minimisation d'une fonctionnelle F avec une régularisation de la forme $R(x) = \|Lx\|_1$ produit souvent un vecteur $\hat{x}$ tel que Lx est parcimonieux pour un opérateur linéaire L quelconque.

Ainsi, si on sait qu'une image est *sparse* (parcimonieuse) dans un base d'ondelettes, et qu'on observe x via un opérateur M de masquage i.e $y = Mx + b$, on peut estimer x en minimisant une fonctionnelle de la forme :

$$F(x) = \frac{1}{2} \|y - Mx\|^2 + \lambda \|Tx\|_1 , \tag{16}$$

où T est une transformée en ondelettes. La valeur de λ étant ajustée en fonction du niveau de bruit.

De même, si on sait qu'un signal est *sparse* dans un dictionnaire D (donné ou créé à partir d'une base de données), et qu'on observe x via un opérateur A, i.e. $y = Ax + b$, on peut estimer x en minimisant

$$F(x) = \frac{1}{2} \|y - Ax\|^2 + \lambda \|Dx\|_1 , \tag{17}$$

On peut appliquer ceci à la séparation de sources :
Si on dispose d'une image $y = S_1 + S_2$ somme de deux sources chacune parcimonieuse via une transformée T_1 et T_2. Ainsi, on peut écrire $y = T_1 x_1 + T_2 x_2$ où x_1 et x_2 sont parcimonieux et on cherche x_1 et x_2.
Si on pose $T = [T_1, T_2]$ on cherche alors $x = [x_1, x_2]^t$ parcimonieux tel que Tx est parcimonieux.
On voudrais alors minimiser

$$F(x) = \frac{1}{2} \|y - Tx\|^2 + \lambda \|x\|_0 , \tag{18}$$

On peut le faire par algorithme greedy (MP) ou en remplaçant la pénalité ℓ^0 par une pénalité ℓ^1.
On peut utiliser la même approche si $y = A(S_1 + S_2)$.

Pour aller plus loin, voici une liste non exhaustive de problèmes pouvant être écrits comme des problèmes d'optimisation, en plus des problèmes de déconvolution, d'inpainting et de séparation de sources.

— Transport optimal.
— Transfert de couleur.
— Estimation de mouvement dans une vidéo.
— Classification.