

Faculté Polytechnique



Optimisation Non Linéaire

Récupération d'une image floutée (deblurring)

Projet d'Optimisation

Sabrine RIAHI
Aymerick SOYEZ



Sous la direction de Monsieur Nicolas GILLIS et Arnaud VANDAELE

Décembre 2017

Table des matières

1	Introduction	2
2	Étude de la convexité du problème	2
3	Le problème admet-il un minimum global ?	2
4	Conditions d'optimalité	3
5	Méthode de descente de coordonnées	4
6	Méthode du gradient	5
7	Comparaison des méthodes	6
8	Étude de la sensibilité de la solution	6

1 Introduction

Le problème posé est de déflouter une image dont chaque pixel a été remplacé par une combinaison linéaire des pixels voisins. La matrice de floutage utilisée est donnée. De plus, un bruit gaussien a été ajouté dans l'image.

L'objectif de ce projet est donc la résolution du problème suivant :

$$\min_{0 \leq x \leq 1} \|Ax - \tilde{x}\|_2^2 + \lambda \|x\|_2^2$$

où :

A est la matrice de floutage

\tilde{x} est le vecteur de pixels flouté

λ est un paramètre positif qui dépend du niveau de bruit

2 Étude de la convexité du problème

Pour qu'un problème soit convexe, il faut que

- Le problème soit sous forme de minimisation,
- Son domaine D soit convexe,
- $\forall x, y \in D, \forall \lambda \in [0; 1], f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$

Le domaine est décrit par $D = \{x \mid c(x) \geq 0\}$ et est convexe si $c(x)$ est concave.

Ici : $D = \{x \mid 0 \leq x \leq 1\}$

On a donc deux inégalités $x \geq 0$ et $1 - x \geq 0$. Ces deux fonctions étant linéaires, elles sont à la fois concaves et convexes, donc ces deux « sous-domaines » sont convexes.

Puisque l'intersection de deux ensembles convexes est un ensemble convexe, l'ensemble D est également convexe.

La norme 2 est une fonction convexe.

Preuve

En utilisant les propriétés de la norme euclidienne :

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &= \|\lambda x + (1 - \lambda)y\|_2^2 \\ &\leq \lambda \|x\|_2^2 + (1 - \lambda) \|y\|_2^2 \\ &= \lambda f(x) + (1 - \lambda)f(y) \end{aligned}$$

On a donc bien une fonction convexe, et on peut conclure que le problème est convexe.

3 Le problème admet-il un minimum global ?

Le domaine du problème est compact (fermé et borné). De plus, la fonction est convexe sur ce domaine. On aura donc soit un ou plusieurs minima locaux à l'intérieur du domaine, soit un minimum local sur l'une de ses bornes.

Le problème étant convexe, tout minimum local est global.

On conclut ainsi que le problème admet un minimum global.

4 Conditions d'optimalité

Conditions d'optimalité de Karush-Kuhn-Tucker

Si x^* est une solution optimale locale telle que l'ensemble des gradients des contraintes actives est linéairement indépendant, alors il existe un vecteur de multiplicateurs de Lagrange α^* tel que :

$$\begin{aligned}\nabla_x \mathcal{L}(x^*, \alpha^*) &= 0 \\ \alpha_i^* &\geq 0 \\ \alpha_i^* c_i(x^*) &= 0\end{aligned}$$

Nous disposons du problème de minimisation suivant :

$$\min_{0 \leq x \leq 1} \|Ax - \tilde{x}\|_2^2 + \lambda \|x\|_2^2$$

En reformulant les contraintes comme suit :

$$\begin{cases} x_i &\geq 0 \\ -x_i + 1 &\geq 0 \end{cases}$$

Et en tenant compte de la définition de la fonction Lagrangienne :

$$\mathcal{L}(x, \alpha) = f(x) - \sum \alpha_i c_i(x)$$

On obtient le Lagrangien de notre problème d'optimisation :

$$\mathcal{L}(x_i, \alpha_i) = \|Ax_i - \tilde{x}_i\|_2^2 + \lambda_i \|x_i\|_2^2 - \sum \alpha_i x_i - \sum \alpha_i (-x_i + 1)$$

En tenant compte des conditions KKT, nous avons finalement :

$$\nabla_x \mathcal{L}(x_i, \alpha_i) = 0 \Leftrightarrow \nabla(\|Ax_i - \tilde{x}_i\|_2^2 + \lambda_i \|x_i\|_2^2 - \sum \alpha_i x_i - \sum \alpha_i (-x_i + 1)) = 0$$

$$\alpha_i \geq 0 \Leftrightarrow \alpha_i \geq 0$$

$$\alpha_i c_i(x_i) = 0 \Leftrightarrow \begin{cases} \alpha_i x_i &= 0 \\ \alpha_i (-x_i + 1) &= 0 \end{cases}$$

$$\begin{aligned}x_i &\geq 0 \\ -x_i + 1 &\geq 0\end{aligned}$$

Pour que les conditions d'optimalité de KKT soient nécessaires et suffisantes, il faut que le problème soit convexe, et qu'il admette un point de Slater.

Il a été démontré plus haut que le problème étudié est convexe. Étant donné qu'il nous est également possible de trouver un point de Slater respectant les contraintes du problème, les conditions d'optimalité de KKT sont donc nécessaires et suffisantes.

5 Méthode de descente de coordonnées

La méthode de descente de coordonnées consiste à minimiser la fonction successivement sur chaque variable séparément en gardant les autres variables fixées.

Pour cela, on choisit un point de départ, qui sera, dans le cadre de ce projet, l'image floutée.

On calcule ensuite, pour chaque variable, la dérivée par rapport à celle-ci, que l'on égalise à zéro. La valeur obtenue remplacera la précédente pour les prochaines itérations.

On effectue ces opérations jusqu'à obtenir une convergence des valeurs obtenues pour chaque variable, c'est à dire jusqu'à ce que l'écart entre deux itérations soit inférieur à une précision fixée.

La fonction que nous devons minimiser est quadratique. Mettons-la sous la forme

$$f(x) = \frac{1}{2}x^T Q x + c^T x + b$$

$$\begin{aligned} \|Ax - \tilde{x}\|_2^2 + \lambda \|x\|_2^2 &= (Ax - \tilde{x})^T (Ax - \tilde{x}) + \lambda x^T x \\ &= (Ax)^T (Ax) - \tilde{x}^T (Ax) - (Ax)^T \tilde{x} + \tilde{x}^T \tilde{x} + \lambda x^T x \\ &= x^T A^T A x - 2\tilde{x}^T A x + \tilde{x}^T \tilde{x} + x^T (\lambda I) x \\ &= x^T (A^T A + \lambda I) x - 2\tilde{x}^T A x + \tilde{x}^T \tilde{x} \end{aligned}$$

On retrouve bien la forme ci-dessus, avec
$$\begin{cases} Q &= 2(A^T A + \lambda I) \\ c &= -2A^T \tilde{x} \\ b &= \tilde{x}^T \tilde{x} \end{cases}$$

Pour calculer la formule de l'actualisation d'une variable, considérons une fonction quadratique en 3 dimensions :

$$f(x_1, x_2, x_3) = \frac{1}{2} \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} c_1 & c_2 & c_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Si on veut actualiser x_2 , on calcule la dérivée de f en cette variable :

$$\frac{\delta f(x_1, x_2, x_3)}{\delta x_2} = q_{21}x_1 + q_{22}x_2 + q_{23}x_3 + c_2$$

La nouvelle valeur de x_2 est donc donnée par :
$$x_2 = \frac{-(q_{21}x_1 + q_{23}x_3 + c_2)}{q_{22}}$$

Pour une fonction à n dimensions, en la variable x_i :

$$\begin{aligned} x_i &= \frac{-\left(\sum_{\substack{j=1 \\ j \neq i}}^n q_{ij}x_j + c_i\right)}{q_{ii}} \\ &= x_i - \frac{\sum_{j=1}^n q_{ij}x_j + c_i}{q_{ii}} \\ &= x_i - \frac{[\nabla f]_i}{q_{ii}} \end{aligned}$$

Pour respecter la contrainte $0 \leq x \leq 1$, il suffit d'utiliser $\min(\max(\dots, 0), 1)$, où " \dots " est la formule d'actualisation ci-dessus.

Nous avons également utilisé un test d'arrêt qui prend en compte la norme de l'écart entre les deux dernières itérations ainsi que le nombre d'itérations.
L'implémentation sous Matlab se trouve en annexes.

6 Méthode du gradient

La méthode du gradient est similaire à la méthode de descente de coordonnées, en prenant comme direction de descente la direction de plus grande pente.
Afin de s'assurer que l'objectif diminue, on détermine un pas optimal pour chaque itération.

Une itération sera donnée par :

$$X_{k+1} = X_k - \alpha \nabla f(X_k) \quad (1)$$

Pour résoudre le problème, nous le reformulerons sous la forme quadratique, tout comme pour la méthode du gradient, ce qui nous donnera également :

$$f(x) = \frac{1}{2}x^T Q x + c^T x + b \quad (2)$$

Avec :

$$\begin{aligned} Q &= 2(A^T A + \lambda I) \\ c &= -2A^T \tilde{x} \\ b &= \tilde{x}^T \tilde{x} \end{aligned}$$

Le gradient de la fonction est donc égal à :

$$\nabla f(x) = Qx + c \quad (3)$$

Déterminer un pas optimal α pour obtenir une itération $k+1$ revient donc à résoudre l'équation :

$$\frac{\partial f(X_{k+1})}{\partial \alpha} = 0$$

En considérant les équations (1), (2) et (3), et en posant $\nabla f(X_k) = g$, nous pouvons déterminer le pas optimal α pour chaque itération :

$$\begin{aligned} f(X_{k+1}) &= \frac{1}{2}X_{k+1}^T Q X_{k+1} + c^T X_{k+1} + b \\ &= \frac{1}{2}(X_k - \alpha g)^T Q (X_k - \alpha g) + c^T (X_k - \alpha g) + b \\ &= \frac{1}{2}(X_k^T - \alpha g^T) Q (X_k - \alpha g) + c^T (X_k - \alpha g) + b \\ &= \frac{1}{2}(X_k^T Q X_k - \alpha X_k^T Q g - \alpha g^T Q X_k + \alpha^2 g^T Q g) + c^T X_k - \alpha c^T g + b \\ &= \frac{1}{2}X_k^T Q X_k - \alpha X_k^T Q g + \frac{1}{2}\alpha^2 g^T Q g + c^T X_k - \alpha c^T g + b \end{aligned}$$

En prenant $\frac{\partial f(X_{k+1})}{\partial \alpha} = 0$, on obtient :

$$-QX_k^T g + \alpha g^T Qg - c^T g = 0$$

$$\alpha = \frac{c^T g + QX_k^T g}{g^T Qg} = \frac{(c^T + QX_k^T)g}{g^T Qg} = \frac{g^T g}{g^T Qg}$$

L'implémentation sous Matlab se trouve en annexe.

7 Comparaison des méthodes

Dans le tableau suivant reprenant l'ensemble des erreurs pour diverses valeurs de λ (limité à 3000 itérations), on peut voir que la méthode du gradient est plus efficace pour l'Exemple1, tandis que la méthode de descente de coordonnées est plus efficace pour la première et la dernière image.

Nous avons également observé les différences au niveau du temps de calcul et du nombre d'itérations pour chaque méthode : (limité à 100 000 itérations, sauf la nuit)

Méthode	Exemple	Nombre d'itérations	Temps de calcul	Erreur
Gradient	0	1 000 000	20m	0.93 %
	1	100 000	45m	11.80 %
	2	212 924	?	1.05 %
Descente de coordonnées	0	1 159	< 1m	0.93 %
	1	44 314	4h	14.00 %
	2	2 900	20m	1.05 %

La méthode de descente de coordonnées nécessite moins d'itérations, mais il faut plus de temps pour les exécuter.

8 Étude de la sensibilité de la solution

Nous avons déflouté en boucle les différents exemples pour des valeurs de λ différentes avec une limite de 3000 itérations. Le tableau suivant montre l'ensemble de nos résultats, pour les deux méthodes.

Pour la première image, on voit que l'erreur diminue constamment avec λ , en se stabilisant petit à petit (λ tend en fait vers 0).

Pour la seconde image, l'erreur est minimale autour de $\lambda = 10^{-4}$ avec la méthode de descente de coordonnées, mais se stabilise après $\lambda = 5 \cdot 10^{-7}$ avec la méthode du gradient.

L'erreur de la dernière image diminue également progressivement jusqu'à se stabiliser après $\lambda = 10^{-7}$.

Méthode	λ	Erreur		
		Example0	Example1	Example2
Descente de coordonnées	1,00E-02	47,14617916	10,1647664	30,72163516
	5,00E-03	38,66568078	9,169684144	26,8902945
	1,00E-03	21,3446921	7,167769831	18,05490584
	5,00E-04	15,28981797	6,453585874	14,62302629
	1,00E-04	5,446041966	6,048189766	7,727587915
	5,00E-05	3,155836022	6,416526121	5,293370035
	1,00E-05	1,250600831	9,522410563	1,868970892
	5,00E-06	1,058235509	10,95138518	1,375975318
	1,00E-06	0,934905051	12,60119727	1,050864016
	5,00E-07	0,921463922	12,84824921	1,017104882
	1,00E-07	0,911350647	13,04974679	0,990150346
	5,00E-08	0,910149684	13,07534406	0,98675349
	1,00E-08	0,909188837	13,09592037	0,984038304
	5,00E-09	0,90907554	13,09849993	0,983700172
	1,00E-09	0,908979452	13,10056444	0,983429913
	5,00E-10	0,908967441	13,10082255	0,98339619
	1,00E-10	0,908957833	13,10102905	0,98336921
	5,00E-11	0,908956632	13,10105487	0,983365838
Gradient	1,00E-02	47,14785368	10,1657138	30,72596525
	5,00E-03	38,66362675	9,169837985	26,89259855
	1,00E-03	21,341731	7,163593487	18,05553437
	5,00E-04	15,29204685	6,476905571	14,63874942
	1,00E-04	5,835811798	5,465286598	8,818225907
	5,00E-05	3,909591126	5,340744131	7,547420843
	1,00E-05	2,24333251	5,2763741	6,386730072
	5,00E-06	2,026906195	5,272683027	6,234823196
	1,00E-06	1,84494184	5,271115461	6,11199903
	5,00E-07	1,821932862	5,271049484	6,096575705
	1,00E-07	1,803569205	5,271017599	6,084220985
	5,00E-08	1,801278858	5,271013461	6,082674497
	1,00E-08	1,799434918	5,271010282	6,08143701
	5,00E-09	1,799204933	5,271009887	6,081282306
	1,00E-09	1,799020708	5,271009571	6,081158539
	5,00E-10	1,798997653	5,271009531	6,081143068
	1,00E-10	1,798979228	5,271009499	6,081130692
	5,00E-11	1,798976924	5,271009495	6,081129144