

Optimisation bi-convexe

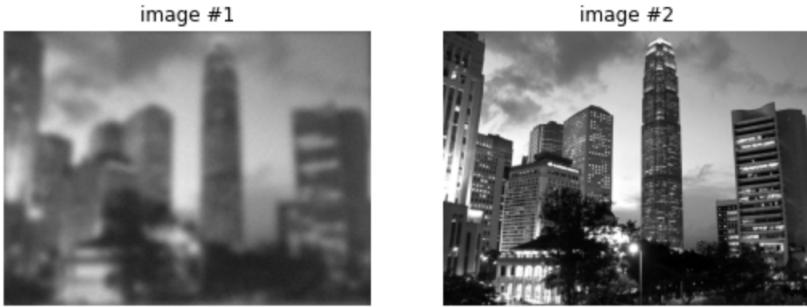
Cecile Della Valle

March 9, 2021

Introduction



Erreur $|x_{\text{pred}} - x_{\text{true}}|_2 : 0.2682$



Erreur $|x_{\text{pred}} - x_{\text{true}}|_2 : 0.2804$

Figure 1: Sur la première ligne, l'image 1 correspond à l'image convoluée. Sur la deuxième ligne, l'image 1 correspond à l'image convoluée et bruitée (avec une std de 0.05, le bruit est à peine visible à l'oeil nu sur une image convoluée). L'image 2 est toujours l'image de référence. C'est à partir des images de la première colonne que l'on va initialiser les algorithmes, et on espère obtenir une erreur finale meilleure que respectivement 26.82% sans bruit et 28.04% avec bruit.

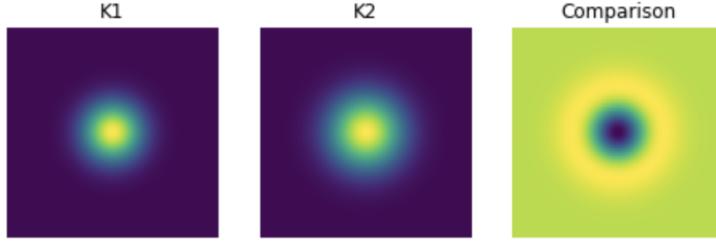
On introduit le problème d'optimisation biconvexe-concave de déconvolution aveugle d'images avec l'a priori de noyau de convolution régulier,

$$\min_{u \in U, K \in \Sigma} \sup_v \frac{1}{2} \|g - K * u\|^2 + \frac{1}{2} \alpha \|DK\|^2 + \langle \nabla u, v \rangle - \chi_{B,\mu}(v) .$$

L'espace Σ correspond à l'ensemble des simplexes (ensemble des noyaux à coefficient positif dont la somme des coefficients vaut 1). Par extension on notera Σ la fonction indicatrice de cet ensemble. L'espace U correspond à l'hypercube $[0, 1]^{N_x \times N_y}$, ou la boule unité pour la norme L^∞ . Sa fonction indicatrice correspond donc à $\chi_{B,1}$.

La fonction à minimiser peut alors se réécrire

$$\min_{u, K} \sup_v \frac{1}{2} \|g - K * u\|^2 + \chi_{B,1}(u) + \frac{1}{2} \alpha \|DK\|^2 + \Sigma(K) + \langle \nabla u, v \rangle - \chi_{B,\mu}(v) .$$



Erreur $|K_{\text{pred}} - K_{\text{true}}|_2 : 0.2811$

Figure 2: Le noyau exact K_1 et un noyau proche K_2 pour l'initialisation de l'algorithme (de même centre mais un peu plus large). L'erreur initiale sur le noyau s'élève donc à environ 28%.

On note

$$L(u, v, K) = \frac{1}{2} \|g - K * u\|^2 + \frac{1}{2} \alpha \|DK\|^2 + \langle \nabla u, v \rangle - \chi_{B, \mu}(v)$$

On souhaite comparer deux stratégies pour résoudre ce problème :

- Minimisation alternée : on scinde la fonction à minimiser en une partie qui dépend de l'image u et une partie qui dépend du noyau K , les deux fonctions sont minimisées jusqu'à convergence, alternativement.
- Minimisation par pas de descente alterné (PDA) : l'algorithme alterne les pas de descente entre l'image u , le noyau K et la variable auxiliaire (ou duale) v .

Pour chacune de ces deux méthodes, on effectue les trois tests suivants :

- Initialisation par le noyau exacte, bruit std = 0.05, $(\alpha, \mu) = (0, 5)$ (cas 1);
- Initialisation par un noyau approché, pas de bruit, $(\alpha, \mu) = (1, 0.5)$ (cas 2) ;
- Initialisation par un noyau approché, bruit std = 0.05, $(\alpha, \mu) = (1, 5)$ (cas 3) ;

Dans le tableau , on indique les erreurs finales des deux algorithmes (avec même minimisation et mêmes paramètres).

| Erreur relative | cas 1 | | cas 2 | | cas 3 | |
|-----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| | noyau K | image u | noyau K | image u | noyau K | image u |
| Initial | 0.0 | 0.2804 | 0.2811 | 0.2682 | 0.2811 | 0.2804 |
| alterné | 0.0026 | 0.2436 | 0.2477 | 0.2491 | 0.2477 | 0.2532 |
| PDA | 0.0001 | 0.2513 | 0.0069 | 0.2667 | 0.2787 | 0.2606 |

Remarque: J'ai retiré la projection sur Simplex finalement car les algorithmes ne marchaient pas bien. En fait le résultat de la projection donnait des noyaux irréguliers. A la place j'ai projeté sur un hypercube avec une valeur max et une valeur min, cela avait l'avantage d'être simple à implémenter, d'avoir un bon comportement et de rester cohérent avec le problème initial (?).

1 Minimisation alternée

On introduit les deux sous problèmes suivants :

$$E_1(K) = \frac{1}{2} \|g - K * u\|^2 + \frac{1}{2} \alpha \|DK\|^2 + \Sigma(K) \quad (1)$$

$$E_2(u) = \frac{1}{2} \|g - K * u\|^2 + \mu \text{TV}(u) \quad (2)$$

Dans la minimisation alternée, on considère d'abord la minimisation de E_2 pour reconstruire l'image u . La minimisation est obtenue par un algorithme de Chambolle-Pock. Puis on minimise E_1 pour reconstruire le noyau K avec un forward-backward. On alterne 5 fois entre chaque minimisation, chacune à 200 pas d'itérations, donc en fait on n'est pas complètement à convergence..

2 Pas de descente alterné (PDA)

Pour résoudre ce problème d'optimisation on se propose de tester l'algorithme suivant,

$$K_{k+1} = \text{prox}_{\Sigma}(K_k - \gamma \partial_K L(u_k, K_k, v_k)) \quad (3)$$

$$u_{k+1} = \text{proj}_{B_1}(u_k - \tau \partial_u L(u_k, K_{k+1}, v_k)) \quad (4)$$

$$v_{k+1} = \text{prox}_{B_\mu}(v_k + \sigma \partial_v L(\bar{u}_k, K_{k+1}, v_k)) \quad (5)$$

$$\bar{u}_{k+1} = u_{k+1} + \theta(u_{k+1} - u_k) \quad (\text{relaxation}) \quad (6)$$

On écrit alors les dérivées partielles de L ,

$$\partial_K L(u_k, K_k, v_k) = -(g - K_k * u_k) * \tilde{u}_k + \alpha D^T D K_k \quad (7)$$

$$\partial_u L(u_k, K_{k+1}, v_k) = \text{div} \cdot v_k - \tilde{K}_k * (g - K_k * u_k) \quad (8)$$

$$\partial_v L(\bar{u}_k, K_{k+1}, v_k) = \nabla \bar{u}_k \quad (9)$$

Pour les opérateurs ∇ et \div j'ai repris les opérateurs discrets du document 120415-deconvolution que tu avais écrit.

On trace les énergies primales et duales au cours de la minimisation

$$Ep_K = \frac{1}{2} \|g - K * u\|^2 + \frac{1}{2} \alpha \|DK\|^2 \quad (10)$$

$$Ep_x = \frac{1}{2} \|g - K * u\|^2 + \mu \cdot \text{TV}(u) \quad (11)$$

$$Ed = -\langle \nabla u, v \rangle \quad (12)$$

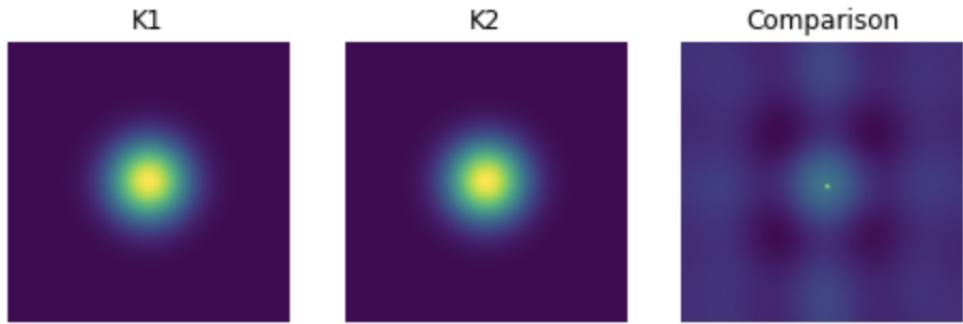
L'algorithme est très sensible aux pas de descente choisi. Soit ℓ_1 le coefficient de Lipschitz de E_1 , le pas de descente pour le noyau K est choisi par

$$\tau = c \cdot \frac{2}{\ell_1} \leq c \cdot \frac{1}{\|u\|^2 + \alpha \|D\|^2} \quad 0 < c \leq 1 .$$

En effet on ne veut pas dépasser la valeur limite du pas de descente pour assurer la convergence de la descente de gradient, mais on ne veut pas nécessairement effectuer un pas optimal non plus, par exemple si le noyau est initialisé de façon exacte. On choisit donc la stratégie suivante, à chaque pas de temps le coefficient c est multiplié par 0.98. La valeur initiale est ajustée à la main. Cela permet notamment de choisir un c supérieur à 1. En effet j'ai choisi cette stratégie quand le noyau est très peu modifié par rapport à l'image, cela améliore nettement la convergence. Pour le pas de descente de l'image et de la variable auxiliaire cette fois, on respecte la condition de convergence de Chambolle-Pock

$$\sigma_u \sigma_v < \frac{1}{L^2} \leq \left(\frac{1}{\sqrt{8} + \|k\|^2} \right)^2 .$$

J'ai testé plusieurs stratégies, mais mettre le pas "optimal" marchait mieux jusqu'à présent.



Erreur $|K_{pred} - K_{true}|_2 : 0.0026$



Erreur $|x_{pred} - x_{true}|_2 : 0.2436$

Initialisation juste, bruit 0.05

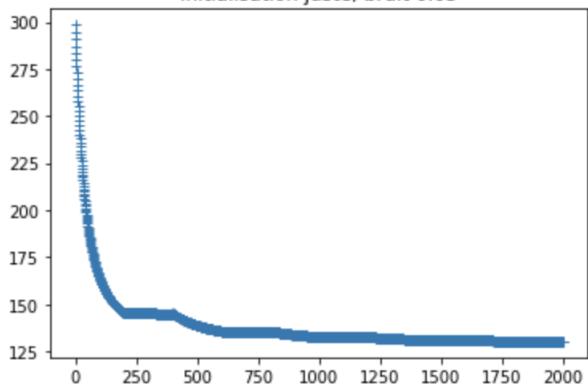
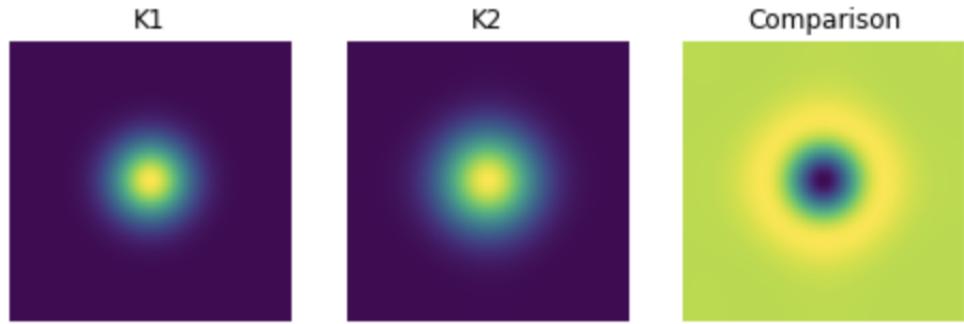


Figure 3: Algorithme de minimisation alternée. Reconstruction correspondant au cas 1 (initialisation noyau exacte, bruit std 0.05). Paramètres : $(\alpha, \mu) = (0, 5)$



Erreur $|K_{pred} - K_{true}|_2 : 0.2744$



Erreur $|x_{pred} - x_{true}|_2 : 0.2491$

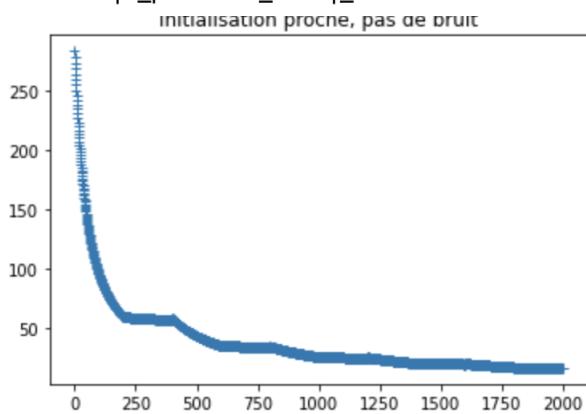
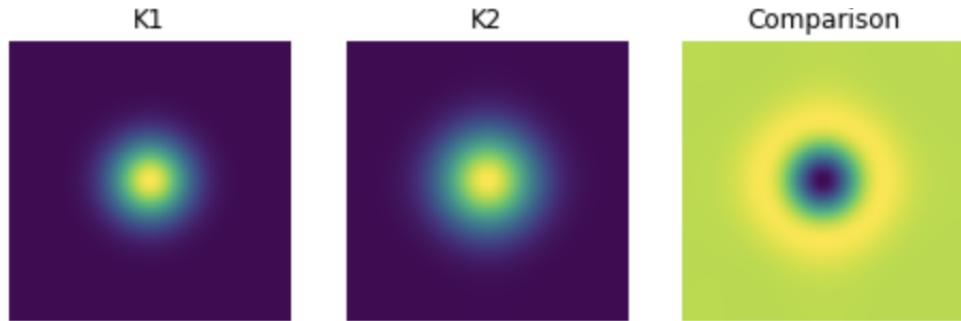


Figure 4: Algorithme de minimisation alternée. Reconstruction correspondant au cas 2 (initialisation noyau proche, pas de bruit). Paramètres : $(\alpha, \mu) = (1, 0.5)$



Erreur $|K_{\text{pred}} - K_{\text{true}}|_2 : 0.2744$



Erreur $|x_{\text{pred}} - x_{\text{true}}|_2 : 0.2532$

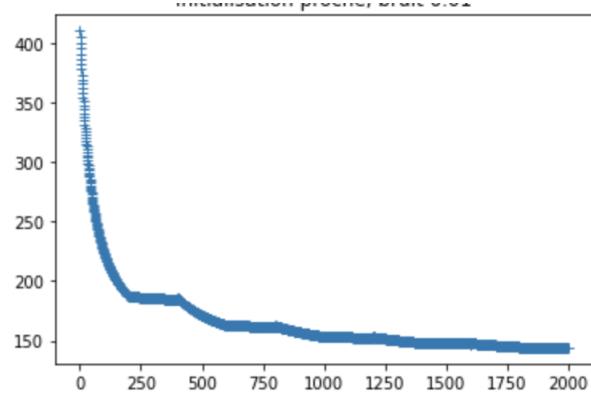
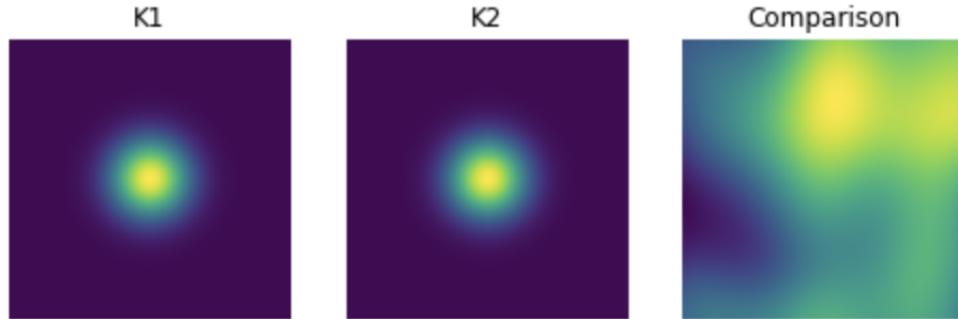


Figure 5: Algorithme de minimisation alterné. Reconstruction correspondant au cas 3 (initialisation noyau proche, bruit std 0.05). Paramètres : $(\alpha, \mu) = (1, 5)$



Erreur $|K_{\text{pred}} - K_{\text{true}}|_2 : 0.0001$



Erreur $|x_{\text{pred}} - x_{\text{true}}|_2 : 0.2513$

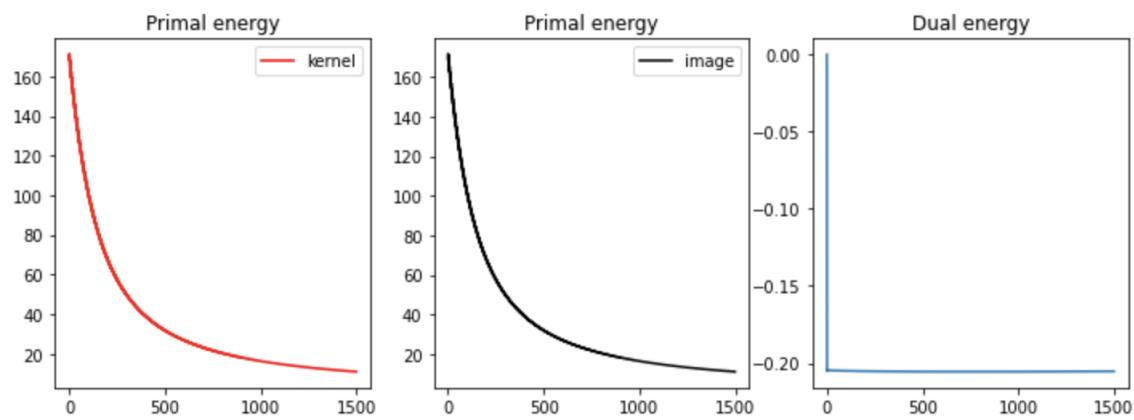
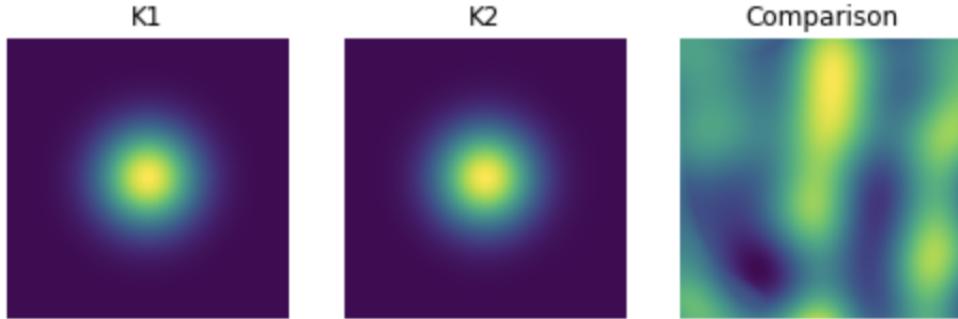


Figure 6: Algorithme de pas de descente alterné PDA. Reconstruction correspondant au cas 1 (initialisation noyau exacte, bruit std 0.05). Paramètres : $(\alpha, \mu) = (0, 5)$



Erreurs $|K_{\text{pred}} - K_{\text{true}}|_2 : 0.0069$

image #1

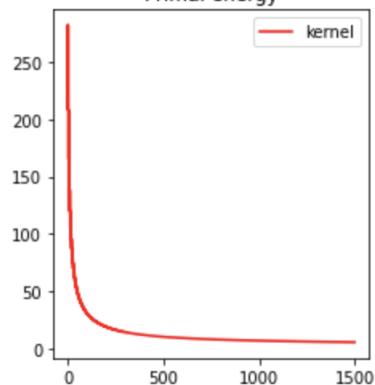


image #2



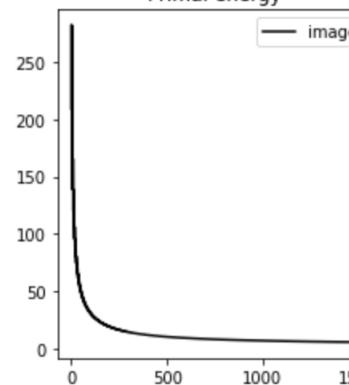
Erreurs $|x_{\text{pred}} - x_{\text{true}}|_2 : 0.2667$

Primal energy



Primal energy

image



Dual energy

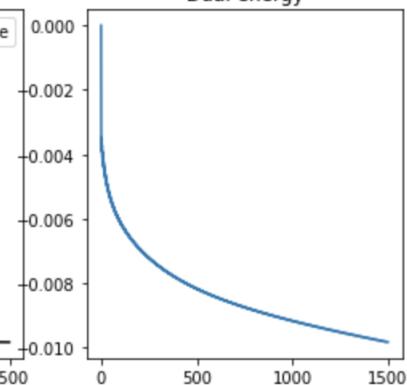
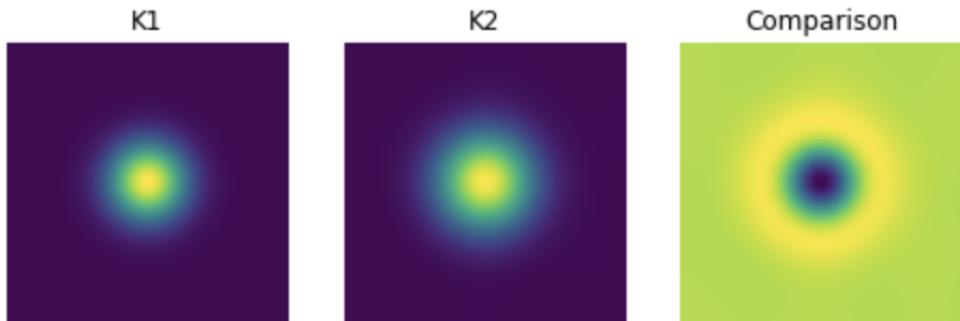


Figure 7: Algorithme de pas de descente alterné PDA. Reconstruction correspondant au cas 2 (initialisation noyau proche, pas de bruit). Paramètres : $(\alpha, \mu) = (1, 0.5)$



Erreurs $|K_{pred} - K_{true}|_2 : 0.2787$



Erreurs $\|x_{pred} - x_{true}\|_2 : 0.2606$

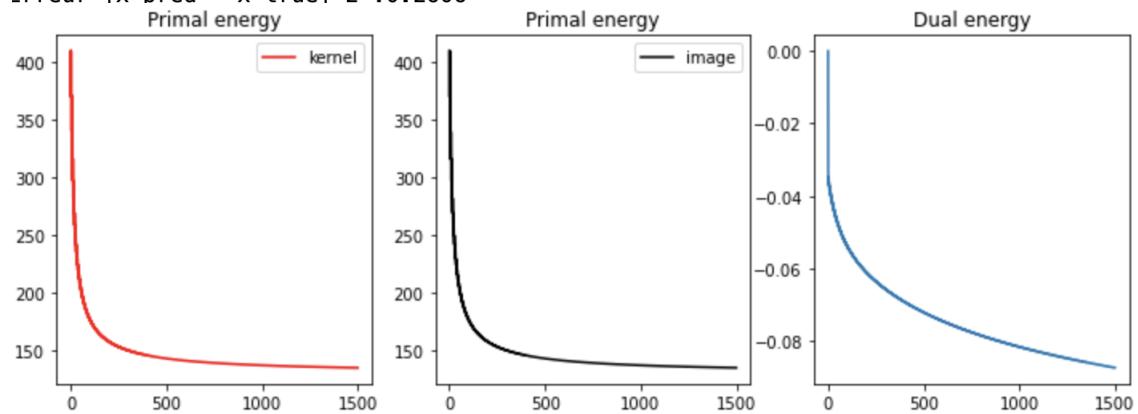


Figure 8: Algorithme de pas de descente alterné PDA. Reconstruction correspondant au cas 3 (initialisation noyau proche, bruit std 0.05). Paramètres : $(\alpha, \mu) = (1, 5)$