Méthodes numériques : optimisation.
L3 2015–2016 — 2$^{\text{e}}$ semestre.
Feuille de TD n$^{\text{o}}$ 3 : Méthode du gradient conjugué — Éléments de correction.

1 Étude de fonction quadratique abstraite.

(a) On a $\mathcal{A}(e_j) = \sum_k a_{kj} e_k$. Et donc en prenant le produit scalaire avec e_i on obtient $\langle e_i, \mathcal{A}(e_j)\rangle_E = a_{ij}$. Si on a $\langle x, \mathcal{A}(y)\rangle_E = \langle y, \mathcal{A}(x)\rangle_E$ pour x et y dans E, on l'applique à e_i et e_j pour obtenir que $a_{ij} = a_{ji}$. Réciproquement si on a $a_{ij} = a_{ji}$ on obtient que $\langle e_i, \mathcal{A}(e_j)\rangle_E = \langle e_j, \mathcal{A}(e_i)\rangle_E$ (pour tous les i et j), et par linéarité de $\mathcal{A}$ et du produit scalaire par rapport à la première variable, si $x = \sum_i x_i e_i$, on obtient que $\langle x, \mathcal{A}(e_j)\rangle_E = \langle e_j, \mathcal{A}(x)\rangle_E$. De même, par linéarité on obtient que si $y = \sum_j y_j e_j$, alors $\langle x, \mathcal{A}(y)\rangle_E = \langle y, \mathcal{A}(x)\rangle_E$.
On écrit alors

$$f(x+h) = f(x) + \frac{1}{2}[\langle h, \mathcal{A}(x)\rangle_E + \langle \mathcal{A}(h), x\rangle_E] + \langle b, h\rangle_E + \frac{1}{2}\langle h, \mathcal{A}(h)\rangle_E.$$

Par la propriété de symétrie on obtient donc que le terme linéaire est égal à $\langle \mathcal{A}(x) + b, h\rangle_E$, et on a bien que $|\langle h, \mathcal{A}(h)\rangle_E| \leqslant \|h\|_E \|\mathcal{A}(h)\|_E$, et comme on est en dimension finie, l'application linéaire $\mathcal{A}$ est continue. On obtient donc bien que $\|\mathcal{A}(h)\|_E \to 0$ lorsque $\|h\|_E \to 0$, ce qui nous donne bien que $\frac{1}{2}\langle h, \mathcal{A}(h)\rangle_E = o(\|h\|_E)$. La fonction f est donc bien différentiable en x et $\langle \nabla f(x), h\rangle_E = \langle \mathcal{A}(x) + b, h\rangle_E$ pour tout $h \in E$, ce qui donne bien le résultat.

(b) On a donc (en prenant $h = -t\,r$ dans le calcul précédent) :

$$f(x - t\,r) = \frac{1}{2}\langle r, \mathcal{A}(r)\rangle_E t^2 - \langle \mathcal{A}(x) + b, r\rangle_E t + f(x).$$

C'est un polynôme du second degré de coefficient dominant $\frac{1}{2}\langle r, \mathcal{A}(r)\rangle_E > 0$, il admet donc un unique minimum en $t = \frac{\langle \mathcal{A}(x)+b, r\rangle_E}{\langle r, \mathcal{A}(r)\rangle_E} = \frac{\langle r, r\rangle_E}{\langle r, \mathcal{A}(r)\rangle_E}$.

(c) La relation de récurrence est donc $x_{k+1} = x_k - \alpha_k r_k$ avec $\alpha_k = \frac{\langle r, r\rangle_E}{\langle r, \mathcal{A}(r)\rangle_E}$. On peut donc en déduire que $r_{k+1} = \mathcal{A}(x_k - \alpha_k r_k) + b$, et donc par linéarité de $\mathcal{A}$ on obtient $r_{k+1} = r_k - \alpha_k \mathcal{A}(r_k)$.
À chaque étape on garde seulement en mémoire les valeurs de x_k et r_k.

On calcule donc seulement $\mathcal{A}(r_k)$, que l'on utilise pour calculer α_k (ce qui nécessite deux calculs de produits scalaires). On calcule ensuite x_{k+1} puis r_{k+1} en utilisant la valeur déjà calculée de $\mathcal{A}(r_k)$, ce qui permet de passer à l'étape suivante.

2 Problème d'application : inpainting.

(a) On obtient que $x_{i,j} = u_{i,j} - f_{i,j}$ si $m_{i,j} = 1$ et $x_{i,j} = u_{i,j}$ si $m_{i,j} = 0$. Donc on obtient que si $U \in C$ alors $x_{i,j} = 0$ dès que $m_{i,j} = 1$ et réciproquement si $x_{i,j} = 0$ dès que $m_{i,j} = 1$ alors $U \in C$. Donc $U \in C$ équivaut à ce que $X \in E$ défini par

$$E = \{E \in M_{n_1,n_2}(\mathbb{R}) | \forall (i,j) \text{ tels que } m_{i,j} = 1, \text{ on a } x_{i,j} = 0\}.$$

On voit bien que E est un espace vectoriel. Sa dimension est égale au nombre de coefficients $x_{i,j}$ que l'on peut choisir arbitrairement (ceux pour lesquels on n'a pas la contrainte $x_{i,j} = 0$), c'est à dire le nombre d'indices (i,j) tels que $m_{i,j} = 0$. Ou de manière équivalente ceux pour lesquels $\overline{m}_{i,j} = 1$. Comme les coefficients de $\overline{M}$ valent 0 ou 1, la dimension de E est donc la somme des coefficients de $\overline{M}$.

(b) On minimise $\frac{1}{2}\varphi_2(U,U)$ pour $U \in C$, soit $\frac{1}{2}\varphi_2(X + F \circledast M, X + F \circledast M)$ pour $X \in E$. Par bilinéarité (et symétrie) de φ_2, on cherche donc à minimiser $\frac{1}{2}\varphi_2(X,X) + \varphi_2(X, F \circledast M) + \frac{1}{2}\varphi_2(F \circledast M, F \circledast M)$. Le dernier terme de dépend pas de X, et donc cela revient à minimiser $\frac{1}{2}\varphi_2(X,X) + \varphi_1(X)$, pour $X \in E$, où $\varphi_1(X) = \varphi_2(X, F \circledast M)$, donc φ_1 est bien une forme linéaire $E \to \mathbb{R}$.

On a $\varphi_2(X,X) = \sum_{i=1}^{n_1-1}\sum_{j=0}^{n_2-1}(x_{i,j} - x_{i-1,j})^2 + \sum_{i=0}^{n_1-1}\sum_{j=1}^{n_2-1}(x_{i,j} - x_{i,j-1})^2$. Donc $\varphi_2(X,X) \geqslant 0$, et si $\varphi_2(X,X) = 0$, on obtient que $x_{i,j} = x_{i-1,j}$ dès que $i \neq 0$, donc par récurrence sur i, on a donc que pour tout i, $x_{i,j} = x_{0,j}$. De même $x_{i,j} = x_{i,j-1}$ dès que $j \neq 0$, et donc par récurrence $x_{0,j} = x_{0,0}$ pour tout j. Au final, pour i, j quelconques, on a donc $x_{i,j} = x_{0,0}$ (tous les coefficients sont égaux).

Enfin si on suppose que $M \neq 0$, alors il existe i_0, j_0 tel que $m_{i_0,j_0} = 1$. Donc si on a $X \in E$ avec $\varphi_2(X,X) = 0$, on obtient $x_{i_0,j_0} = 0 = x_{0,0}$, et donc tous les coefficients sont nuls, donc $X = 0$. Donc si $X \in E \setminus \{0\}$, on a bien que $\varphi_2(X,X) > 0$.

(c) On utilise le résultat suivant : dans un espace euclidien (ici E muni de son produit scalaire), toute forme linéaire φ s'écrit de manière unique comme le produit scalaire avec un vecteur.

Il existe donc $B \in E$ tel que $\varphi_1(X) = \langle B, X\rangle_E$ pour tout $X \in E$. De même pour tout Y la fonction $X \mapsto \varphi_2(X,Y)$ est une forme linéaire, et donc il existe un unique vecteur de E, que l'on note $\mathcal{A}(Y)$ tel que $\varphi_2(X,Y) = \langle \mathcal{A}(Y), X\rangle_E$. En utilisant les propriétés de bilinéarité et de symétrie de φ_2 on obtient donc que pour $Y, Y' \in E$ et $\lambda \in \mathbb{R}$, on a

$$\begin{aligned}\langle \mathcal{A}(Y+\lambda Y'), X\rangle_E &= \varphi_2(X, Y+\lambda Y') = \varphi_2(Y+\lambda Y', X)\\ &= \langle \mathcal{A}(X), Y+\lambda Y'\rangle_E = \langle \mathcal{A}(X), Y\rangle + \lambda\langle \mathcal{A}(X), Y'\rangle_E\\ &= \varphi_2(X,Y)+\lambda\varphi_2(X,Y') = \varphi_2(Y,X)+\lambda\varphi_2(Y',X)\\ &= \langle \mathcal{A}(Y), X\rangle_E + \lambda\langle \mathcal{A}(Y'), X\rangle_E\\ &= \langle \mathcal{A}(Y)+\lambda\mathcal{A}(Y'), X\rangle_E\end{aligned}$$

Ceci étant valable pour tout $X \in E$, on a bien $\mathcal{A}(Y+\lambda Y') = \mathcal{A}(Y)+\lambda\mathcal{A}(Y')$, on obtient donc bien une application linéaire $\mathcal{A}$ qui vérifie les propriétés données. Si $M \neq 0$, comme $\varphi_2(X,X) > 0$ dès que $X \in E$ et $X \neq 0$, ceci donne que $\langle X, \mathcal{A}(X)\rangle_E > 0$, c'est à dire que la matrice correspondante (dans une base orthonormale de E) est symétrique définie positive, et que donc il y a bien une solution unique au problème de minimisation (dès que $M \neq 0$).

Pour stocker $\mathcal{A}$ sous forme de sa matrice dans une base orthonormale de E, on aurait donc une matrice de taille $N \times N$, où N est la dimension de E.

(d) Par exemple, si on imagine une image de 500×400 pixels, en considérant que la moitié des pixels est transmise, on obtient $N = 10^5$. On aurait donc à stocker $N \times N = 10^{10}$ coefficients. À raison de 8 octets par coefficients, cela correspondrait à 80 gigaoctets, simplement pour stocker une matrice A. Le calcul de AX nécessite également de faire $N \times N$ multiplications, ce qui est trop coûteux : en réalité il y a un grand nombre de zéros dans la matrice A, et certains calculs ne sont pas nécessaires.

On pose D_n la matrice $(n-1) \times n$ avec $d_{i,i} = -1$, $d_{i,i+1} = 1$ et $d_{i,j} = 0$ dès que $j \neq i$ et $j \neq i+1$.

Si on note $G = D_{n_1}X$, on obtient que $g_{i,j} = x_{i+1,j} - x_{i,j}$ pour $0 \leqslant i \leqslant n_1 - 2$ et $0 \leqslant j \leqslant n_2 - 1$.

De même si on note $H = D_{n_1}Y$, on obtient que $h_{i,j} = y_{i+1,j} - y_{i,j}$. Donc

$$\begin{aligned}\langle D_{n_1}X, D_{n_1}Y\rangle_{M_{n_1-1,n_2}} &= \sum_{i=0}^{n_1-2}\sum_{j=0}^{n_2-1}(x_{i+1,j}-x_{i,j})(y_{i+1,j}-y_{i,j})\\ &= \sum_{i=1}^{n_1-1}\sum_{j=0}^{n_2-1}(x_{i,j}-x_{i-1,j})(y_{i,j}-y_{i-1,j}).\end{aligned}$$

De même en posant $G' = XD_{n_2}^T$ et $H' = YD_{n_2}^T$, on obtient $g'_{i,j} = x_{i,j+1} - x_{i,j}$ et $h_{i,j} = y_{i,j+1} - y_{i,j}$ pour $0 \leqslant i \leqslant n_1 - 1$ et $0 \leqslant j \leqslant n_2 - 2$. Donc

$$\begin{aligned}\langle XD_{n_2}^T, YD_{n_2}^T\rangle_{M_{n_1,n_2-1}} &= \sum_{i=0}^{n_1-1}\sum_{j=0}^{n_2-2}(x_{i,j+1}-x_{i,j})(y_{i,j+1}-y_{i,j})\\ &= \sum_{i=0}^{n_1-1}\sum_{j=1}^{n_2-1}(x_{i,j}-x_{i,j-1})(y_{i,j}-y_{i,j-1}).\end{aligned}$$

On a donc bien

$$\varphi_2(X,Y) = \langle D_{n_1}X, D_{n_1}Y\rangle_{M_{n_1-1,n_2}} + \langle XD_{n_2}^T, YD_{n_2}^T\rangle_{M_{n_1,n_2-1}}.$$

On obtient donc

$$\begin{aligned}\varphi_2(X,Y) &= \mathrm{Tr}(X^T D_{n_1}^T D_{n_1} Y) + \mathrm{Tr}(D_{n_2} X^T Y D_{n_2}^T)\\ &= \mathrm{Tr}(X^T D_{n_1}^T D_{n_1} Y) + \mathrm{Tr}(X^T Y D_{n_2}^T D_{n_2})\\ &= \mathrm{Tr}(X^T (D_{n_1}^T D_{n_1} Y + Y D_{n_2}^T D_{n_2}) = \langle X, D_{n_1}^T D_{n_1} Y + Y D_{n_2}^T D_{n_2}\rangle.\end{aligned}$$

On serait tenté de vouloir poser $\mathcal{A}(Y) = D_{n_1}^T D_{n_1} Y + Y D_{n_2}^T D_{n_2}$, mais on n'est pas assuré que ceci définit une application $E \to E$.
Dès que $X \in E$, on a $x_{i,j} = \overline{m}_{i,j} x_{i,j}$ (si $\bar{m}_{i,j} = 0$, alors $m_{i,j} = 1$, et on a bien $x_{i,j} = 0$ et sinon $\overline{m}_{i,j} = 1$). Donc pour $Z \in M_{n_1,n_2}(\mathbb{R})$, on a en fait

$$\langle X, Z\rangle = \sum_{i,j} x_{i,j} \overline{m}_{i,j} z_{i,j} = \langle X, \overline{M} \circledast Z\rangle.$$

Et on a donc bien un élément $\overline{M} \circledast Z$ de E, tel que $\langle X, Z\rangle = \langle X, \overline{M} \circledast Z\rangle_E$ pour tout X dans E ($\overline{M} \circledast Z$ est en fait la projection orthogonale de Z sur E).
On a donc pour tout $X \in E$:

$$\varphi_2(X,Y) = \langle X, \overline{M} \circledast [D_{n_1}^T D_{n_1} Y + Y D_{n_2}^T D_{n_2}]\rangle.$$

En posant $\mathcal{A}(Y) = \overline{M} \circledast [D_{n_1}^T D_{n_1} Y + Y D_{n_2}^T D_{n_2}]$, ceci définit bien une application linéaire $E \to E$ telle qu'on ait $\varphi_2(X,Y) = \langle X, \mathcal{A}(Y)_E$ pour X et Y dans E.
De même comme on a $\varphi_1(X) = \varphi_2(X, M \circledast F)$, on obtient que pour $X \in E$, on a

$$\varphi_2(X, M \circledast F) = \langle X, \overline{M} \circledast [D_{n_1}^T D_{n_1}(M \circledast F) + (M \circledast F) D_{n_2}^T D_{n_2}]\rangle,$$

c'est à dire $\varphi_1(X) = \langle X, B\rangle_E$ avec

$$B = \overline{M} \circledast [D_{n_1}^T D_{n_1}(M \circledast F) + (M \circledast F) D_{n_2}^T D_{n_2}].$$

(e) On cherche à minimiser $\frac{1}{2}\langle X, \mathcal{A}(X)\rangle_E + \langle B, X\rangle_E$ pour $X \in E$. On peut donc, comme pour le premier exercice, écrire les itérations du gradient conjugué en utilisant seulement le produit scalaire $\langle \cdot, \cdot\rangle_E$, et la fonction $\mathcal{A}$. On obtient donc $R_0 = \mathcal{A}(X_0) + B$, $P_0 = -R_0$, on calcule $\langle R_0, R_0\rangle$ et ensuite à chaque étape
- on calcule $\mathcal{A}(P_k)$ (une évaluation de fonction),
- on utilise le calcul de $\langle R_k, R_k\rangle_E$ de l'étape précédente et le calcul précédent de $\mathcal{A}(P_k)$ pour calculer $\alpha_k = \frac{\langle R_k, R_k\rangle_E}{\langle P_k, \mathcal{A}(P_k)\rangle_E}$, ce qui nécéssite un calcul de produit scalaire de plus.
- on calcule $X_{k+1} = X_k + \alpha_k P_k$ et $R_{k+1} = R_k + \alpha_k \mathcal{A}(P_k)$ (grâce au calcul précédent de $\mathcal{A}(P_k)$),
- on calcule $\langle R_{k+1}, R_{k+1}\rangle_E$ que l'on garde en mémoire pour l'étape suivante (ce qui nécessite un produit scalaire), et on utilise le calcul de l'étape précédente pour calculer $P_{k+1} = -R_{k+1} + \frac{\langle R_k, R_k\rangle_E}{\langle R_{k+1}, R_{k+1}\rangle_E} P_k$.

On utilise donc par étape une seule évaluation de la fonction $\mathcal{A}$ et deux calculs de produits scalaires dans E.