Cours MAP434, «Contrôle de modèles dynamiques»
Séance 4, 6 mai 2015
OPTIMISATION ET ALGORITHMES

I - Descente de Gradient On considère une fonction F convexe définie sur un espace de Hilbert X. On suppose que ∇F est L-Lipschitz, et que F admet un minimiseur.
1. Montrer que pour tout $x, y \in X$,

$$F(x) + \nabla F(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2 \;\geq\; F(y) \tag{1}$$

Correction: Ecrire un développement de Taylor de $t \mapsto F(x + t(y - x))$.

2. On propose l'algorithme suivant : pour un point donné x, on appelle $y = T(x)$ le minimiseur de la fonction quadratique

$$y \;\mapsto\; F(x) + \nabla F(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2$$

qui majore F. $x^0 \in X$ étant donné on pose $x^n = T^n(x^0)$. Quel est cet algorithme ?

Correction: Le minimum est donné par $\nabla F(x) + L(y - x) = 0$ donc $y = x - (1/L)\nabla F(x)$: l'algorithme est une descente de Gradient à pas fixe $\mu = 1/L$.

3. Montrer que si Φ est α-convexe et x^* un minimiseur de Φ, alors pour tout x,

$$\Phi(x) \geq \Phi(x^*) + \frac{\alpha}{2}\|x - x^*\|^2$$

Correction: C'est la définition non ? On peut aussi écrire pour tout t

$$\Phi(x^*) \leq \Phi(tx^* + (1-t)x) \leq t\Phi(x^*) + (1-t)\Phi(x) - \alpha\frac{t(1-t)}{2}\|x - x^*\|^2$$

donc

$$(1-t)\Phi(x^*) \leq (1-t)\Phi(x) - \alpha\frac{t(1-t)}{2}\|x - x^*\|^2.$$

En divisant par $(1-t)$ et en faisant tendre t vers 1 on obtient l'inégalité. Ou encore simplement que

$$\Phi(x) \geq \Phi(x^*) + \nabla\Phi(x^*) \cdot (x - x^*) + \frac{\alpha}{2}\|x - x^*\|^2$$

et utiliser que $\nabla\Phi(x^*) = 0$ (mais l'autre preuve ne suppose pas Φ dérivable).

4. Montrer la formule suivante, vraie pour tous $x, y \in X$:

$$F(y) + \frac{L}{2}\|y - x\|^2 \geq F(Tx) + \frac{L}{2}\|y - Tx\|^2. \tag{2}$$

On rappelle que F étant convexe, $F(y) \geq F(x) + \nabla F(x) \cdot (y - x)$. On utilisera le fait que Tx est minimiseur d'un problème L-convexe, et l'inégalité (1).

Correction: On a

$$F(y) + \frac{L}{2}\|y-x\|^2 \geq F(x) + \nabla F(x)\cdot(y-x) + \frac{L}{2}\|y-x\|^2.$$

Or par définition, Tx minimise la fonction ci-dessus, qui est L-convexe. Donc d'après la question précédente,

$$F(y) + \frac{L}{2}\|y-x\|^2 \geq F(x) + \nabla F(x)\cdot(Tx-x) + \frac{L}{2}\|Tx-x\|^2 + \frac{L}{2}\|y-Tx\|^2.$$

Enfin, grâce à (1),

$$F(x) + \nabla F(x)\cdot(Tx-x) + \frac{L}{2}\|Tx-x\|^2 \geq F(Tx)$$

ce qui permet de conclure.

4. En prenant $x = y = x^n$ (et $Tx = x^{n+1}$) dans (2), vérifier que $F(x^n)$ décroit. En prenant $y = x^*$, un minimiseur de F dans (2), et $x = x^n$, $Tx = x^{n+1}$, montrer que

$$\frac{2}{L}(F(x^{n+1}) - F(x^*)) \;\leq\; \|x^n - x^*\|^2 - \|x^{n+1} - x^*\|^2\,. \tag{3}$$

En déduire que pour tout $k \geq 0$, $F(x^k) - F(x^*) \leq L\|x^0 - x^*\|^2/(2k)$. Que peut-on dire de ce type de convergence ?

Correction: Il faut sommer la dernière équation de $n = 0$ à $k-1$ pour conclure. C'est assez lent.

5. On suppose maintenant que F est aussi α-convexe : que devient (3) ? En déduire la convergence de x^n vers x^*, et la vitesse de convergence.

Correction: (a) Dans la preuve de (2) on a utilisé

$$F(y) \;\geq\; F(x) + \nabla F(x)\cdot(y-x)$$

qui est améliorée dans ce cas :

$$F(y) \;\geq\; F(x) + \nabla F(x)\cdot(y-x) + \frac{\alpha}{2}\|y-x\|^2\,.$$

Par conséquent, (2) devient :

$$F(y) + \frac{L}{2}\left(1 - \frac{\alpha}{L}\right)\|y-x\|^2 \geq F(Tx) + \frac{L}{2}\|y-Tx\|^2$$

(b) On a de plus $F(x^{n+1}) - F(x^*) \geq (\alpha/2)\|x^{n+1} - x^*\|^2$. Par conséquent, en reprenant $x = x^n, y = x^*$, l'équation (3) devient maintenant

$$\frac{\alpha}{L}\|x^{n+1} - x^*\|^2 \;\leq\; \left(1 - \frac{\alpha}{L}\right)\|x^n - x^*\|^2 - \|x^{n+1} - x^*\|^2\,,$$

donc $\|x^{n+1} - x^*\|^2 \leq \omega^2\|x^n - x^*\|^2$ avec $\omega = \sqrt{1-\alpha/L}/\sqrt{1+\alpha/L}$. Par conséquent, $\|x^n - x^*\| \leq \omega^n\|x^0 - x^*\|$.

6. "Gradient projeté" : on suppose que l'on s'intéresse maintenant au problème

$$\min_{x\in K} F(x)$$

où K est un convexe de X sur lequel "on sait projeter" (par exemple, $X = \mathbb{R}^N$ et $K = (\mathbb{R}_+)^N$). Répondre à la question **2.** (en rajoutant la contrainte $x \in K$ dans les minimisations) et montrer que (2) (question **4.**) est encore vraie lorsque $y \in K$. Qu'en déduit-on ?

Correction: L'algorithme est désormais

$$\begin{aligned}\min_{y\in K} F(x) + \nabla F(x)\cdot(y-x) + \frac{L}{2}\|x-y\|^2 \\ = F(x) - \frac{1}{2L}\|\nabla F(x)\|^2 + \min_{y\in K}\frac{L}{2}\left\|y - \left(x - \frac{1}{L}\nabla F(x)\right)\right\|^2\end{aligned}$$

et c'est donc un gradient projeté. De plus, la preuve de (2) est inchangée si on choisit $y \in K$. On montre à nouveau la même convergence de l'algorithme.

II - Double algorithme de gradient à pas fixe.

On considère une matrice A symétrique définie positive d'ordre n et un vecteur $b \in \mathbb{R}^n$. On introduit la fonctionnelle quadratique $J : \mathbb{R}^n \to \mathbb{R}$ telle que, pour tout $v \in \mathbb{R}^n$, $J(v) := \frac{1}{2}v^T Av - b^T v$. On considère une matrice rectangulaire C d'ordre $m \times n$ et un vecteur $d \in \mathbb{R}^m$. On suppose que $m < n$ et que la matrice C est de rang maximal. On rappelle (cf. TD3) que le problème de minimisation de J dans $K := \{v \in \mathbb{R}^n;\ Cv = d\}$ admet une et une seule solution $u \in \mathbb{R}^n$ et qu'il existe un et un seul vecteur $\lambda \in \mathbb{R}^m$ tel que le couple $(u, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ soit solution du système linéaire

$$Au + C^T\lambda = b, \tag{4a}$$
$$Cu = d, \tag{4b}$$

où C^T désigne la matrice transposée de C. L'objectif de cet exercice est d'étudier la convergence de l'algorithme d'approximation suivant : on se donne $(u^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m$, ainsi que deux paramètres positifs ρ_1, ρ_2, et pour tout $k \geq 0$, (u^k, λ^k) étant connu, on détermine (u^{k+1}, λ^{k+1}) par

$$u^{k+1} = u^k - \rho_1(Au^k - b + C^T\lambda^k), \tag{5a}$$
$$\lambda^{k+1} = \lambda^k + \rho_1\rho_2(Cu^{k+1} - d). \tag{5b}$$

1. Montrer l'identité suivante (I désigne la matrice identité de $\mathbb{R}^n$) :

$$\begin{aligned}\|\lambda^{k+1} - \lambda\|^2 = \|\lambda^k - \lambda\|^2 &+ \rho_1^2\rho_2^2\|C(u^{k+1} - u)\|^2 \\ &- 2\rho_2(u^{k+1} - u)\cdot\big(u^{k+1} - u - (I - \rho_1 A)(u^k - u)\big).\end{aligned}$$

(Indication : observer que $\lambda^{k+1} - \lambda = (\lambda^k - \lambda) + \rho_1\rho_2 C(u^{k+1} - u)$.)

Correction: L'indication se montre en utilisant (5b) et le fait que $Cu = d$. On élève chaque membre de l'égalité ainsi obtenue au carré, et pour le double produit, on constate que

$$\begin{aligned}\rho_1\rho_2 C(u^{k+1} - u)\cdot(\lambda^k - \lambda) &= \rho_1\rho_2(u^{k+1} - u)\cdot C^T(\lambda^k - \lambda) \\ &= -\rho_2(u^{k+1} - u)\cdot\big(u^{k+1} - u - (I - \rho_1 A)(u^k - u)\big),\end{aligned}$$

où on utilise $\rho_1 C^T\lambda = \rho_1(b - Au)$ et $\rho_1 C^T\lambda^k = u^k - u^{k+1} - \rho_1(Au^k - b)$ de par (5a).

2. On peut montrer que pour $\rho_1 > 0$ suffisamment petit, on a $\beta := \|I - \rho_1 A\| < 1$. Par la suite, la valeur de ρ_1 est fixée à une telle valeur. Montrer que pour $\rho_2 > 0$ suffisamment petit, il existe une constante $\gamma > 0$, indépendante de k, telle que

$$\gamma\|u^{k+1} - u\|^2 \leq \frac{1}{\rho_2}\|\lambda^k - \lambda\|^2 + \beta\|u^k - u\|^2 - \left(\frac{1}{\rho_2}\|\lambda^{k+1} - \lambda\|^2 + \beta\|u^{k+1} - u\|^2\right). \tag{6}$$

(Indication : observer que $\frac{1}{\rho_2}\|\lambda^k - \lambda\|^2 - \frac{1}{\rho_2}\|\lambda^{k+1} - \lambda\|^2 \geq 2\|u^{k+1} - u\|^2 - \rho_1^2\rho_2\|C(u^{k+1} - u)\|^2 - 2\beta\|u^{k+1} - u\|\|u^k - u\|$.) En déduire, pour un tel choix de ρ_2, que $\lim_{k\to\infty} u^k = u$.

Correction: L'indication se montre en divisant l'identité obtenue à la question 1 par ρ_2, en développant le dernier terme du membre de droite et en majorant la norme de la matrice $(I-\rho_1 A)$ par β. On utilise ensuite la fait que

$$-2\beta\|u^{k+1} - u\|\|u^k - u\| \geq -\beta\|u^{k+1} - u\|^2 - \beta\|u^k - u\|^2,$$

et $\|C(u^{k+1}-u)\|^2 \leq \|C\|^2\|u^{k+1}-u\|^2$. En combinant ces relations, on obtient l'inégalité (6) avec $\gamma = 2-2\beta-\rho_1^2\rho_2\|C\|^2$ et on choisit ρ_2 suffisamment petit pour que $\gamma > 0$ (ce qui est possible car $\beta < 1$). En sommant l'inégalité (6) pour k allant de 0 à $(l-1)$, et en constatant que le second membre télescope, il vient

$$\gamma \sum_{k=0}^{l-1} \|u^{k+1}-u\|^2 + \frac{1}{\rho_2}\|\lambda^l - \lambda\|^2 + \beta\|u^l - u\|^2 \leq \frac{1}{\rho_2}\|\lambda^0 - \lambda\|^2 + \beta\|u^0 - u\|^2,$$

pour tout $l \geq 1$. On conclut en faisant tendre $l \to \infty$ que la série au membre de gauche converge, si bien que $\|u^k - u\|$ tend vers zéro.

3. Vérifier que la matrice CC^T est inversible, puis montrer que

$$\lambda^k = (CC^T)^{-1}C\big(\rho_1^{-1}(u^k - u^{k+1}) + b - Au^k\big).$$

Que peut-on dire de la convergence de la suite λ^k ?

Correction: Si $v \in \mathrm{Ker}(CC^T)$, on a $0 = v \cdot CC^Tv = \|C^Tv\|^2$. On en déduit que $v = 0$ car C^T est injective (puisque C est de rang maximal). Donc, CC^T est bien inversible. D'après (5a), on a

$$C^T\lambda^k = \rho_1^{-1}(u^k - u^{k+1}) + b - Au^k.$$

En multipliant à gauche par C puis par $(CC^T)^{-1}$, on obtient l'expression demandée pour λ^k. Enfin, comme $u^k \to u$, le membre de droite de cette expressoin converge vers $(CC^T)^{-1}C(b - Au) = \lambda$ grâce à (4a). Ainsi, la suite λ^k converge vers λ.

III - Optimisation de portefeuille Supposons que nous pouvons acheter un actif financier au prix de X_0 euros à la date $T = 0$ puis le revendre pour X_1 euros à la date $T = 1$. On appelle la quantité

$$R = \frac{X_1 - X_0}{X_0}$$

le *taux de rendement* de l'actif.

Supposons maintenant que nous disposons de n actifs et souhaitons repartir un capital donné entre ces actifs. Notons par w_i la proportion de notre richesse investi dans le i-ème actif et par v le vecteur $v := (w_1, \ldots, w_n)^T$. Puisque toute la richesse initiale doit être investie,

$$\sum_{i=1}^{n} w_i = 1.$$

Le taux de rendement de notre portefeuille entre les dates $T = 0$ et $T = 1$ est alors donné par

$$R^v = \sum_{i=1}^{n} w_i R_i,$$

où R_i est le taux de rendement du i-ème actif.

Dans l'approche de H. Markowitz (1952), les rendements des actifs sont modélisés par des variables aléatoires, dont la moyenne et les covariances sont supposées connues par l'investisseur. Le risque d'un portefeuille est mesuré par sa variance, et le portefeuille optimal est celui qui minimise la variance en maintenant l'espérance du rendement à un niveau souhaité.

Soit $\mu_i := \mathbb{E}[R_i]$, $m := (\mu_1, \ldots, \mu_n)^T$, $b_{ij} := \mathrm{Cov}\,(R_i, R_j)$ et $B := (b_{ij})_{i,j=1}^n$. La variance du taux de rendement du portefeuille est alors $\mathrm{Var}\,R^v = v^T B v$ et le portefeuille optimal dans la théorie de Markowitz est la solution du problème d'optimisation suivant :

$$\min_{v \in \mathbb{R}^n} \frac{1}{2} v^T B v \tag{7}$$

$$\text{sous contraintes} \quad m^T v \geq \mu_b \quad \text{et} \quad \mathbf{1}^T v = 1, \tag{8}$$

où $\mathbf{1}$ est un vecteur de dimension n dont tous les éléments valent 1.

1. Donner une condition nécessaire et suffisante pour que l'ensemble défini par les contraintes (8) soit non vide pour tout μ_b. Dans la suite de l'exercice on supposera que cette condition est vérifiée.

Correction: Il faut que le couple $(m, \mathbf{1})$ soit linéairement indépendant.

2. Ecrire les conditions de Kuhn-Tucker pour le problème (7) et calculer la composition du portefeuille optimal en supposant que la matrice B est définie positive. Montrer que la stratégie optimale est de repartir sa richesse entre deux portefeuilles universels, à caractériser.

Correction: Les conditions de Kuhn-Tucker sont

$$\begin{aligned}
&Bv - \lambda m - \gamma \mathbf{1} = 0, \\
&m^T v \geq \mu_b, \quad \mathbf{1}^T v = 1, \quad \lambda \geq 0, \\
&\lambda(m^T v - \mu_b) = 0.
\end{aligned}$$

En résolvant la première équation sous contrainte $\mathbf{1}^T v = 1$, on trouve

$$v = \lambda B^{-1}m + (1 - \lambda \mathbf{1}^T B^{-1}m)\frac{B^{-1}\mathbf{1}}{\mathbf{1}^T B^{-1}\mathbf{1}} = \lambda B^{-1}m + (1 - \lambda \mathbf{1}^T B^{-1}m)v_{minvar},$$

où $v_{minvar} = \frac{B^{-1}\mathbf{1}}{\mathbf{1}^T B^{-1}\mathbf{1}}$ est le portefeuille qui résout le problème (7) sans la contrainte d'inégalité, c'est-à-dire, le portefeuille de variance minimale. Si

$$m^T v_{minvar} \geq \mu_b \quad \iff \quad \frac{m^T B^{-1}\mathbf{1}}{\mathbf{1}^T B^{-1}\mathbf{1}} \geq \mu_b,$$

la contrainte d'inégalité n'est pas active, et la solution optimale correspond à $\lambda = 0$, c'est-à-dire $v_{opt} = v_{minvar}$. Sinon, on calcule λ à partir de l'équation $m^T v = \mu_b$ ce qui donne

$$\lambda = \frac{\mu_b\, \mathbf{1}^T B^{-1}\mathbf{1} - m^T B^{-1}\mathbf{1}}{m^T B^{-1}m\, \mathbf{1}^T B^{-1}\mathbf{1} - (m^T B^{-1}\mathbf{1})^2}.$$

Si $m^T B^{-1}\mathbf{1} \neq 0$, on pose

$$\alpha = \lambda\, m^T B^{-1}\mathbf{1},$$

et le portefeuille optimal est représenté via

$$v_{opt} = \alpha \frac{B^{-1}m}{m^T B^{-1}\mathbf{1}} + (1 - \alpha)v_{minvar}.$$

Chaque agent va donc investir une fraction de sa richesse dans le portefeuille de variance minimale et une fraction dans un autre portefeuille qui dépend des rendements espérés des actifs mais est le même pour tous les agents du marché.

Nous supposons maintenant que l'agent a la possibilité d'investir dans l'actif sans risque. Le poids de l'actif sans risque dans le portefeuille sera noté w_0 et le taux de rendement de l'actif sans risque (déterministe) sera noté par r. Le problème d'optimisation de portefeuille devient alors

$$\min_{v \in \mathbb{R}^n, w_0 \in \mathbb{R}} \frac{1}{2} v^T B v \tag{9}$$

$$\text{sous contraintes} \quad m^T v + r w_0 \geq \mu_b \quad \text{et} \quad \mathbf{1}^T v + w_0 = 1, \tag{10}$$

3. Ecrire les conditions de Kuhn-Tucker pour le problème (9) et calculer la composition du portefeuille optimal en supposant que la matrice B est définie positive. Montrer que la stratégie optimale est de repartir sa richesse entre un portefeuille universel, à caractériser, et l'actif sans risque (ce résultat est connu sous le nom de "mutual fund theorem").

Correction: Les conditions de Kuhn-Tucker prennent la forme

$$\begin{aligned}
&Bv - \lambda m - \gamma \mathbf{1} = 0, \quad -\lambda r - \gamma = 0, \\
&m^T v + r w_0 \geq \mu_b, \quad \mathbf{1}^T v + w_0 = 1, \quad \lambda \geq 0, \\
&\lambda(m^T v + r w_0 - \mu_b) = 0.
\end{aligned}$$

En éliminant γ et w_0, on trouve

$$\begin{aligned}
&v = \lambda B^{-1}(m - r\mathbf{1}), \\
&v^T(m - r\mathbf{1}) \geq \mu_b - r, \quad \lambda \geq 0, \\
&\lambda(v^T(m - r\mathbf{1}) - \mu_b + r) = 0.
\end{aligned}$$

Si $\mu_b \leq r$, la contrainte d'inégalité n'est pas active et la solution optimale est $v_{opt} = 0$ et $w_0^{opt} = 1$, toute la richesse est investie en l'actif sans risque. Sinon, la valeur optimale de λ est donnée par

$$\lambda = \frac{\mu_b - r}{(m - r\mathbf{1})^T B^{-1}(m - r\mathbf{1})},$$

les proportions optimales des actifs risqués sont

$$v_{opt} = (\mu_b - r)\frac{B^{-1}(m - r\mathbf{1})}{(m - r\mathbf{1})^T B^{-1}(m - r\mathbf{1})}$$

et la proportion optimale de l'actif sans risque est

$$w_0^{opt} = 1 - (\mu_b - r)\frac{\mathbf{1}^T B^{-1}(m - r\mathbf{1})}{(m - r\mathbf{1})^T B^{-1}(m - r\mathbf{1})}.$$

Si $\mathbf{1}^T B^{-1}(m - r\mathbf{1}) \neq 0$, tous les agents repartissent leur richesse entre l'actif sans risque (qui joue le rôle du portefeuille de variance minimale) et le portefeuille

$$w_m = \frac{B^{-1}(m - r\mathbf{1})}{\mathbf{1}^T B^{-1}(m - r\mathbf{1})},$$

qui est le même pour tout le monde et s'appelle pour cette raison "le portefeuille de marché".