Optimisation sous contraintes

1 Gradient projeté

Soit A une matrice inversible de $M_N(\mathbb{R})$, et $b \in \mathbb{R}^N$. On cherche $x \in \mathbb{R}^N$ minimisant :

$$J(x) = \|Ax - b\|^2$$

sous les contraintes $x_i \geq 0$ $(1 \leq i \leq N)$.

1. Cacluler ∇J.
2. Montrer que J est elliptique, et que ∇J est lipshitzienne (préciser les constantes associées).
3. Expliciter l'algorithme du gradient projeté associé au problème de minimisation considéré ici. Pour quelles valeurs du pas τ est-on assuré de la convergence ?
4. Ecrire une fonction *x=minimise(A,b,τ,r)* qui retourne le résultat x de l'algorithme à l'itération r.
5. Utiliser la fonction précédente pour trouver la solution du problème quand :

$$A = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} \quad \text{et} \quad b = \begin{pmatrix} -3 \\ 4 \end{pmatrix}$$

6. Vérifier à la main la solution obtenue à la question précédente en appliquant les relations de Kuhn et Tucker.

2 Gradient projeté (bis)

On considère dans $\mathbb{R}^3$ l'ensemble :

$$C = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3,\ x_1^2 + x_2^2 \leq 1\right\}$$

1. (a) Quelle est la forme géométrique de C ? Montrer que C est convexe.
   (b) Expliciter la projection orthogonale P_C sur C.
   (c) Ecrire une fonction *z=projection(x)* qui en entrée prend un vecteur x de $\mathbb{R}^3$ et en sortie renvoie le vecteur z (de même taille que que x) projection de x sur C.
2. Soit A une matrice réelle 3×3 inversible, et $b \in \mathbb{R}^3$. On considère la fonction $J : \mathbb{R}^3 \to \mathbb{R}$ définie par $J(x) = \|Ax - b\|^2$, et le problème (P) : *minimiser $J(x)$ sous la contrainte $x \in C$*
   (a) Montrer que (P) admet une unique solution.
   (b) Expliciter l'algorithme du gradient projeté associé à (P). Pour quelles valeurs du pas est-on assuré de la convergence ?
   (c) Ecrire une fonction *x=minimise(A,b,τ,r)*, faisant appel à la fonction *projection*.
   (d) Calculer alors numériquement le minimum de J sur C quand :

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & 1 \\ 1 & 0 & 2 \end{pmatrix} \quad \text{et} \quad b = \begin{pmatrix} 3 \\ 1 \\ 3 \end{pmatrix}$$

   (e) Trouver, dans ce cas particulier, la valeur empirique à 10^{-4} près du pas critique τ_c au-delà duquel la convergence ne se produit plus.

3 Uzawa

On considère le problème (P) :

$$\min \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$$

sous la contrainte $Cx \leq d$, avec $x \in \mathbb{R}^n$, $d \in \mathbb{R}^p$, C une matrice $p \times n$ à coefficients réels et A une matrice symétrique $n \times n$ réelle définie positive.

Plus précisément, on prend $n = 6$, $p = 3$, et :

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 & \dots & \dots \\ -1 & 2 & -1 & 0 & \dots & \dots \\ 0 & -1 & 2 & -1 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & 0 & -1 & 2 & -1 \\ \dots & \dots & \dots & 0 & -1 & 2 \end{pmatrix} \quad \text{et} \quad b = \begin{pmatrix} 1 \\ 1 \\ \dots \\ \dots \\ 1 \\ 1 \end{pmatrix}$$

$$C = -\begin{pmatrix} 3 & 1 & 0 & -1 & 0 & 0 \\ -1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix} \quad \text{et} \quad d = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

1. Vérifier que A est définie positive. A quel opérateur classique pourrait correspondre A ?
2. Ecrire le Lagrangien $\mathcal{L}$ associé à ce problème.
3. Résoudre (P) avec la méthode d'Uzawa (en choisissant judicieusement le pas).
4. Vérifier la solution obtenue avec les relations de Kuhn et Tucker (modifiez votre programme afin d'afficher les informations suffisantes pour vérifier ces relations). Remplacer b par $-b$, que constatez-vous sur les multiplicateurs de Lagrange ?

4 Rappels sur l'algorithme d'Uzawa

4.1 Ideas from duality

Let V and M two sets, and a L a function :

$$L : V \times M \to \mathbb{R} \tag{1}$$

Definition :

A point (u, λ) is said to be a saddle point (*point selle*) of L if u is a minimizer of $L(., \lambda) : v \in V \to L(v, \lambda) \in \mathbb{R}$ and if λ is a maximizer of $L(u, .) : \mu \in M \to L(u, \mu) \in \mathbb{R}$, i.e. if :

$$\sup_{\mu \in M} L(u, \mu) = L(u, \lambda) = \inf_{v \in V} L(v, \lambda) \tag{2}$$

We consider $J : V \to \mathbb{R}$ and $\phi_i : V \to \mathbb{R}$, $1 \leq i \leq m$, and

$$U = \{v \in V = \mathbb{R}^N, \phi_i(v) \leq 0, 1 \leq i \leq m\} \tag{3}$$

We consider the problem (primal problem) (P) :

$$\text{Find } u \in U \text{ such that } J(u) = \inf_{v \in U} J(v) \tag{4}$$

Under some specific conditions, any solution u of (P) is the first argument of a saddle point (u, λ) of a certain function L called *Lagrangian* associated to problem (P). The second argument λ is called *generalized Lagrange multiplier* associated to u (since as we will see it is the vector given by the Kuhn and Tucker relations).

Let us define the *Lagrangien* associated to problem (P) as :

$$L : (v, \mu) \in V \times \mathbb{R}^m_+ \to L(v, \mu) = J(v) + \sum_{i=1}^{m} \mu_i \phi_i(v) \tag{5}$$

Under some hypotheses, a solution u of the primal problem (P) is the first argument of a saddle point of its associated Lagrangien.

Assume that we know one of the second argument λ of the saddle points of L. Then the constrained problem (P) would be replaced by an unconstrained problem (P_λ) : find u_λ such that :

$$u_\lambda \in V \text{ and } L(u_\lambda, \lambda) = \inf_{v \in V} L(v, \lambda) \tag{6}$$

Now the question is how to find such a $\lambda \in \mathbb{R}^m_+$. Recall the equality verified by saddle points :

$$L(u_\lambda, \lambda) = \inf_{v \in V} L(v, \lambda) = \sup_{\mu \in \mathbb{R}^m_+} \inf_{v \in V} L(v, \mu) \tag{7}$$

We are therefore led to find λ as a solution of the dual problem (Q) :

$$\text{Find } \lambda \in \mathbb{R}^m_+ \text{ such that } G(\lambda) = \sup_{\mu \in \mathbb{R}^m_+} G(\mu) \tag{8}$$

where $G : \mathbb{R}^m_+ \to \mathbb{R}$ is defined by :

$$G : \mu \in \mathbb{R}^m_+ \to G(\mu) = \inf_{v \in V} L(v, \mu) \tag{9}$$

$\mu \in \mathbb{R}^m_+$ is called the dual variable of the primal variable $v \in V = \mathbb{R}^N$.

Notice that the dual problem (Q) is also a constrained problem, but the constraints $\mu_i \geq 0$ are very easy to handle (since we know explicitely the projection operator). On the contrary, the constraints $\phi_i(u) \leq 0$ are in general impossible to handle numerically. This is the basic idea of Uzawa algorithm.

4.2 Uzawa algorithm

The idea of the method is that the projection operator $P_+ : \mathbb{R}^m \to \mathbb{R}^m_+$ in the dual problem (Q) is very simple :

$$(P_+\lambda)_i = \max(\lambda_i, 0) \tag{10}$$

Uzawa algorithm is in fact the projected gradient method applied to the dual problem (Q) : $\lambda_0 \in \mathbb{R}^m_+$ arbitrary, and the sequence λ_k in $\mathbb{R}^m_+$ is defined by :

$$\lambda_{k+1} = P_+(\lambda_k - \rho \nabla G(\lambda_k)) \tag{11}$$

Since in the dual problem (Q), one is interested in a maximum (and not a minimum), it is therefore natural to change the sign of the parameter ρ with respect to the classical method.

Under some hypotheses, it is possible to compute the gradient of G :

$$(\nabla G(\mu))_i = \phi_i(u_\mu) \ , \ 1 \leq i \leq m \tag{12}$$

the vector u_μ being the solution of the *unconstrained* minimization problem :

$$u_\mu \in V, \ J(u_\mu) + \sum_{i=1}^{m} \mu_i \phi_i(u_\mu) = \inf_{v \in V} \left\{ J(v) + \sum_{i=1}^{m} \mu_i \phi_i(v) \right\} \tag{13}$$

Uzawa algorithm : $\lambda_0 \in \mathbb{R}_+^m$ arbitrary. We define by induction $(\lambda^k, u^k) \in \mathbb{R}_+^m \times V$ by (for the sake of clarity, we write $u^k = u_{\lambda_k}$) :

$$\left\{ \begin{array}{ll} \text{Computation of } u^k : & J(u^k) + \sum_{i=1}^m \lambda_i^k \phi_i(u^k) = \inf_{v \in V} \left\{ J(v) + \sum_{i=1}^m \lambda_i^k \phi_i(v) \right\} \\ \text{Computation of } \lambda_i^{k+1} : & \lambda_i^{k+1} = \max\left\{ \lambda_i^k + \rho \phi_i(u^k), 0 \right\} \end{array} \right. \tag{14}$$

Uzawa method is a way to replace a constrained minimization problem by a sequence of unconstrained minimization problem.

Notice that u^k can converge while λ^k does not.

Convergence of Uzawa method

$V = \mathbb{R}^N$. J elliptic (with contant α), and :

$$U = \{v \in \mathbb{R}^N, Cv \leq d\},\ C \in \mathcal{A}_{m,N}(\mathbb{R}),\ d \in \mathbb{R}^m \tag{15}$$

is non empty. Then, if

$$0 < \rho < \frac{2\alpha}{\|C\|^2} \tag{16}$$

the sequence u_k converges to the unique solution of the primal problem (P).

If the rank of C is m, then the sequence λ^k also converges towards the unique solution of the dual problem (Q).

Notice that $\|C\| = \sup_{v \in\ \mathbb{R}^N} \frac{\|Cv\|_m}{\|v\|_n}$.

Case of a quadratic functional :

$$J(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle \tag{17}$$

with A symmetric definite positive matrix of size N^2, and $b \in \mathbb{R}^N$.

$$U = \left\{ v \in \mathbb{R}^N, \sum_{j=1}^N c_{i,j} v_j \leq d_i, 1 \leq i \leq m \right\} = \left\{ v \in \mathbb{R}^N, Cv \leq d \right\} \tag{18}$$

where C is a real $m \times N$ matrix, and $d \in \mathbb{R}^m$. We assume U non empty (and thus the constraints are qualified).

An iteration of Uzawa algorithm is :

$$\left\{ \begin{array}{ll} \text{Computation of } u^k : & Au^k - b + C^T \lambda^k = 0 \\ \text{Computation of } \lambda_i^{k+1} : & \lambda_i^{k+1} = \max\{(\lambda^k + \rho(Cu^k - d))_i, 0\} \end{array} \right. \tag{19}$$

And the method converges if

$$0 < \rho < \frac{2\lambda_1(A)}{\|C\|^2} \tag{20}$$

where $\lambda_1(A)$ is the smallest eigenvalue of A (it is the ellipticity constant).