RESUME DE RECHERCHE OPERATIONNELLE

RO – Résumé

I. Dérivées

1. Ligne de niveau

$$N_c = \{x \in \mathbb{R}^n | J(x) = c\}$$

2. Dérivée première

a. Gradient

$$\nabla J(x_0) = \left[\frac{\partial J}{\partial x_1}, \dots, \frac{\partial J}{\partial x_n}\right]^\top$$

Propriété : Au point x_0, $\nabla J(x_0)$ est $\perp$ à la ligne de niveau, son sens va dans le sens de J croissant.

b. Dérivée directionnelle

$$\boxed{D_x J(x,d) \quad = \quad \lim_{\varepsilon \to 0} \frac{J(x+\varepsilon d) - J(x)}{\varepsilon} \quad = \quad \frac{d \overbrace{J(x+\varepsilon d)}^{\varphi(\varepsilon)}}{d\varepsilon}\Bigg|_{\varepsilon=0} \quad = \quad \varphi'(0) \quad = \quad \nabla J(x)^\top d}$$

c. Plan tangent

Passe par $(x_0, J(x_0))$ et vecteur directeur $\nabla J(x_0)$

$$\underline{P = \{x \in \mathbb{R}^n | \nabla J(x_0)^\top (x - x_0) = 0\}}$$

Approximation locale de J

$$J(x) \approx P(x) = J(x_0) + \nabla J(x_0)^\top x$$

d. Développement limité au premier ordre

$$J(x) \approx J(x+\varepsilon d) = J(x) + \varepsilon \underbrace{\nabla J(x)^\top d + o(\varepsilon)}_{DJ_x(x,d)}$$

e. Règles de calcul pour la dérivée

$$D(J_1 + \alpha J_2) = DJ_1 + \alpha J_2 \;\Big|\; D\big(J_1(J_2)\big) = D_z J_1(z = J_2)\, D_x J_2(x) \;\Big|\; \nabla a^\top x = a \;\Big|\; \nabla x^\top A x = 2Ax$$

3. Dérivées secondes

a. Dérivée directionnelle au sens de Gâteaux

$$D^2 J(x,d) = \lim_{\varepsilon \to 0} \frac{DJ_x(x+ed) - DJ_x(x)}{\varepsilon}$$

b. Matrice Hessienne

$$H_J(x) = \begin{pmatrix} \frac{\partial^2 J(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 J(x)}{\partial x_1 \partial x_j} & \cdots & \frac{\partial^2 J(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J(x)}{\partial x_i \partial x_1} & \cdots & \frac{\partial^2 J(x)}{\partial x_i \partial x_j} & \cdots & \frac{\partial^2 J(x)}{\partial x_i \partial x_n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 J(x)}{\partial x_n \partial x_j} & \cdots & \frac{\partial^2 J(x)}{\partial x_n \partial x_n} \end{pmatrix}$$

Calcul pratique :

A partir de la dérivée de

$$\varphi(\varepsilon) = D_x J(x+\varepsilon d)$$

Identifier H dans $\varphi'(0)$:

$$\varphi'(0) = d^\top H_X(x)^\top d$$

c. Développement limité au second ordre

$$J(x+d) = J(x) + D_x J(x,d) + \frac{1}{2} D_x^2 J(x,d) + o(\|d\|^2)$$

$$J(x+d) = J(x) + \nabla_x J(x)^\top d + \frac{1}{2} d^\top H d \qquad + o(\|d\|^2)$$

II. Généralités

Variables $\boldsymbol{V}$	$V \in \mathbb{R}^p$
Objectifs $\boldsymbol{O}$	$V \mapsto \mathbb{R}^q$ $\quad \min\limits_{x\in\Omega} J(x)$
Contraintes $\boldsymbol{C}$	$V \mapsto \mathbb{R}^{m+n}$ $\quad \begin{cases} h_i(x) = 0 \\ g_j(x) \leq 0 \end{cases} \Leftrightarrow \begin{cases} H(x) = (h_1, \dots, h_n)^\top = 0 \\ G(x) = (g_1, \dots, g_m)^\top \leq 0 \end{cases}$
Domaine de faisabilité $\boldsymbol{\Omega}$	$\Omega = \{x \in \mathbb{R}^d \,; H(x) = 0 \text{ et } G(x) \leq 0\}$
Fonction coût	$J : \Omega \mapsto \mathbb{R}$
Domaine de la fonction coût	$\operatorname{dom} J = \{x \in \Omega \,; -\infty < J(\theta) < +\infty\}$ *Si $\emptyset$, fonction coût impropre : pas de solution*
Minimum global $\boldsymbol{\theta^*}$	$J(\theta^*) \leq J(\theta) \quad \forall\, \theta$
Minimum local $\boldsymbol{\hat{\theta}}$	$J(\hat{\theta}) \leq J(\theta) \quad \forall\, \theta \mid \lVert\hat{\theta} - \theta\rVert \leq \varepsilon$

III. Optimisation convexe sous contraintes

1. Avec contraintes d'égalités

Problème	Lagrangien	Conditions d'optimalité
$\begin{cases} \min\limits_{x\in\mathbb{R}^n} J(x) \\ s.c.\, H(x) = 0 \end{cases}$	$\boxed{L(x,\lambda) = J(x) + \sum\limits_{j=1}^{p} \lambda_j H_j(x)}$ *λ_j multiplicateurs de Lagrange*	$\nabla_x L = 0 \quad \Rightarrow \quad \nabla_x J(x) + \sum\limits_{j=1}^{p} \lambda_i \nabla_x H_j(x) = 0$ $\nabla_{\lambda_j} L = 0 \quad \Rightarrow \quad H_j(x) = 0$

2. Avec contraintes d'inégalités

a. Lagrangien et conditions d'optimalité

Problème	Lagrangien	Conditions d'optimalité KKT
$\begin{cases} \min\limits_{x\in\mathbb{R}^n} \; J(x) \\ \text{s.c.} \quad H(x) = 0 \\ \text{et} \quad G(x) \leq 0 \end{cases}$	$\boxed{L(x,\lambda,\mu) = J(x) + \sum\limits_{j=1}^{p} \lambda_j H_j(x) + \sum\limits_{i=1}^{q} \mu_i G_i(x)}$ *λ_j multiplicateurs de Lagrange*	Stationarité : $\nabla L(x,\lambda) = 0$ Adm. primale : $G(x) \leq 0$, $H(x) = 0$ Adm. duale : $\mu_i \geq 0$ Complémentarité : $\mu_i g_i(x) = 0$

b. Problème dual

Le problème dual est $\boxed{\mathcal{L}(\mu,\lambda) = \min\limits_{x} L(z,\lambda,\mu)}$

IV. Optimisation convexe sans contraintes

1. Matrice définie positive

A définie positive	A définie négative	A quadratique
$\lambda_i \leq 0$ J convexe	$\lambda_i \geq 0$	$\begin{cases}\lambda_i \leq 0\\ \lambda_j \geq 0\end{cases}$

2. Condition d'optimalité

a. Existence de solution

J est coercivité	$\lVert x \rVert \to \infty \Rightarrow J(x) \to \infty$ (fonction infinie à l'infini)
J est propre	$\exists\, x \mid J(x) \in \mathbb{R}$
∃ solution globale	Si J continue, propre, coercive, $\min_x J(x)$ admet une solution globale

b. Conditions d'optimalité

1er ordre	x_0 solution $\Rightarrow \nabla J(x_0) = 0$
2ème ordre	x_0 solution $\Rightarrow \nabla J(x_0) = 0$ et $H(x_0)$ définie positive $\nabla J(x_0) = 0$ et $H(x_0)$ définie positive $\Rightarrow$ x_0 solution locale
Convexité	J est convexe et x_0 respecte condition d'optimalité $\Rightarrow x_0 = x^*$ solution globale

3. Optimisation itérative

a. Recherche linéaire *(line search)*

$$x_{k+1} = x_k + \rho_k d_k \qquad \lim_{k\to\infty} x_k = x^* = \operatorname*{argmin}_{x\in\mathbb{R}^n} J(x)$$

i. Direction de descente d $\quad (\nabla J^\top d < 0)$

- **Gradient :** $d_k = -\nabla J$ $\quad \mathcal{O}(n)$
- **Gradient conjugué :** $d_k = -\nabla J + \beta_k d_{k-1}$ $\quad \mathcal{O}(n^2)$
- **Quasi-Newton :** $d_k = -B\nabla J$ $\quad$ (Ex : $B = diag(H)^{-1}$) $\quad \mathcal{O}(n^2)$
- **Newton :** $d_k = -H^{-1}\nabla J$ $\quad \mathcal{O}(n^3)$

ii. Choix du pas ρ

- **Pas fixe :** $\rho_k = c$ $\quad \mathcal{O}(1)$
- **Pas variable :** $\rho_k = \begin{cases}\alpha\rho_{k-1} \text{ si } J \text{ diminue} & \alpha = 1{,}15\\ \alpha\rho_{k-1} \text{ si } J \text{ augmente} & \beta = 0{,}5\end{cases}$ $\quad \mathcal{O}(1)$
- **Pas optimal :** $\rho_k = \begin{cases}\operatorname*{argmin}_{\rho\in\mathbb{R}^+} J(x+\rho d) & < \mathcal{O}(n^2)\\ \dfrac{d^\top d}{d^\top H d} & \mathcal{O}(n^2)\end{cases}$

b. Région de confiance *(trust region)*

Comme recherche linéaire avec saut maximum limité par une région de confiance

$$x_{k+1} = x_k + d_k \quad \text{avec} \quad d_k = \operatorname*{argmin}_{d\in \boldsymbol{R}(x_k)} M(x_k + d)$$

$R(x)$: région de confiance autour de x $\qquad$ $M(x+d)$: modèle dans la région de confiance (DL)

V. Programmation linéaire

1. Forme standard

$$\begin{cases} \min\limits_{z\in\mathbb{R}^m} c^\top z \\ Az = b \\ z \geq 0 \end{cases}$$

z : m inconnues
c : m couts
A : p contraintes ($p < m$)
b : p seconds membres des contraintes

2. Méthodes de réduction à la forme standard

a. Inégalités ⇒ variables d'écart (positives)

$$z = \begin{bmatrix} x \\ e \end{bmatrix} \qquad \begin{array}{l} f_i(x) \geq b_i \Leftrightarrow f_i(x) - e_i = b_i \\ f_i(x) \leq b_i \Leftrightarrow f_i(x) + e_i = b_i \end{array}$$

b. Variables non contraintes ⇒ décomposition

x_i sans contraintes remplacé par $\begin{cases} x_i = x_i^+ - x_i^- \\ x_i^+, x_i^- \geq 0 \end{cases}$

3. Forme duale standard

$$\begin{cases} \max\limits_{y\in\mathbb{R}^p} b^\top y \\ A^\top y \leq c \end{cases} \qquad c^\top x^* = b^\top y^* \qquad y = \lambda$$

4. Méthode du point intérieur

a. Principe

Résolution d'un problème à la forme standard en résolvant simultanément le primal et le dual à travers les KKT, en les écrivant sous la forme $F(z, \lambda, \mu) = 0$ avec $z, \mu \geq 0$.

C'est une méthode itérative de Newton projeté appliquée à $F\begin{pmatrix} \Delta z \\ \Delta\lambda \\ \Delta\mu \end{pmatrix}$.

b. Maths et algorithme

$$\begin{cases} \min\limits_{z\in\mathbb{R}^m} c^\top z \\ Az = b \\ z \geq 0 \end{cases} \quad \begin{array}{l} \left.\begin{array}{l} A^\top\lambda + \mu - c = 0 \\ Az = b \\ diag(\mu) z = 0 \end{array}\right\} \Leftrightarrow F(z, \lambda, \mu) = 0 \\ z \geq 0 \\ \mu \geq 0 \end{array}$$

$$\boxed{\begin{pmatrix} z \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} z \\ \lambda \\ \mu \end{pmatrix} = pas \times \underbrace{\begin{pmatrix} \Delta z \\ \Delta\lambda \\ \Delta\mu \end{pmatrix}}_{d}}$$

Avec un DL au 1er ordre :

$$\underbrace{F(x + d)}_{=0} = F(x) + d^\top \nabla_x F(x) + \underbrace{o(\|d\|)}_{=0}$$

$$\Rightarrow \boxed{d = \nabla_x F(x)^{-1} F(x)}$$

$$\boxed{\nabla_x F(x) = \begin{pmatrix} 0 & A^\top & I \\ A & 0 & 0 \\ diag(\mu) & 0 & diag(z) \end{pmatrix}}$$

Forme matricielle de F :

$$F(z) = A_F \times x_F + b_F$$

$$\begin{array}{c|ccc|} & m & p & m \\ m & 0 & A^\top & I \\ p & A & 0 & 0 \\ m & diag(\mu) & 0 & 0 \end{array} \times \begin{bmatrix} z \\ \lambda \\ \mu \end{bmatrix} + \begin{bmatrix} -c \\ -b \\ \delta\sigma \end{bmatrix}$$

Pour que ça marche, on met $\delta\sigma$ dans b_F avec $\delta = \mu^\top z$ et $\sigma \ll 1$

On calcule le pas tel que $z \geq 0, \mu \geq 0$.