RESUME D'OPTIMISATION

Optim – 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

$$\boxed{\nabla(J_1 + \alpha J_2) = \nabla J_1 + \alpha \nabla J_2 \;\Big|\; \nabla\big(J_1(J_2)\big) = \nabla J_1|_{x=J_2} \nabla J_2 \;\Big|\; \nabla a^\top x = a \;\Big|\; \nabla x^\top A x = (A + A^\top)x}$$

3. Dérivées secondes

a. Dérivée directionnelle au sens de Gâteaux

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

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. Moindres carrés

$$\hat{y} = X\theta + \varepsilon$$

Moindres carrés

$$\min_\theta \|\varepsilon\|_2^2 = \min_\theta (X\theta - y)^\top (X\theta - y)$$
$$= \min_\theta \theta^\top X^\top X\theta - 2y^\top X\theta + y^\top y$$
$$\boxed{\hat{\theta}_{MC} = (X^\top X)^{-1} X^\top y}$$

Moindres carrés pondérés

$$\min_\theta \|W\varepsilon\|_2^2 = \min_\theta (X\theta - y)^\top W (X\theta - y)$$
$$= \min_\theta \theta^\top X^\top W X\theta - 2y^\top W X\theta + y^\top W y$$
$$\boxed{\hat{\theta}_{MC} = (X^\top W X)^{-1} X^\top W y}$$

Moindres carrés récursifs

$$\hat{\theta} = \underbrace{(X_n^\top X_n)^{-1}}_{P_n} \underbrace{X_n Y_n}_{Q_n}$$

$$P_n = \left(\underbrace{P_{n-1}^{-1}}_{A} + \underbrace{x_n}_{B} \underbrace{I}_{C} \underbrace{x_n^\top}_{D}\right)^{-1} = P_{n-1} - k_n x_n^\top P_{n-1}$$

$$k_n = P_{n-1} x_n (1 + x_n^\top P_{n-1} x_n)^{-1}$$

$$Q_n = Q_{n-1} + x_n y_n$$

$$\boxed{\hat{\theta}_n = \hat{\theta}_{n-1} + k_n\left(Y_n - x_n^\top \hat{\theta}_{n-1}\right)}$$

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

III. Descente de gradient

1. Condition 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

2. Optimisation itérative

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

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

- **Gradient :** $d_k = -\nabla J$ .... $\mathcal{O}(n)$
- **Gradient conjugué :** $d_k = -\nabla J + \beta_k d_{k-1} \quad \beta_k = {\nabla J\, A d_{k-1}}/{d_{k-1} A d_{k-1}}$ .... $\mathcal{O}(n^2)$
- **Grad. conj. $\overline{\text{quad}}$:** $d_k = -\nabla J + \beta_k d_{k-1} \quad \beta_k = {\|\nabla J\|^2}/{\|\nabla J_{k-1}\|^2}$ .... $\mathcal{O}(n^2)$
- **Quasi-Newton :** $d_k = -B\,\nabla J$ .... (Ex : $B = \operatorname{diag}(H)^{-1}$) .... $\mathcal{O}(n^2)$
- **Newton :** $d_k = -H^{-1}\nabla J$ .... $\mathcal{O}(n^3)$

b. Choix du pas ρ

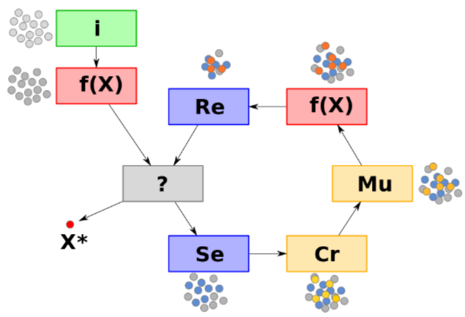
- **Pas fixe :** $\rho_k = c$ .... $\mathcal{O}(1)$
- **Pas adaptatif :** $\rho_k = \begin{cases} \alpha\rho_{k-1} \text{ si } J \text{ diminue} & \alpha = 1{,}15 \\ \alpha\rho_{k-1} \text{ si } J \text{ augmente} & \beta = 0{,}5 \end{cases}$ .... $\mathcal{O}(1)$
- **Pas grad. conjugué :** $\rho_k = {-\nabla J\, d_k}/{d_k^\top A d_k}$ .... $\mathcal{O}(1)$
- **Pas quasi-optimal :** $\rho_k = {-b}/{2a} \mid a\rho_k^2 + b\rho_k + c \approx J(x_k + \rho_k d_k)$ .... $\mathcal{O}(1)$
- **Pas optimal :** $\rho_k = \begin{cases} \underset{\rho\in\mathbb{R}^+}{\operatorname{argmin}} J(x + \rho d) \\ \dfrac{d^\top d}{d^\top H d} \end{cases}$ .... $\begin{cases} < \mathcal{O}(n^2) \\ \mathcal{O}(n^2) \end{cases}$

IV. Méthodes itératives globales

1. Recuit simulé

Selon une température T qui décroit, on tolère plus ou moins de conserver des solutions θ_{k+1} qui dégradent le critère $\left(J(\theta_{k+1}) > J(\theta_k)\right)$.

Init : $\theta_0, T_0 > 0$
while (on a pas convergé)
  $\theta'_k = f(\theta_k)$ *// nouvelle solution hypothétique (potentiellement aléatoirement)*
  $\delta = J(\theta'_k) - J(\theta_k)$
  if $\delta \leq 0$ **or** $U_{[0,1]} < e^{-\delta/T_k}$
    $\theta_{k+1} = \theta'_k$
  else
    $\theta_{k+1} = \theta_k$
  end
  $T_{k+1} = g(T_k)$
end while

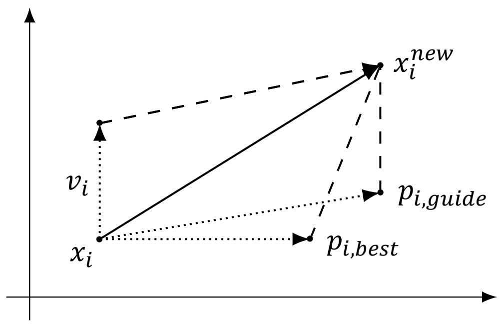


2. Méthodes évolutionnaires

- **Sélection (Se) :**
  - **Roue de loterie (choix probabiliste) :**
    - $\mathbb{P}(x_i) = J(x_i), \quad aJ(x_i) + b, \quad J(x_k)^k$
    - $\mathbb{P}(x_i) = nb + 1 - \text{rang}\, x_i$
    - $\mathbb{P}(x_i) = \frac{J(x_i)}{m}$ (m pénalise les solutions avec voisins proches)
  - **Tournois :**
    - Déterministes : $n \in [1, T]$ meilleurs parmi T
    - Stochastique : Meilleur parmi n avec une proba t
    - Evolutionnaire : $\forall\, x_i$, tournoi avec $T - 1$ concurrents
- **Croisement (Cr) :** $x_i^{new} = (1-\alpha)x_i + \alpha x_j \quad \alpha = U_{[0,1]}, \alpha' = (1+2\alpha)U_{[0,1]} - \alpha$
- **Mutation (Mu) :** modification aléatoire des enfants
- **Réduction (Re) :**
  - **Générationnel :** p enfants remplacent p parents
  - **Steady-state :** chaque enfant remplace 1 de ses parents

3. Essaims particulaires

$$v_i^{new} = \omega r_0 v_i + c_1 r_1 \left(p_{i,best} - x_i\right) + c_2 r_2 \left(p_{i,guide} - x_i\right)$$
$$x_i^{new} = x_i + \chi(v_i^{new})$$



- x_i position de la particule i
- v_i vitesse de la particule i
- $p_{i,best}$ meilleure pos. vue par i
- $p_{i,guide}$ position d'un guide
- r_k valeurs aléatoires
- c_1 facteur individuel
- c_2 facteur social
- ω facteur d'inertie
- $\chi(\cdot)$ modification de la vitesse (contraintes min/max, aléatoire, …)

V. Optimisation multicritère

Une solution x domine une solution $y \Leftrightarrow x \prec y \Leftrightarrow \forall\, k, f_k(x) \leq f(y), \exists k, f_k(x) < f(y)$

1. Approches

- Méthodes monocritères (pondérées, lexicographiques, ϵ-contraintes)
- Méthodes « Pareto » à base population (génétique, essaims, colonies de fourmis)

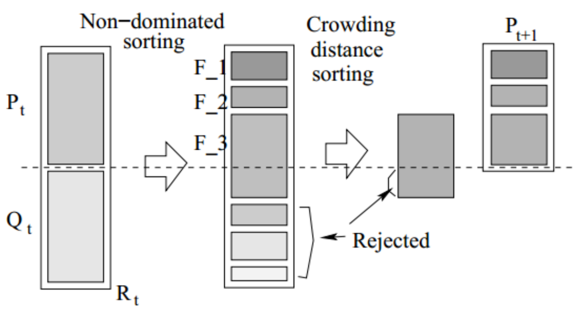
2. Méthodes monocritères

Méthode pondérée	**Méthodes lexicographiques**	**Méthode ϵ contrainte**
Critère combiné $J = \sum w_i J_i$. Variation des w_i.	Objectifs triés par ordre d'importance, optimisations successives sous contraintes.	Minimiser l'objectif principal. Autre objectifs exprimés comme contraintes $J_k \leq \epsilon_k$. Variation de ϵ_k.

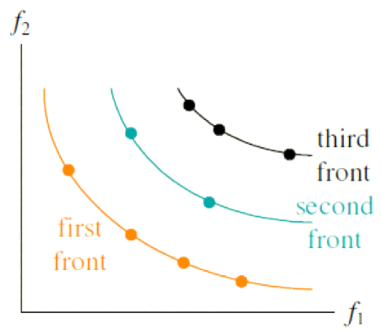
3. Méthodes « Pareto »

- Utilisation de la notion de rang de dominance
- Utilisation d'une archive conservant les solutions non dominées
- Gestion de la diversité

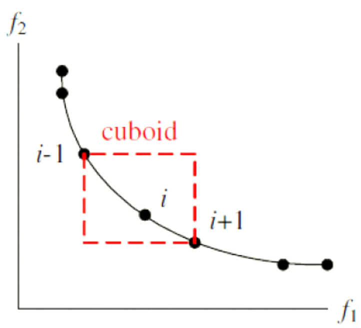
VEGA	**SPEA**	**NSGA-II**	**Essaims**
Algorithme génétique où parmi p parents, on fait une pré-sélection selon chaque critère. *Individus bons dans 1 domaine particulier ⇒ non-Pareto.*	Suppression des individus semblables dans l'archive. Force S_i, Fitness F_i $S_i = \frac{n_{dominés}}{n+1}$ $F_i = 1 + \sum_{j \in dom^{ant}} S_j$	Fitness basé sur rang de non dominance. Distance de crowding pour préserver div. (taille max du cube contenant uniquement x_i)	Elimin. de l'archive si autre sol. proches (inclus dans la dom.). Sélection probabiliste du guide $\frac{1}{densité}$.



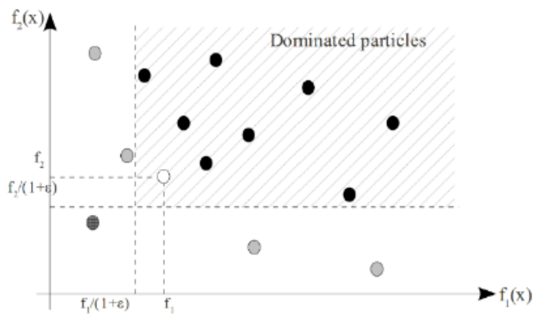
NSGA-II



Rangs de non-dominance



Distance de crowding = Cuboïde de x_i



Diversité ∈ Dominance