Fiche

Charles Vin

Date

1 Formule et définition

- Produit scalaire : $\langle x, y\rangle = x^T y = \sum x_i y_i$
- Norme : $\|x\| = \sqrt{\langle x, x\rangle}$
- Identité remarquable : $\|a+b\| = \|a\| + \|b\| + 2\langle a, b\rangle$
- Inégalité de Cauchy : $|\langle x, y\rangle| \leq \|x\| \|y\|$
- k-lipschitzienne : $|f(x) - f(y)| \leq k|x - y|$ Bouger dans l'espace d'arriver fait bouger k fois plus dans l'espace de départ.
- L-Smooth : = gradient Lipschitz$\forall \theta, \theta', \|\nabla F(\theta) - \nabla F(\theta')\| \leq L\|\theta - \theta'\|$
- Bilinéarité du produit scalaire :
  - $k\langle x, y\rangle = \langle kx, y\rangle = \langle x, ky\rangle$
  - $\langle z, x + y\rangle = \langle z, x\rangle + \langle x, y\rangle$
- Inégalité triangulaire : $\|x + y\| \leq \|x\| + \|y\|$
- GD : $\theta_{t+1} = \theta_t - \gamma \nabla F(\theta_t)$
- Polyak-Ruppert averaging : $\bar{\theta}_T = \frac{1}{T}\sum_{t=1}^{T} \theta_t$
- Sub gradient : $f(x) - f(x_0) \geq \langle v, (x - x_0)\rangle$
- KKT :
- Dérivé norme : $\frac{d}{dt}\|f(t)\| = \frac{\langle f(t), f'(t)\rangle}{\|f(t)\|}$
- Dérivé norme au carré : $\frac{d}{dt}\|f(t)\|^2 = 2\langle f(t), f'(t)\rangle$
- Convexity props
  - Convexity : under chords : $F(\eta\theta + (1-\eta)\theta') \leq \eta F(\theta) + (1-\eta)F(\theta'), \forall \theta, \theta', \forall \eta \in [0, 1]$
  - Convexity + diff (tangent lie below) $F(\theta') \geq F(\theta) + \langle \nabla F(\theta), \theta' - \theta\rangle, \forall \theta, \theta'$
  - Convexity : increasing slopes $\langle \nabla F(\theta) - \nabla F(\theta'), \theta - \theta'\rangle \geq 0$ (∇F is said to be a monotone operator)
  - Convexity + $\mathcal{C}^2$: curves upwards $\forall \theta, Hess_F(\theta) \succeq 0$ (SDP)
- μ-strongly convex, $\mu > 0$.
  - μ-convexity : **well** under chords : $F(\eta\theta + (1-\eta)\theta') \leq \eta F(\theta) + (1-\eta)F(\theta') - \frac{\eta(1-\eta)\mu}{2}\|\theta - \theta'\|_2^2, \forall \theta, \theta', \forall \eta \in [0, 1]$
  - μ-convexity + diff (tangent lie **well** below) : $F(\theta') \geq F(\theta) + < \nabla F(\theta), \theta' - \theta > + \frac{\mu}{2}\|\theta - \theta'\|_2^2, \forall \theta, \theta'$
  - μ-convexity : **well** increasing slopes : $\langle \nabla F(\theta) - \nabla F(\theta'), \theta - \theta'\rangle \geq 0 + \mu\|\theta - \theta'\|_2^2$
  - μ-convexity + $\mathcal{C}^2$: curves upwards : $\forall \theta, Hess_F(\theta) \succeq \mu Id$ (SDP)
- Co-coercivity = L-Smooth + Convexe : $\frac{1}{L}\|\nabla F(\theta) - \nabla F(\theta')\|_2^2 \leq \langle \nabla F(\theta) - \nabla F(\theta'), \theta - \theta'\rangle$

2 Technique de preuve

- Penser au $\pm$ pour faire apparaitre un terme voulu
- $\nabla F(\theta^\infty) \approx \nabla F(\theta^\star) = 0$
- Trick de l'intégrale

$$\begin{aligned} F(x - \gamma y) - F(x) &= F(x - \gamma y) - F(\theta - 0 \times y) \\ &= [F(x - \tau y)]_0^\gamma \\ &= \int_0^\gamma \langle\rangle \end{aligned}$$

- Si on a des inégalités avec du θ_1 et des sommes, potentiel somme d'inégalités

— On utilise souvent la cocoercivity du gradient avec $\nabla F(\theta^\star) = 0$

$$\|\nabla F(\theta_t)\| \leq L \langle \nabla F(\theta_t), \theta_t - \theta^\star \rangle .$$

3 Théorèmes importants

Lemme 3.1 (Descent lemma)**.** *Assume that F is L-Smooth. Therefore $\forall \theta, \theta' \in$ domain of F*

$$F(\theta') \leq F(\theta) + \langle \nabla F(\theta), \theta' - \theta \rangle + \frac{L}{2} \|\theta' - \theta\|^2 .$$

HEAVYBALL [Polyak, 64]

$$\begin{aligned} \beta_k &= \theta_k + (1 - \alpha_k)(\theta_k - \theta_{k-1}) \\ \theta_{k+1} &= \beta_k - \gamma \nabla F(\theta_k) \end{aligned}$$

NESTEROV ALGO [83]

$$\begin{aligned} \beta_k &= \theta_k + (1 - \alpha_k)(\theta_k - \theta_{k-1}) \\ \theta_{k+1} &= \beta_k - \gamma \nabla F(\beta_k) \end{aligned}$$

4 Gros gros plan du cours

— Basic of deterministic optim
  — GD when L-Smooth
  — GD when not L-Smooth
— SGD
  — Tourne autour de la solution ($Var(\nabla F_i(\theta^\star)) = 1/3$)
  — Polyak averaging fix ça
  — Vanishing step size