# Descente de gradient sous contraintes
# Gradient projeté et Gradient conditionnel

10 février 2020

# Formalisation du problème

### Problème d'optimisation convexe

$$\begin{array}{ll} \min_{\mathbf{w}} & f(\mathbf{w}) \\ \text{tel que} & \mathbf{w} \in \mathcal{X} \end{array}$$

où $f$ est différentiable et $\mathcal{X}$ un ensemble convexe.

### Condition d'optimalité

- $\mathbf{w}^\star$ est solution du problème ssi

$$\nabla f(\mathbf{w}^\star)^\top(\mathbf{y} - \mathbf{w}^\star) \geq 0 \quad \forall y \in \mathcal{X}$$

# Projection sur un ensemble convexe

### Définition

- Soit $\mathcal{X}$ un ensemble convexe de $\mathbb{R}^d$, et $y$ un point de $\mathbb{R}^d$, on appelle la projection de $y$ sur $\mathcal{X}$, la solution du problème

$$\begin{array}{ll} \min_{\mathbf{w}} & \frac{1}{2}\|\mathbf{y} - \mathbf{w}\|_2^2 \\ \text{tel que} & \mathbf{w} \in \mathcal{X} \end{array}$$

### Exemples de projection simple

- projection sur l'orthant positif
- projection sur la boule unité de norme 2

# Algorithme de gradient projeté

## Iteration $k$

$$w^{k+1} = P_{\mathcal{X}}[\mathbf{w}^k - \alpha_k \nabla f(\mathbf{w}^k)]$$

où $P[\cdot]$ est l'opérateur de projection sur $\mathcal{X}$, et le pas $\alpha_k = \beta^{t_k}$ et $t_k$ le premier entier positif tel que

$$f(\mathbf{w}^{k+1}) \leq f(\mathbf{w}^k) + \sigma \nabla f(\mathbf{w}^k)^\top (\mathbf{w}^{k+1} - \mathbf{w}^k)$$

## Intuitions

- on réalise un pas de descente ...
- et on projette la solution sur l'ensemble admissible

## Propriété

- si le problème admet une solution, alors la séquence des $\mathbf{w}_{k+1}$ converge vers cette solution

# Interprétation du Gradient projeté

## Approximation quadratique

- si $f(\mathbf{w})$ est différentiable, on peut approximer $f$ en $\mathbf{w}_0$par

$$\tilde{f}(\mathbf{w}) = f(\mathbf{w}_0) + \nabla_{\mathbf{w}_0} f(\mathbf{w})^\top (\mathbf{w} - \mathbf{w}_0) + (\mathbf{w} - \mathbf{w}_0)^\top \mathbf{H} (\mathbf{w} - \mathbf{w}_0)$$

avec $\mathbf{H} = \frac{1}{2t}\mathbf{I}$

## Propriété

- lien avec la descente de gradient

$$\mathbf{w}_{k+1} = \arg\min_{\mathbf{w}} \tilde{f}(\mathbf{w}) = \mathbf{w}_k - t\nabla_{\mathbf{w}_k} f$$

- lien avec le gradient projeté

$$\mathbf{w}_{k+1} = \arg\min_{\mathbf{w}} \tilde{f}(\mathbf{w}) = P_{\mathcal{X}}(\mathbf{w}_k - t\nabla_{\mathbf{w}_k} f)$$

# Gradient conditionnel aka Algorithme de Franke-Wolfe

## Principe

- utilise une approximation linéaire de $f$
- a l'itératon $k$, on cherche un point qui vérifie

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}\in\mathcal{X}} \nabla_{\mathbf{w}^k} f^\top \mathbf{z}$$

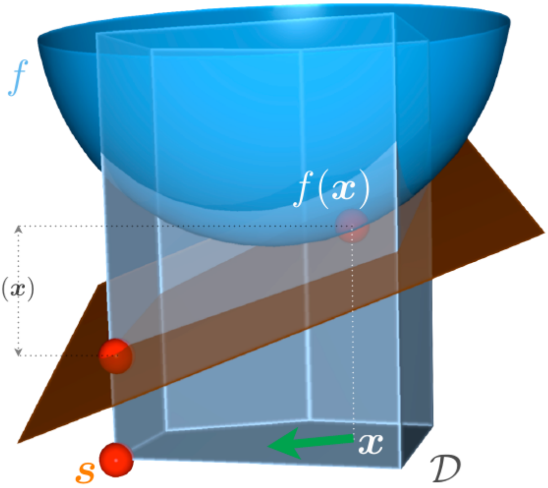
  $\mathbf{z}^{k+1}$ est un point admissible qui définit une direction de descente
- l'itération est

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \gamma_k(\mathbf{z}^{k+1} - \mathbf{w}^k) = (1-\gamma_k)\mathbf{w}^k + \gamma_k \mathbf{z}^{k+1}$$

- le pas $\gamma_k$ peut etre optimisé mais on peut également choisir un pas par défaut :

$$\gamma_k = \frac{2}{k+1}$$

# Illustration




(jaggi,2011)

# Propriétés

- Utile si
$$\arg\min_{\mathbf{z}\in\mathcal{X}} \nabla_{\mathbf{w}^k} f^\top \mathbf{z}$$
est facile à calculer. C'est le cas pour les contraintes de type $\|\mathbf{w}\|_p \leq t$
- Certificat d'optimalité
$$\max_{\mathbf{z}\in\mathcal{X}} \nabla_{\mathbf{w}} f(\mathbf{w}^k)^\top(\mathbf{w}^k - \mathbf{z}) \geq f(\mathbf{w}^{k-1}) - f^\star$$
- Vitesse de convergence
$$f(\mathbf{w}^{k-1}) - f^\star \leq \frac{2M}{k+2}$$
où $M$ est une constante dépendant de la courbure de $f$.