Problème d'optimisation sous contraintes

Conditions d'optimalité

Bo-Yuan HUANG

November 10, 2022

Plan

1. Introduction
2. Condition necéssaire d'optmialité
    - Théorème Karush-Kuhn-Tucker
    - Interprétation géométrique des conditions KKT
    - Qualification de contrainte
3. Conditions suffisantes d'optimalité
    - Cas 1 : Optimisation convexe
    - Cas 2 : Condition d'optimalité de second ordre
4. Dualité
    - Théorème de dualité
    - Interprétation géométrique
5. Application : un problème de classification par SVM

Problème d'optimisation

Un problème d'optimisation sous contraintes a pour forme générale :

$$(P) \quad \begin{cases} \min_x f(x) \\ g_i(x) \leq 0, i \in \mathcal{I} := \{1, ..., m\} \\ h_j(x) = 0, j \in \mathcal{E} := \{1, .., p\} \\ f, g, h \ sont \ \mathbb{R}^n \to \mathbb{R} \end{cases}$$

L'ensemble réalisable et le Lagrangien sont définis comme suite :

$$\mathcal{F} := \{x \in \mathbb{R}^n | \quad g_i(x) \leq 0 \ for \ i \in \mathcal{I}, h_j(x) = 0 \ for \ j \in \mathcal{E}\}$$

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_i h_j(x)$$

Condition necéssaire d'optimalité

Théorème Karush-Kuhn-Tucker

Si $x^* \in \mathcal{F}$ est un minimum local pour le problème (P) et qu'une "certaine qualification de contrainte" est satifaite, alors qu'il existe $\lambda \in \mathbb{R}_+^m$ and $\mu \in \mathbb{R}^p$ tel que:

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_i \nabla h_j(x^*) = 0$$

$$\lambda_i g_i(x^*) = 0 \ for \ i = 1, ..., m$$

Interprétation géométrique des conditions KKT

Il existe $\lambda_1, \lambda_2 \in \mathbb{R}_+$ tel que: $\nabla f(x^*) + \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*) = 0$

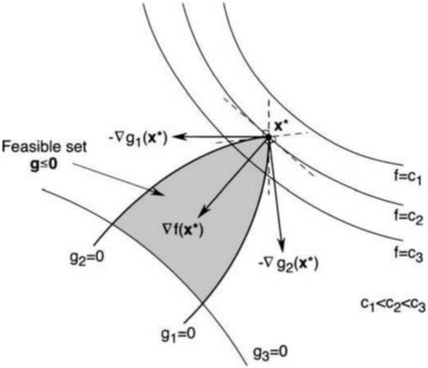


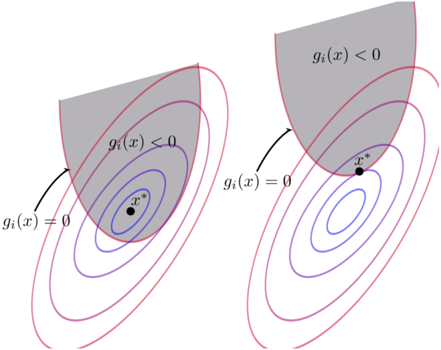
Figure: Condition de premier ordre

Interprétation géométrique des conditions KKT

Condition de relâchement supplémentaire: $\lambda_i g_i(x^*) = 0$ *for* $i = 1, ..., m$

Si x^* est situé à l'intérieur de $\mathcal{F}$, $g_i(x^*) < 0$, il suffit de choisir $\lambda_i = 0$ (contrainte inactive).
Si x^* est situé à la frontière de $\mathcal{F}$, $g_i(x^*) = 0$, quelconque λ_i réalise $\lambda_i g_i(x^*) = 0$ (contrainte active).
On note l'ensemble des indices actifs :

$$\mathcal{A}(x^*) := \{i \in \mathcal{I} | g_i(x^*) = 0\}$$



Qualification de contrainte

En fait, pour utiliser le théorème KKT, il suffit d'assurer l'égalité du "cône linéarisanté" et "cône tangent" (plus fort). Mais celle-ci est souvent difficile à réaliser.
On recherche avec compromis des conditions de la régularité appelées "qualification de contrainte" (QC) qui sont plus facile à vérifier .
Les QCs les plus couramment utilisés sont :

1. Slater : les g_i soient convexes, $\forall i = 1..m$ et $\exists \tilde{x} \in \mathcal{F}, g_i(\tilde{x}) < 0$.
2. Indépendance linéaire (QC-IL) : les vecteurs $(\nabla g_i(x^*)|i \in \mathcal{A}(x^*)$ et $(\nabla h_j(x^*)|j \in \mathcal{E})$ soient linéairement indépendents.

Cas 1 : Optimisation convexe

Précédemment, on a vu que les points de minimum local satisfont les conditions KKT. Si, de plus, les deux conditions suivantes sont vérifiées (optimisation convexe) :

- f et g_i sont convexes
- h_i sont linéaires (affines)

Alors les conditions KKT deviennent suffisantes et qu'un minimum local x^* trouvé avec KKT est un minimum global.

Si on trouve un point x^* satisfaisant KKT, mais la condition (1) ou (2) ci-dessus n'est pas vérifiée. Peut-t-on vérifier d'autre condition pour confirmer l'optimalité de x^* ?

Cas 2 : Condition d'optimalité de second ordre

Théorème : condition suffisante de second ordre

Si on a $x^* \in \mathcal{F}$ et $\lambda \in \mathbb{R}_+^m \geq 0$, $\mu \in \mathbb{R}^p$, qui vérifie les conditions KKT. Si on a de plus :

$$s^T \nabla_{xx}^2 L(x^*, \lambda, \mu) s > 0, \ \forall s \in \mathcal{F}_2$$

Alors x^* atteint un minimum local strict.

$$\begin{aligned}\mathcal{F}_2 := \{s | \nabla g_i(x^*)^T s \leq 0 \ &for \ i \in \mathcal{I}^0 \\ \nabla g_i(x^*)^T s = 0 \ &for \ i \in \mathcal{I}^+ \\ \nabla h_j(x^*) = 0 \ &for \ j \in \mathcal{E}\}\end{aligned}$$

où $\mathcal{I}^+ = \{i | \lambda_i > 0, for \ i \in \mathcal{I}\} \quad \mathcal{I}^0 = \{i | \lambda_i = 0, for \ i \in \mathcal{I}\}$

Dualité

Etant donné le problème primaire (P), quitte à convertir les contraintes d'égalité en inégalité, on introduit le problème (P') :

$$(P') \qquad \begin{cases} \min_x f(x) \\ g_i(x) \leq 0, i \in \mathcal{I} := 1, ..., m \\ x \in X \end{cases}$$

X pourrait être R^n, Z^n, Z_+^n ou bien $\{0,1\}^n$ qui rend le problème compliqué à résoudre.

Dualité

Pourquoi considère-t-on le problème dual ?

L'optimum du problème primaire (P), noté v(P) peut être difficile à résoudre. Le problème dual (D) est un problème fortement lié au problème primaire, il est toujours convexe donc plus simple à résoudre. Résoudre le problème dual nous aide à analyser le problème primaire, voire dans certains cas, à résoudre le problème primaire.

Construction du problème dual

On définit la fonction duale de Lagrange pour tout $\lambda \in \mathbb{R}_+$:

$$d(\lambda) = \min_{x \in X} L(x, \lambda) = \min_{x \in X} \{f(x) + \lambda g(x)\},$$

on a

$$d(\lambda) \leq \min_{x \in \mathcal{F}} \{f(x) + \lambda g(x)\} \leq \min_{x \in \mathcal{F}} f(x) = v(P)$$

Construction du problème dual

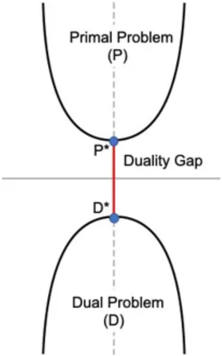
On en déduit $\forall \lambda \geq 0, d(\lambda) \leq v(P)$, qui donne une borne inférieur au problème (P). On s'intéresse donc à pouser la borne inférieur vers le haut pour "serrer" l'inégalité, donc on considère le problème :

$$(D) \qquad \begin{cases} \max_\lambda d(\lambda) = \max_\lambda \min\limits_{x \in X} L(x, \lambda) \\ s.t.\lambda > 0, i \in \mathcal{I} := 1, ..., m \end{cases}$$

On verra qu'il est possible de reformuler le problème (P') en $\min\limits_{x \in X} \max\limits_{\lambda \geq 0} L(x, \lambda)$

Théorème de dualité

On note $v(D)$ l'optimum du problème (D).



Théorème de dualité faible

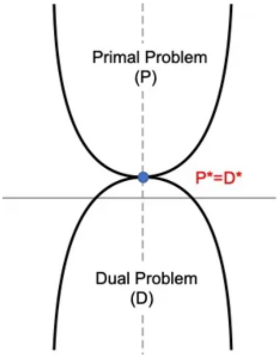
On a $\forall x \in \mathcal{F}, \lambda \geq 0$

$$d(\lambda) \leq v(D) \leq v(P) \leq f(x)$$

Théorème de dualité

Théorème de dualité forte

Si la qualification de contrainte Slater est vérifiée pour un problème convexe (P), on a $\exists \lambda \geq 0, x \in \mathcal{F}$, $d(\lambda) = v(D) = v(P) = f(x)$



Interprétation géométrique

Pour visualiser le domain de valeur de f et g, on introduit

$$G = \{(y, z) | g(x) = y, f(x) = z, x \in X\}$$

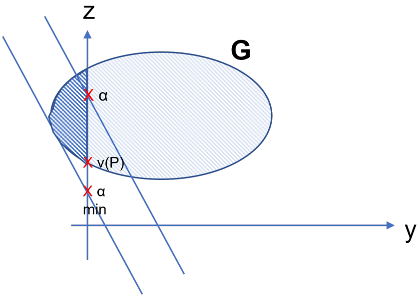
On rappel que

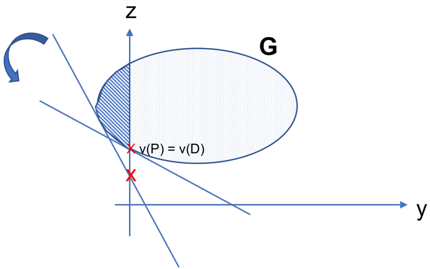
$$(P') \qquad \begin{cases} \min f(x) \\ s.t. g(x) \leq 0, x \in X \end{cases}$$

$$(D) \qquad \begin{cases} \max\limits_{\lambda \geq 0} d(\lambda) \\ d(\lambda) = \min\limits_{x \in X} \{f(x) + \lambda g(x)\} \end{cases}$$

Interprétation géométrique

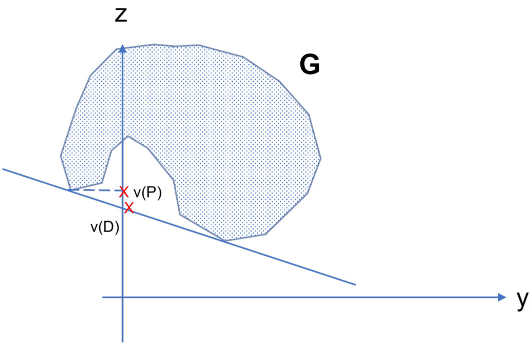
Soit $\alpha = z + \lambda y$, $-\lambda$ est la pente et α est l'intercept. Si G est convexe :





Interprétation géométrique

Soit $\alpha = z + \lambda y$, λ est la pente et α est l'intercept. Si G n'est pas convexe :



Interprétation géométrique

Réflexion 1

Si on trouve un point $\bar{x} \in \mathcal{F}, \lambda \geq 0$ tels que $d(\lambda) = f(\bar{x})$
Comme $d(\lambda) \leq v(D) \leq v(P) \leq f(\bar{x})$, on aura alors $v(P) = v(D)$ et $\bar{x}, \lambda$ sont l'optimum de (P) et de (D)

Réflexion 2

- Si $v(P) = -\infty$, alors $d(\lambda) = -\infty, \forall \lambda \geq 0$
- Si $v(D) = +\infty$, alors $v(P) = +\infty$, c'est-à-dire, (P) n'a pas de solution réalisable

Classification par SVM

Problème classification de tumeur

On dispose de 699 points de données (x_i, y_i), i $= 1..699$, avec 9 dimension pour chaque x_i, et $y_i \in \{1, -1\}$

Approche SVM

Trouver un hyperplan decrit par $< \omega, x > + \beta = 0$ avec $< \omega, x_i > + \beta \geq 1$ si $y_i = 1$ et $< \omega, x_i > + \beta \leq -1$ si $y_i = -1$ pour separer les points x_i dans deux groupes.

Classification par SVM

Problème optimisation quadratique

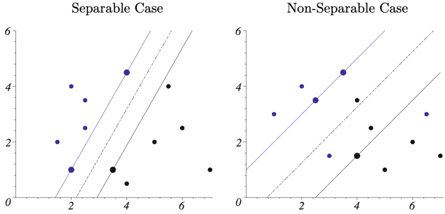
Maximiser la "marge" (distance) $2/\sqrt{<\omega,\omega>}$ entre les deux hyperplans $<\omega, x> +\beta = 1$ et $<\omega, x> +\beta = -1$

$$\begin{cases} \min_{\omega} \frac{1}{2} <\omega,\omega> \\ s.c. y_i(<\omega, x_i> +\beta) \geq 1, i = 1..m \end{cases}$$

SVM - Soft margin

Pour éviter les cas des points "inséparables", on introduit les terms de pénalités ξ_i et C. Le problème d'optimisation avec "soft margins" est formulé comme le suivant:

$$(P) \qquad \begin{cases} \min\limits_{\omega,\xi} f(\omega,\xi) = \frac{1}{2} < \omega, \omega > +C \sum\limits_{i=1}^{m} \xi_i \\ y_i(< \omega, x_i > +\beta) \geq 1 - \xi_i, \ \xi_i \geq 0, \ i \in \{1, ..., m\} \end{cases}$$



On remarque que dans (P), f et g sont convexes, et que la condition Slater est satisfaite.

SVM - Soft margin

Lagrangien du problème avec pénalité est

$$\begin{aligned}\mathcal{L} &= \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{m}\xi_i + \sum_{i=1}^{m}\lambda_i(1-\xi_i - y_i(\omega^T x_i + \beta)) + \sum_{i=1}^{m}\mu_i(0-\xi_i) \\ &= \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\lambda_i(y_i(\omega^T x_i + \beta) - 1 + \xi_i) - \sum_{i=1}^{m}\mu_i\xi_i\end{aligned}$$

En prenant les dérivées par rapport à ω, βet ξ_i

$$\partial_\omega\mathcal{L} = \omega - \sum_{i=1}^{m}\lambda_i y_i x_i = 0 \Rightarrow \omega = \sum_{i=1}^{m}\lambda_i y_i x_i$$

$$\partial_\beta\mathcal{L} = \sum_{i=1}^{m}\lambda_i y_i = 0$$

$$\partial_{\xi_i}\mathcal{L} = C - \lambda_i - \mu_i = 0$$

SVM - Soft margin

En remplaçant ω, on obtient la fonction duale:

$$\begin{aligned}\mathcal{L}_{dual} &= \frac{1}{2}||\omega||^2 - \omega^T(\sum_{i=1}^{m}\lambda_i y_i x_i) - \beta\sum_{i=1}^{m}\lambda_i y_i + \sum_{i=1}^{m}\lambda_i + \sum_{i=1}^{m}(C-\lambda_i-\mu_i)\xi_i \\ &= \sum_{i=1}^{m}\lambda_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j x_i^T x_j\end{aligned}$$

Le problème dual s'écrit donc :

$$\begin{cases} \max_{\lambda} \mathcal{L}_{dual} = \sum_{i=1}^{m}\lambda_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j x_i^T x_j \\ 0 \leq \lambda_i \leq C, i \in \{1, ..., m\} \\ \sum_{i=1}^{m}\lambda_i y_i = 0 \end{cases}$$

SVM - Solution numéric

Application du theorème de dualité forte

Grâce à la condition Slater qui est vérifiée, la solution du problème dual λ est telle que $\mathcal{L}_{dual}(\lambda) = \mathcal{L}(\omega, \xi, \lambda)$

On résout les λ_i du problème dual (convexe) grâce à un solveur. On peut ensuite obtenir

$$\omega = \sum_{i=1}^{m} \lambda_i y_i x_i$$

$$\beta = y_i - < \omega, x_i > \text{ , si } 0 < \lambda_i < C$$

On arrive finalement au classifieur : $f(x) := sign\{< \omega, x > +\beta\}$

SVM - Solution numéric

Table: Prediction results on testing sample (563 points)

		Predicted	
		Negative	Positive
Actual	Negative	True negative = 376	False positve = 3
	Positive	False negative = 46	True positive = 138

Table: Indicateurs clés

Precision	Recall	F1 score
$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FN}$	$2 \times \frac{Precision \times Recall}{Precision + Recall}$
$\frac{138}{138+3} = 97.87\%$	$\frac{138}{138+46} = 75\%$	$2 \times \frac{0.9787 * 0.75}{0.9787 + 0.75} = 84.9\%$

Merci pour votre attention