Metodi del Sottogradiente per la determinazione di Punti di Sella

Seminario di
Teoria e Metodi di Ottimizzazione

Docente: Prof. G.Bigi

A.A. 2015-2016

Carmine Frascella
28 giugno 2016

Indice

Capitolo 1

Introduzione e preliminari

1.1 Introduzione

Siano $X \subseteq \mathbb{R}^n$ e $M \subseteq R^m$ insiemi convessi e chiusi, e sia:

$$L : X \times M \to \mathbb{R}$$

una funzione convessa nella prima variabile, concava nella seconda.

Un **punto di sella** per la funzione L è una coppia $(x^*, \mu^*) \in X \times M$ tale che per ogni
$(x, \mu) \in X \times M$ valga:

$$L(x^*, \mu) \leq L(x^*, \mu^*) \leq L(x, \mu^*) \tag{1.1}$$

L'obiettivo del nostro seminario, allora, sarà quello di formulare degli algoritmi per determinare i punti di sella. Il problema appena formulato è interessante in diversi ambiti dell'ottimizzazione; in più, è noto da diversi anni che i metodi del sottogradiente risultano molto efficienti per il problema appena introdotto. Queste teorie si possono usare anche nell'ambito della teoria dei giochi (nell'articolo [4] è trattato un approccio alla programmazione lineare attraverso la teoria dell'equilibrio di Nash).

Alcuni articoli affrontano il problema sotto ipotesi aggiuntive (ad esempio nell'ipotesi in cui la funzione L si della forma $L(x, \mu) = L_1(x)L_2(\mu)$), in modo che il calcolo dei sottogradienti della funzione duale sia facilmente calcolabile.

In questo seminario presenteremo un algoritmo per la determinazione dei punti di sella, basato sul metodo del sottogradiente, che genererà una successione approssimante il punto di sella. Ci baseremo principalmente sui risultati esposti nell'articolo [6].

Nelle restanti parti del primo capitolo esporremo alcuni risultati preliminari, e formuleremo il problema che ci vogliamo porre.

Nel secondo capitolo esporremo un algoritmo basato sul metodo del sottogradiente proiettato utile per affrontare il problema della determinazione di un punto di sella, ed esporremo alcuni risultati a riguardo.

Nel terzo capitolo, dopo aver introdotto la nozione di dualità lagrangiana, mostreremo come entrambi il problema primale e quello duale possono essere risolti sfruttando la struttura della lagrangiana del problema, con i metodi esposti nel secondo capitolo. Daremo infine un risultato di limitatezza sotto un'ipotesi aggiuntiva ben precisa, ossia la condizione di Slater.

Nel quarto capitolo, quindi, modificheremo l'algoritmo presentato nel secondo capitolo per risolvere il problema primale-duale, e vedremo un importante risultato a riguardo.

Nell'ultimo capitolo, poi, evidenzieremo con un esempio esplicito alcuni limiti dei metodi proposti nei capitoli precedenti.

1.2 Nozioni preliminari

Nel seguito, chiamiamo $\mathbb{R}^n_+$ il sottoinsieme di $\mathbb{R}^n$ dei vettori con coordinate tutte nonnegative. Per un vettore $u \in \mathbb{R}^n$, indichiamo con u^+ la proiezione di u su $\mathbb{R}^n_+$:

$$u^+ = \begin{bmatrix} \max(u_1, 0) & \dots & \max(u_n, 0) \end{bmatrix}^t$$

Data una funzione convessa $F : \mathbb{R}^n \to \tilde{\mathbb{R}}$, si indica con $\mathrm{Dom}(F)$ il seguente insieme:

$$\mathrm{Dom}(F) \stackrel{\textbf{def}}{=} \{x \in \mathbb{R}^n \mid F(x) < +\infty\}$$

Dato un punto $\bar{x} \in \mathrm{Dom}(F)$, un vettore $v \in \mathbb{R}^n$ è detto **sottogradiente** di F in $\bar{x}$ se per ogni $x \in \mathrm{Dom}(F)$ vale la seguente stima dal basso:

$$F(x) \geq F(\bar{x}) + \langle v, x - \bar{x} \rangle \tag{1.2}$$

Dati F e $\bar{x}$ come sopra, l'insieme dei sottogradienti è detto **sottodifferenziale**, e si indica con $\partial F(\bar{x})$.

Data invece una funzione concava $q : \mathbb{R}^m \to \tilde{\mathbb{R}}$, si indica con $\mathrm{Dom}(q)$ il seguente insieme:

$$\mathrm{Dom}(q) \stackrel{\textbf{def}}{=} \{x \in \mathbb{R}^m \mid F(x) > -\infty\}$$

Dato un punto $\bar{x} \in \mathrm{Dom}(q)$, un vettore $v \in \mathbb{R}^m$ è detto **sopragradiente** di q in $\bar{x}$ se per ogni $x \in \mathrm{Dom}(q)$ vale la seguente stima dall'alto:

$$q(x) \leq q(\bar{x}) + \langle v, x - \bar{x} \rangle \tag{1.3}$$

Dati q e $\bar{x}$ come sopra, l'insieme dei sopragradienti è detto **sopradifferenziale**, e si indica con $\partial q(\bar{x})$.

Dato un insieme convesso e chiuso $C \subseteq \mathbb{R}^n$, indichiamo con $\mathbf{P}_C$ la (ben definita) **proiezione ortogonale** su C. Alcune proprietà di tale funzione sono le seguenti:

- Se $x = \mathbf{P}_C(y)$, allora per ogni $z \in C$ vale:
$$\langle y - x, z - x \rangle \leq 0 \; ; \tag{1.4}$$
- $\mathbf{P}_C$ è **non-espansiva**, ossia 1-Lipschitziana. Per ogni $y, z \in \mathbb{R}^n$, allora:
$$||\mathbf{P}_C(y) - \mathbf{P}_C(z)|| \leq ||y - z|| \; . \tag{1.5}$$

1.3 Formulazione del problema

Siano $X \subseteq \mathbb{R}^n$ e $M \subseteq R^m$ insiemi convessi e chiusi, e sia:
$$L : X \times M \to \mathbb{R}$$
una **funzione convessa-concava**, ossia convessa nella prima variabile per ogni $\mu \in M$ fissato, e concava nella seconda variabile per ogni $x \in X$ fissato. Il **problema del punto di sella** consiste nel determinare, se esiste:
$$\min_{x \in X} \max_{\mu \in M} L(x, \mu) \tag{1.6}$$
Una soluzione è una coppia $(x^*, \mu^*) \in X \times M$ tale che per ogni $(x, \mu) \in X \times M$ valga l'equazione (1.1). Essa è allora detta **punto di sella**.

Per un punto $(\bar{x}, \bar{\mu}) \in X \times M$, denotiamo con $\partial_x L(\bar{x}, \bar{\mu})$ il sottodifferenziale rispetto alla prima variabile, e con $\partial_\mu L(\bar{x}, \bar{\mu})$ il sopradifferenziale rispetto alla seconda variabile. Nell'ipotesi in cui questi insiemi siano non vuoti per ogni $(\bar{x}, \bar{\mu}) \in X \times M$, indichiamo un eventuale sottogradiente rispetto alla prima variabile con $L_x(\bar{x}, \bar{\mu})$, e un eventuale sopragradiente rispetto alla seconda variabile con $L_\mu(\bar{x}, \bar{\mu})$.

Capitolo 2

Algoritmo del sottogradiente per il problema del punto di sella

Presenteremo ora un metodo basato sul **metodo del sottogradiente proiettato**, per risolvere il problema del punto di sella. La generica iterazione è allora della forma:

$$\begin{cases} x_{k+1} = \mathbf{P}_X \left[x_k - \alpha L_x(x_k, \mu_k) \right] \\ \mu_{k+1} = \mathbf{P}_M \left[\mu_k + \alpha L_\mu(x_k, \mu_k) \right] \end{cases} , \tag{2.1}$$

ove $\mathbf{P}_X$ e $\mathbf{P}_M$ sono le proiezioni ortogonali su X e M rispettivamente, e:

$$L_x(x_k, \mu_k) \in \partial_x L(x_k, \mu_k) \quad , \quad L_\mu(x_k, \mu_k) \in \partial_\mu L(x_k, \mu_k)$$

Per inizializzare il metodo, si considera un punto arbitrario (x_0, μ_0).

Si può subito notare che il metodo è a passo costante ($\alpha > 0$), e non a passo variabile: siamo infatti interessati a quantificare la bontà dell'algoritmo dopo un certo numero finito di passi. Un passo variabile, invece, può essere utile se si vogliono dimostrare proprietà di convergenza al punto di sella nel caso in cui il numero di iterazioni tende all'infinito.

Chiameremo d'ora in avanti **algoritmo del sottogradiente proiettato** l'algoritmo dato dal sistema (2.1).

2.1 Primi risultati

Esaminiamo ora alcune prime proprietà delle successioni $(x_k)_{k \in \mathbb{N}}$ e $(\mu_k)_{k \in \mathbb{N}}$ ottenute utilizzando l'algoritmo del sottogradiente proiettato: tali risultati torneranno utili più avanti. Cominciamo con un lemma, che fornisce un primo risultato sotto ipotesi davvero semplici.

Lemma 2.1.1. Siano $(x_k)_{k\in\mathbb{N}}$ e $(\mu_k)_{k\in\mathbb{N}}$ le successioni ottenute usando l'algoritmo del sottogradiente proiettato. Allora:

- Per ogni $x \in X$ e per ogni $k \geq 0$:
$$||x_{k+1} - x||^2 \leq ||x_k - x||^2 - 2\alpha\left[L(x_k, \mu_k) - L(x, \mu_k)\right] + \alpha^2||L_x(x_k, \mu_k)||^2 \ ;$$
- Per ogni $\mu \in M$ e per ogni $k \geq 0$:
$$||\mu_{k+1} - \mu||^2 \leq ||\mu_k - \mu||^2 + 2\alpha\left[L(x_k, \mu_k) - L(x_k, \mu)\right] + \alpha^2||L_\mu(x_k, \mu_k)||^2 \ .$$

Dimostrazione. Dimostriamo la prima parte. Usando le formule (1.2), (1.5) e (2.1) e la convessità di L rispetto alla prima variabile si ha, per ogni $x \in X$ e per ogni $k \geq 0$:

$$\begin{aligned}||x_{k+1} - x||^2 = ||\mathbf{P}_X\left[x_k - \alpha L_x(x_k, \mu_k)\right] - \mathbf{P}_X(x)||^2 \leq ||x_k - \alpha L_x(x_k, \mu_k) - x||^2 \leq \\ \leq ||x_k - x||^2 + \alpha^2||L_x(x_k, \mu_k)||^2 - 2\alpha\langle L_x(x_k, \mu_k), x_k - x\rangle \leq \\ \leq ||x_k - x||^2 + \alpha^2||L_x(x_k, \mu_k)||^2 - 2\alpha\left[L(x_k, \mu_k) - L(x, \mu_k)\right] \ ,\end{aligned}$$

dato che vale:

$$\langle L_x(x_k, \mu_k), x - x_k\rangle \leq L(x, \mu_k) - L(x_k, \mu_k) \ ,$$

ossia:

$$-\langle L_x(x_k, \mu_k), x_k - x\rangle \leq -\left[L(x_k, \mu_k) - L(x, \mu_k)\right]$$

La dimostrazione della seconda parte è molto simile, perciò omettiamo i dettagli. Usando le formule (1.3), (1.5) e (2.1) e la concavità di L rispetto alla seconda variabile si ha, per ogni $\mu \in M$ e per ogni $k \geq 0$:

$$\begin{aligned}||\mu_{k+1} - \mu||^2 = ||\mathbf{P}_M\left[\mu_k + \alpha L_\mu(x_k, \mu_k)\right] - \mathbf{P}_M(\mu)||^2 \leq ||\mu_k + \alpha L_\mu(x_k, \mu_k) - \mu||^2 \leq \\ \leq ||\mu_k - \mu||^2 + \alpha^2||L_\mu(x_k, \mu_k)||^2 + 2\alpha\langle L_\mu(x_k, \mu_k), \mu_k - \mu\rangle \leq \\ \leq ||\mu_k - \mu||^2 + \alpha^2||L_\mu(x_k, \mu_k)||^2 + 2\alpha\left[L(x_k, \mu_k) - L(x_k, \mu)\right] \ ,\end{aligned}$$

come si voleva. □

Si possono ora mostrare altri risultati riguardo le successioni generate dall'algoritmo del gradiente proiettato, sotto un'ipotesi aggiuntiva.

D'ora in avanti, allora, supponiamo che ci sia **limitatezza dei sottogradienti e dei sopragradienti**. Supponiamo quindi che esista $h > 0$ tale che per ogni $k \geq 0$ valga:

$$\begin{cases} ||L_x(x_k, \mu_k)|| \leq h \\ ||L_\mu(x_k, \mu_k)|| \leq h \end{cases} \tag{2.2}$$

L'ipotesi è verificata se ad esempio X e M sono compatti, e L è una funzione convessa-concava definita su tutto $\mathbb{R}^n \times \mathbb{R}^m$.

Lemma 2.1.2. Siano $(x_k)_{k\in\mathbb{N}}$ e $(\mu_k)_{k\in\mathbb{N}}$ le successioni ottenute usando l'algoritmo del sottogradiente proiettato. Supponiamo poi che valga l'ipotesi di limitatezza dei sottogradienti e dei sopragradienti, data dal sistema (2.2). Siano allora, per ogni $k > 0$:

$$\hat{x}_k \stackrel{\textbf{def}}{=} \frac{1}{k}\sum_{j=0}^{k-1} x_j \quad , \quad \hat{\mu}_k \stackrel{\textbf{def}}{=} \frac{1}{k}\sum_{j=0}^{k-1} \mu_j$$

Allora, per ogni $k \geq 1$:

- Per ogni $x \in X$:

$$\frac{1}{k}\sum_{j=0}^{k-1} L(x_j,\mu_j) - L(x,\hat{\mu}_k) \leq \frac{||x_0 - x||^2}{2\alpha k} + \frac{\alpha h^2}{2} \tag{2.3}$$

- Per ogni $\mu \in M$:

$$-\frac{||\mu_0 - \mu||^2}{2\alpha k} - \frac{\alpha h^2}{2} \leq \frac{1}{k}\sum_{j=0}^{k-1} L(x_j,\mu_j) - L(\hat{x}_k,\mu) \tag{2.4}$$

Dimostrazione. Dimostriamo prima la formula (2.3). Usando la prima parte del lemma (2.1.1) e l'ipotesi di limitatezza dei sottogradienti data dal sistema (2.2), si ha per ogni $x \in X$ e per ogni $j \geq 0$:

$$||x_{j+1} - x||^2 \leq ||x_j - x||^2 - 2\alpha\left[L(x_j,\mu_j) - L(x,\mu_j)\right] + \alpha^2||L_x(x_j,\mu_j)||^2 \leq$$

$$\leq ||x_j - x||^2 - 2\alpha\left[L(x_j,\mu_j) - L(x,\mu_j)\right] + \alpha^2 h^2 \ ,$$

da cui:

$$L(x_j,\mu_j) - L(x,\mu_j) \leq \frac{1}{2\alpha}\left[||x_j - x||^2 - ||x_{j+1} - x||^2\right] + \frac{\alpha h^2}{2}$$

Sommando per $j = 0, \dots, k-1$:

$$\sum_{j=0}^{k-1}\left[L(x_j,\mu_j) - L(x,\mu_j)\right] \leq \frac{1}{2\alpha}\left[||x_0 - x||^2 - ||x_k - x||^2\right] + \frac{k\alpha h^2}{2}$$

Dunque, dividendo per k:

$$\frac{1}{k}\sum_{j=0}^{k-1} L(x_j,\mu_j) - \frac{1}{k}\sum_{j=0}^{k-1} L(x,\mu_j) \leq \frac{1}{2\alpha k}\left[||x_0 - x||^2 - ||x_k - x||^2\right] + \frac{\alpha h^2}{2} \leq$$

$$\leq \frac{||x_0 - x||^2}{2\alpha k} + \frac{\alpha h^2}{2}$$

Sfruttando ora la concavità di L rispetto alla seconda variabile ($x \in X$ è fissato), si ha:

$$L(x, \hat{\mu}_k) \geq \frac{1}{k} \sum_{j=0}^{k-1} L(x, \mu_j) \ ,$$

e da qui si ottiene la relazione (2.3):

$$\frac{1}{k} \sum_{j=0}^{k-1} L(x_j, \mu_j) - L(x, \hat{\mu}_k) \leq \frac{||x_0 - x||^2}{2\alpha k} + \frac{\alpha h^2}{2}$$

Dimostriamo ora la formula (2.4). Usando la seconda parte del lemma (2.1.1) e l'ipotesi di limitatezza dei sopragradienti data dal sistema (2.2), si ha per ogni $\mu \in M$ e per ogni $j \geq 0$:

$$||\mu_{j+1} - \mu||^2 \leq ||\mu_j - \mu||^2 + 2\alpha \left[L(x_j, \mu_j) - L(x_j, \mu)\right] + \alpha^2 ||L_\mu(x_j, \mu_j)||^2 \leq$$
$$\leq ||\mu_j - \mu||^2 + 2\alpha \left[L(x_j, \mu_j) - L(x_j, \mu)\right] + \alpha^2 h^2 \ ,$$

da cui:

$$L(x_j, \mu_j) - L(x_j, \mu) \geq -\frac{1}{2\alpha} \left[||\mu_j - \mu||^2 - ||\mu_{j+1} - \mu||^2\right] - \frac{\alpha h^2}{2}$$

Sommando per $j = 0, \dots, k-1$:

$$\sum_{j=0}^{k-1} \left[L(x_j, \mu_j) - L(x_j, \mu)\right] \geq -\frac{1}{2\alpha} \left[||\mu_0 - \mu||^2 - ||\mu_k - \mu||^2\right] - \frac{k\alpha h^2}{2}$$

Dunque, dividendo per k:

$$\frac{1}{k} \sum_{j=0}^{k-1} L(x_j, \mu_j) - \frac{1}{k} \sum_{j=0}^{k-1} L(x_j, \mu) \geq -\frac{1}{2\alpha k} \left[||\mu_0 - \mu||^2 - ||\mu_k - \mu||^2\right] - \frac{\alpha h^2}{2} \geq$$
$$\geq -\frac{||\mu_0 - \mu||^2}{2\alpha k} - \frac{\alpha h^2}{2}$$

Sfruttando ora la convessità di L rispetto alla prima variabile ($\mu \in M$ è fissato), si ha:

$$L(\hat{x}_k, \mu) \leq \frac{1}{k} \sum_{j=0}^{k-1} L(x_j, \mu) \ ,$$

e da qui si ottiene la relazione (2.4):

$$\frac{1}{k} \sum_{j=0}^{k-1} L(x_j, \mu_j) - L(\hat{x}_k \, \mu) \geq -\frac{||\mu_0 - \mu||^2}{2\alpha k} - \frac{\alpha h^2}{2} \ ,$$

e ciò conclude la dimostrazione. □

Si può vedere con un esempio che la stima ottenuta è *sharp*, nel senso che vi sono in casi in cui si ha esattamente quell'errore. Siano $\alpha > 0, h > 0$, e sia $f : \mathbb{R} \to \mathbb{R}$ data da:

$$f(x) \stackrel{\textbf{def}}{=} h|x|$$

Nel nostro esempio, sia allora $L(x, \mu) = f(x)$. Dunque, per $x \neq 0$:

$$\begin{cases} f'(x) = h, & x > 0 \\ f'(x) = -h, & x < 0 \end{cases},$$

da cui se $x_0 = \frac{\alpha h}{2}$, si ha:

$$x_1 = x_0 - \alpha h = -\frac{\alpha h}{2} = -x_0$$

$$x_2 = x_1 + \alpha h = \frac{\alpha h}{2} = x_0$$

La successione generata, dunque, è periodica con periodo 2. In più per ogni $k \geq 0$ vale:

$$f(x_k) = \frac{\alpha h^2}{2} ,$$

da cui la stima sull'errore risulta ottima, dato che il punto di minimo è chiaramente $x^* = 0$.

2.2 Punti di sella approssimati

In questa sezione porremo l'attenzione sulla costruzione di una successione di **punti di sella approssimati**, usando il metodo del sottogradiente proiettato (presentato all'inizio di questo capitolo, e dato dalla formula (2.1)).

Per fissare la notazione, date due successioni $(x_k)_{k\in\mathbb{N}}$ e $(\mu_k)_{k\in\mathbb{N}}$, chiameremo $(\hat{x}_k)_{k\in\mathbb{N}^+}$ e $(\hat{\mu}_k)_{k\in\mathbb{N}^+}$ le successioni delle medie aggiornate, definite dalle seguenti formule:

$$\hat{x}_k \stackrel{\textbf{def}}{=} \frac{1}{k}\sum_{j=0}^{k-1} x_j \quad , \quad \hat{\mu}_k \stackrel{\textbf{def}}{=} \frac{1}{k}\sum_{j=0}^{k-1} \mu_j$$

Proposizione 2.2.1. Supponiamo che valga l'ipotesi di limitatezza dei sottogradienti e dei sopragradienti, data dal sistema (2.2). Siano $(x_k)_{k\in\mathbb{N}}$ e $(\mu_k)_{k\in\mathbb{N}}$ le successioni generate dal metodo del sottogradiente proiettato; siano $(\hat{x}_k)_{k\in\mathbb{N}^+}$ e $(\hat{\mu}_k)_{k\in\mathbb{N}^+}$ le successioni delle medie aggiornate; sia $(x^*, \mu^*) \in X \times M$ un punto di sella per L. Allora per ogni $k \geq 1$ vale:

$$-\frac{||\mu_0 - \mu^*||^2}{2\alpha k} - \frac{\alpha h^2}{2} \leq \frac{1}{k}\sum_{j=0}^{k-1} L(x_j, \mu_j) - L(x^*, \mu^*) \leq \frac{||x_0 - x^*||^2}{2\alpha k} + \frac{\alpha h^2}{2} \tag{2.5}$$

e:

$$-\frac{||\mu_0 - \mu^*||^2}{2\alpha k} - \frac{||x_0 - \hat{x}_k||^2}{2\alpha k} - \alpha h^2 \leq L(\hat{x}_k, \hat{\mu}_k) - L(x^*, \mu^*) \leq$$
$$\leq \frac{||x_0 - x^*||^2}{2\alpha k} + \frac{||\mu_0 - \hat{\mu}_k||^2}{2\alpha k} + \alpha h^2 \tag{2.6}$$

Dimostrazione. Innanzitutto (x^*, μ^*) è un punto di sella, dunque vale per ogni $k \geq 1$:

$$L(x^*, \hat{\mu}_k) \leq L(x^*, \mu^*) \leq L(\hat{x}_k, \mu^*)$$

Sulla scorta di ciò, sfruttando il lemma (2.1.2) e in particolare le relazioni (2.3) e (2.4) con $x = x^*$ e $\mu = \mu^*$ si ottiene allora la prima parte della tesi:

$$-\frac{||\mu_0 - \mu^*||^2}{2\alpha k} - \frac{\alpha h^2}{2} \leq \frac{1}{k}\sum_{j=0}^{k-1} L(x_j, \mu_j) - L(x^*, \mu^*) \leq \frac{||x_0 - x^*||^2}{2\alpha k} + \frac{\alpha h^2}{2} \tag{2.7}$$

Ora, per ogni $k \geq 1$ si ha $\hat{x}_k \in X$ e $\hat{\mu}_k \in M$, dunque usando il lemma (2.1.2), e in particolare le relazioni (2.3) e (2.4) con $x = \hat{x}_k$ e $\mu = \hat{\mu}_k$, si ottengono le seguenti relazioni:

$$\begin{cases} \frac{1}{k}\sum_{j=0}^{k-1} L(x_j, \mu_j) - L(\hat{x}_k, \hat{\mu}_k) \leq \frac{||x_0 - \hat{x}_k||^2}{2\alpha k} + \frac{\alpha h^2}{2} \\ -\frac{||\mu_0 - \hat{\mu}_k||^2}{2\alpha k} - \frac{\alpha h^2}{2} \leq \frac{1}{k}\sum_{j=0}^{k-1} L(x_j, \mu_j) - L(\hat{x}_k, \hat{\mu}_k) \end{cases}$$

Cambiando i segni e invertendo le disuguaglianze, si ottiene allora:

$$-\frac{||x_0 - \hat{x}_k||^2}{2\alpha k} - \frac{\alpha h^2}{2} \leq L(\hat{x}_k, \hat{\mu}_k) - \frac{1}{k}\sum_{j=0}^{k-1} L(x_j, \mu_j) \leq \frac{||\mu_0 - \hat{\mu}_k||^2}{2\alpha k} + \frac{\alpha h^2}{2} \;, \tag{2.8}$$

e sommando membro a membro le catene di disuguaglianze (2.7) e (2.8) si ottiene la seconda parte della tesi. □

La proposizione appena dimostrata ci dà allora i primi risultati di approssimazione. In particolare abbiamo dimostrato che:

- La media delle immagini delle coppie della successione $((x_k, \mu_k))_{k\in\mathbb{N}}$ convergono a $L(x^*, \mu^*)$ con un errore quantificabile in $\frac{\alpha h^2}{2}$ e con velocità di convergenza $\frac{1}{k}$;
- L'immagine della coppia delle medie aggiornate converge a $L(x^*, \mu^*)$ con un errore quantificabile in αh^2 e con velocità di convergenza $\frac{1}{k}$.

L'errore è dovuto all'uso di un passo costante $\alpha > 0$, e può essere regolato usando un passo costante più piccolo in modulo.

Capitolo 3

Dualità lagrangiana

In questo capitolo, introdurremo la nozione di **dualità lagrangiana**: solitamente si ha un problema di ottimizzazione in ipotesi di convessità, detto **problema primale**, cui si associa un altro problema di ottimizzazione, detto **problema duale**, definito a partire da una funzione, detta **lagrangiana** del problema, la quale risulta convessa-concava. Iniziamo dunque introducendo le nozioni di problema primale e di problema duale: nel prossimo capitolo, poi, vedremo come applicare il metodo del sottogradiente proiettato (o meglio, una sua leggera variazione) in questo contesto.

3.1 Problema primale e problema duale

Consideriamo il seguente problema di ottimizzazione vincolata:

$$\min_{x \in X \cap C} f(x) \ ,$$

ove:

- $X \in \mathbb{R}^n$ è un insieme non vuoto, convesso e chiuso;
- $f : \mathbb{R}^n \to \mathbb{R}$ è una funzione convessa;
- $C = \{ \ x \in \mathbb{R}^n \mid \ \forall \ j = 1, \ldots, m : g_j(x) \leq 0 \ \}$, ove per ogni $j = 1, \ldots, m$ la funzione $g_j : \mathbb{R}^n \to \mathbb{R}$ è convessa.

Questo problema sarà chiamato, d'ora in poi, **problema primale**, e il suo valore ottimale (il minimo, o comunque l'estremo inferiore) sarà denotato con f^*.

In tale contesto, si definisce **lagrangiana** del problema la funzione:

$$L : X \times \mathbb{R}^m_+ \to \mathbb{R}$$

data da:

$$L(x,\mu) \overset{\mathbf{def}}{=} f(x) + \langle \mu, \mathbf{g}(x) \rangle = f(x) + \sum_{j=1}^{m} \mu_j g_j(x)$$

Tramite tale lagrangiana si definisce allora la **funzione obiettivo duale**:

$$q : \mathbb{R}^m_+ \to \mathbb{R} \ ,$$

data da:

$$q(\mu) \overset{\mathbf{def}}{=} \inf_{x \in X} L(x,\mu) \ ,$$

quindi il **problema duale**:

$$\max_{\mu \in \mathbb{R}^m_+} q(\mu)$$

Denoteremo d'ora in avanti il valore ottimale del problema duale con q^*, e l'insieme dei punti di minimo del problema duale con M^*.

Il seguente teorema è facilmente dimostrabile.

Teorema 3.1.1 (di Dualità debole). Siano $x \in X \cap C$, e $\mu \in \mathbb{R}^m_+$. Allora:

$$q(\mu) \leq f(x)$$

Come conseguenza, $q^* \leq f^*$.

Dimostrazione. Vale innanzitutto:

$$q(\mu) \leq L(x,\mu) \leq f(x) \ ,$$

per ogni $\mu \in \mathbb{R}^m_+$ e $x \in X \cap C$. Ora, fissato $x \in X \cap C$, si ha:

$$q^* = \sup_{\mu \in \mathbb{R}^m_+} q(\mu) \leq \sup_{\mu \in \mathbb{R}^m_+} L(x,\mu) \leq f(x) \ ,$$

da cui $q^* \leq f(x)$. Passando allora all'estremo inferiore si ha allora $q^* \leq f^*$, e con ciò la dimostrazione è conclusa. $\square$

Come conseguenza, se $(\bar{x}, \bar{\mu}) \in (X \cap C) \times \mathbb{R}^m_+$ è tale che $f(\bar{x}) = q(\bar{\mu})$, allora $\bar{x}$ è punto di minimo per il problema primale, $\bar{\mu}$ è punto di massimo per il problema duale (quindi $M^* \neq \varnothing$) e $q^* = f^*$. Si dice allora che il **gap di dualità** $(f^* - q^*)$, a priori nonnegativo, è nullo. Al contrario, se il gap di dualità è strettamente positivo, non esiste una coppia $(\bar{x}, \bar{\mu})$ con le caratteristiche appena citate.

Un eventuale coppia ottimale è certamente un punto di sella per la lagrangiana L, dato che di sicuro vale:

$$L(\bar{x}, \mu) \leq L(\bar{x}, \bar{\mu}) \leq L(x, \bar{\mu})$$

per ogni $x \in X \cap C$ e per ogni $\mu \in \mathbb{R}^m_+$. Dato inoltre che f è convessa, che q è concava (ciò si dimostra facilmente), e gli insiemi $X \cap C$ e $\mathbb{R}^m_+$ sono convessi e chiusi, possiamo usare il metodo del sottogradiente proiettato, presentato nello scorso capitolo, per determinare un punto di sella (ossia una soluzione del problema primale-duale).

L'uso del metodo del sottogradiente per la risoluzione del problema, però, presuppone che il calcolo dei sopradifferenziali per la funzione concava q sia agevole. Notiamo però che, per come è definita q, i sopragradienti sono dipendenti dai punti di minimo della lagrangiana rispetto alla prima variabile. In altre parole, fissato $\mu \in \mathbb{R}^m_+$, la determinazione di $\partial q(\mu)$ presuppone la conoscenza dell'insieme seguente:

$$X_\mu \stackrel{\textbf{def}}{=} \{\ x \in X \cap C \mid q(\mu) = L(x, \mu)\ \}$$

Una volta noto quest'insieme, il **teorema di Danskin** afferma che vale:

$$\mathrm{Conv}(\mathbf{g}(x) \mid x \in X_\mu) \subseteq \partial q(\mu)$$

Non vediamo la dimostrazione di ciò: se si è interessati, si veda [2]. Il metodo del sottogradiente proiettato, dunque, presuppone che la determinazione di X_μ sia agevole per ogni $\mu \in \mathbb{R}^m_+$, e purtroppo in molti casi la lagrangiana $L(x, \mu)$ non presenta strutture particolari tali per cui tale determinazione sia agevole, computazionalmente parlando. Vogliamo allora, nel resto del seminario, approfondire questa questione, sotto ipotesi aggiuntive.

Cominciamo enunciando il seguente teorema, che in effetti abbiamo dimostrato pocanzi.

Teorema 3.1.2 (del Punto di Sella). Una coppia $(\bar{x}, \bar{\mu}) \in (X \cap C) \times \mathbb{R}^m_+$ è una soluzione ottimale per il problema primale-duale se e solo se essa è un punto di sella per la lagrangiana del problema L.

L'obiettivo è allora quello di usare quanto fatto nella sezione (2.2) per determinare soluzioni approssimate per il problema primale-duale, e per stimare l'errore commesso dopo un certo numero di iterazioni del metodo.

3.2 Un risultato di limitatezza

In questa sezione mostreremo che, sotto un'opportuna ipotesi aggiuntiva, l'insieme dei punti di minimo del problema duale M^* risulta limitato.

L'ipotesi aggiuntiva è detta **condizione di Slater**, e assume l'esistenza di un **vettore di Slater**, ossia di un vettore $\tilde{x} \in \mathbb{R}^n$ tale che per ogni $j = 1, \ldots, m$ valga $g_j(\tilde{x}) < 0$.

Sappiamo già che, se vale la condizione di Slater, il gap di dualità è nullo, e quindi esiste una soluzione del problema primale-duale (per delle dimostrazioni di ciò, si veda [1] oppure [2]). Vogliamo ora dimostrare che, in aggiunta, l'insieme dei punti di minimo del problema duale M^* risulta limitato. In effetti, con la proposizione che segue mostreremo un risultato più generale: la dimostrazione è tratta dall'articolo [5].

Proposizione 3.2.1. Sia $\bar{\mu} \in \mathbb{R}^m_+$ un vettore, e sia:

$$Q_{\bar{\mu}} \stackrel{\textbf{def}}{=} \{\ \mu \in \mathbb{R}^m_+ \mid q(\mu) \geq q(\bar{\mu})\ \}$$

Supponiamo poi che valga la condizione di Slater. Allora l'insieme $Q_{\bar{\mu}}$ è limitato, e in più per ogni $\mu \in Q_{\bar{\mu}}$ vale:

$$||\mu||_1 \leq \frac{1}{\gamma}\left[f(\tilde{x}) - q(\bar{\mu})\right]\ ,$$

dove $\tilde{x} \in \mathbb{R}^n$ è un vettore di Slater, e:

$$\gamma \stackrel{\textbf{def}}{=} \min_{j=1,\dots,m}\{-g_j(\tilde{x})\}$$

Dimostrazione. Sia $\mu \in Q_{\bar{\mu}}$. Allora vale:

$$q(\bar{\mu}) \leq q(\mu) = \inf_{x \in X} L(x,\mu) \leq L(\tilde{x},\mu) = f(\tilde{x}) + \sum_{j=1}^{m} \mu_j g_j(\tilde{x})\ ,$$

da cui:

$$-\sum_{j=1}^{m} \mu_j g_j(\tilde{x}) \leq f(\tilde{x}) - q(\bar{\mu})$$

Dato che $\mu \in \mathbb{R}^m_+$ e che $\tilde{x}$ è un vettore di Slater, vale allora:

$$\gamma \sum_{j=1}^{m} \mu_j \leq -\sum_{j=1}^{m} \mu_j g_j(\tilde{x}) \leq f(\tilde{x}) - q(\bar{\mu})\ ,$$

e da qui si ottiene la tesi mediante una facile divisione. □

In particolare, se $\bar{\mu}$ appartiene all'insieme M^*, allora per ogni $\mu \in M^*$ vale:

$$||\mu||_1 \leq \frac{1}{\gamma}\left[f(\tilde{x}) - q^*\right] \tag{3.1}$$

Tale stima tornerà utile più avanti.

Capitolo 4

Metodo del sottogradiente primale-duale

Vogliamo ora applicare il metodo del sottogradiente proiettato alla funzione lagrangiana $L(x, \mu)$ incontrata nel capitolo precedente: d'ora in avanti, dunque, sarà $M = \mathbb{R}^m_+$. In tal caso, non è detto che la successione $(\mu_k)_{k\in\mathbb{N}}$ generata dall'algoritmo sia limitata, quindi non è detto che vi sia limitatezza dei sopragradienti (e questa era un'assunzione che si faceva, prima di applicare il metodo).

L'idea è allora quella di guadagnare la limitatezza della successione assumendo la validità della condizione di Slater. In tale ipotesi, possiamo allora considerare un **algoritmo del sottogradiente primale-duale**, della forma:

$$\begin{cases} x_{k+1} = \mathbf{P}_X \left[x_k - \alpha L_x(x_k, \mu_k) \right] \\ \mu_{k+1} = \mathbf{P}_D \left[\mu_k + \alpha L_\mu(x_k, \mu_k) \right] \end{cases} , \tag{4.1}$$

ove $D \subseteq M$ è un opportuno sottoinsieme convesso e compatto. Dato $\mu' \in M$ arbitrario, se $q' = q(\mu')$, allora per ogni $\mu^* \in M^*$ vale, ricordando la relazione (3.1):

$$||\mu^*||_1 \leq \frac{1}{\gamma} \left[f(\tilde{x}) - q' \right] \ ,$$

dato che $q' \leq q^*$. Possiamo allora scegliere il seguente insieme:

$$D \overset{\mathbf{def}}{=} \left\{ \mu \in M \ \middle| \ ||\mu|| \leq \frac{1}{\gamma} \left[f(\tilde{x}) - q' \right] + r \right\} \ ,$$

con $r > 0$ arbitrario (si ricordi che, per $v \in \mathbb{R}^m$, vale $||v|| \leq ||v||_1$).

Definito dunque l'insieme D, possiamo allora affermare che esiste una costante $h > 0$ tale che per ogni $k \geq 0$ valga:

$$\begin{cases} ||L_x(x_k, \mu_k)|| \leq h \\ ||L_\mu(x_k, \mu_k)|| \leq h \end{cases}, \tag{4.2}$$

ove $(x_k)_{k\in\mathbb{N}}$ e $(\mu_k)_{k\in\mathbb{N}}$ sono generate dall'algoritmo del sottogradiente primale-duale.

4.1 Soluzioni approssimate per il problema primale

Useremo ora l'algoritmo del sottogradiente proiettato per generare una successione di soluzioni approssimate per il problema primale.

Fissiamo allora la notazione, prima di enunciare e dimostrare una proposizione. Fissiamo un punto arbitrario $(x_0, \mu_0) \in X \times M$. Sia poi $\tilde{x}$ un vettore di Slater (stiamo supponendo che valga la condizione di Slater), e siano $\gamma > 0$, $\tilde{q} \leq q^*$, $r > 0$, $D \subseteq M$ definiti come nella sezione (3.2). Generiamo quindi, mediante l'algoritmo primale-duale (dato dal sistema (4.1)), le successioni $(x_k)_{k\in\mathbb{N}}$ e $(\mu_k)_{k\in\mathbb{N}}$. Sia infine $h > 0$ la costante che limita i sottogradienti e i sopragradienti relativi ai punti delle due successioni (si veda il sistema (4.2)).

Se consideriamo la successione $(\hat{x}_k)_{k\in\mathbb{N}}$ data dalle medie correnti (si veda il lemma (2.1.2) per la definizione), non è detto che valga, per ogni $k \in \mathbb{N}^+$ e per ogni $j = 1, \ldots, m$, $g_j(\hat{x}_k) \leq 0$, in quanto ciò non vale nemmeno per la successione $(x_k)_{k\in\mathbb{N}}$, dato che la proiezione è ogni volta svolta su X, e non su $X \cap C$ (si veda la sezione (3.1) per la definizione di C).

Proposizione 4.1.1. Nelle ipotesi appena enunciate, per ogni $k \geq 1$ valgono le seguenti proprietà:

- Una limitazione superiore alla violazione del vincolo $||\mathbf{g}(\hat{x}_k)^+||$, data da:

$$||\mathbf{g}(\hat{x}_k)^+|| \leq \frac{2}{k\alpha r}\left(\frac{f(\tilde{x}) - \tilde{q}}{\gamma} + r\right)^2 + \frac{||x_0 - x^*||^2}{2k\alpha r} + \frac{\alpha h^2}{2r} \ ;$$

- Una limitazione superiore a $f(\hat{x}_k) - f^*$, data da:

$$f(\hat{x}_k) - f^* \leq \frac{||\mu_0||^2}{2k\alpha} + \frac{||x_0 - x^*||^2}{2k\alpha} + \alpha h^2 \ ;$$

- Una limitazione inferiore a $f(\hat{x}_k) - f^*$, data da:

$$f(\hat{x}_k) - f^* \geq -\frac{f(\tilde{x}) - \tilde{q}}{\gamma} ||\mathbf{g}(\hat{x}_k)^+|| \ .$$

Dimostrazione. La dimostrazione della prima proprietà è molto tecnica e noiosa, perciò limitiamoci a dimostrare le ultime due, relative alla quantità $f(\hat{x}_k) - f^*$.

Partiamo quindi dalla seconda proprietà. Per convessità di f e per definizione di L, si ha:

$$f(\hat{x}_k) \leq \frac{1}{k}\sum_{j=0}^{k-1} f(x_j) = \frac{1}{k}\sum_{j=0}^{k-1} L(x_j, \mu_j) - \frac{1}{k}\sum_{j=0}^{k-1} \langle \mu_j, \mathbf{g}(x_j) \rangle$$

Dunque:

$$f(\hat{x}_k) - f^* \leq \frac{1}{k}\sum_{j=0}^{k-1} L(x_j, \mu_j) - f^* - \frac{1}{k}\sum_{j=0}^{k-1} \langle \mu_j, \mathbf{g}(x_j) \rangle \leq$$

$$\leq \frac{||x_0 - x^*||^2}{2\alpha k} + \frac{\alpha h^2}{2} - \frac{1}{k}\sum_{j=0}^{k-1} \langle \mu_j, \mathbf{g}(x_j) \rangle \ ,$$

ove l'ultima disuguaglianza è giustificata dalla formula (2.5).

Resta da stimare l'ultima disuguaglianza. Notiamo che $\mathbf{0} \in D$, perciò applicando la seconda formula del lemma (2.1.1) si ha, per ogni $j \geq 0$:

$$||\mu_{j+1}||^2 \leq ||\mu_j||^2 + 2\alpha \langle \mu_j, \mathbf{g}(x_j) \rangle + \alpha^2 h^2 \ ,$$

da cui:

$$-2\alpha \langle \mu_j, \mathbf{g}(x_j) \rangle \leq ||\mu_j||^2 - ||\mu_{j+1}||^2 + \alpha^2 h^2$$

Sommando per $j = 0, \ldots, k-1$ e dividendo per k si ha dunque:

$$-\frac{1}{k}\sum_{j=0}^{k-1} \langle \mu_j, \mathbf{g}(x_j) \rangle \leq \frac{||\mu_0||^2}{2\alpha k} + \frac{\alpha h^2}{2} \ ,$$

quindi si ottiene la proprietà voluta.

Passiamo ora alla terza proprietà. Se μ^* è una soluzione ottimale, allora certamente vale, per $k \geq 1$:

$$f(\hat{x}_k) = L(\hat{x}_k, \mu^*) - \langle \mu^*, \mathbf{g}(\hat{x}_k) \rangle$$

Per il teorema del Punto di Sella (teorema (3.1.2)) si ha inoltre:

$$L(\hat{x}_k, \mu^*) \geq L(x^*, \mu^*) = f^*$$

Quindi:

$$f(\hat{x}_k) \geq f^* - \langle \mu^*, \mathbf{g}(\hat{x}_k) \rangle \geq f^* - \langle \mu^*, \mathbf{g}(\hat{x}_k)^+ \rangle \geq -||\mu^*|| ||\mathbf{g}(x_k)^+|| \ ,$$

dove la penultima disuguaglianza è banale da dimostrare, e l'ultima è dovuta alla disuguaglianza di Cauchy-Schwarz.

Per concludere, allora, basta notare che $\tilde{q} \leq q^*$, e applicare la formula (3.1), ricordando che $||\mu^*|| \leq ||\mu^*||_1$. □

In alcune applicazioni pratiche, può essere utile sostituire l'insieme D con il seguente insieme

$$D_\infty \stackrel{\mathbf{def}}{=} \left\{ \mu \in M \;\middle|\; ||\mu||_\infty \leq \frac{1}{\gamma} \left[f(\tilde{x}) - q' \right] + r \right\} \ ,$$

In effetti, la proposizione (4.1.1) continua a valere, e si può dimostrare esattamente nel modo già visto, sfruttando il fatto che per ogni $v \in \mathbb{R}^n$ vale $||v||_\infty \leq ||v||$.

4.2 Scelte ottimali per l'insieme D

Cominciamo osservando che il secondo membro della prima delle tre proprietà della proposizione (4.1.1) può essere interpretato come una funzione del parametro r (parametro che compare nella definizione dell'insieme D). Possiamo allora porci il problema di minimizzare tale funzione al variare di $r > 0$, per ogni $k \geq 1$. Fissato $k \geq 1$, non è difficile dimostrare che il punto di minimo è:

$$r_k^* \stackrel{\mathbf{def}}{=} \sqrt{\left(\frac{f(\tilde{x}) - \tilde{q}}{\gamma} \right)^2 + \frac{||x_0 - x^*||^2}{4} + \frac{k\alpha^2 h^2}{4}} \tag{4.3}$$

Possiamo allora pensare di utilizzare una variante del metodo del sottogradiente primale-duale (si veda il sistema (4.1)), usando il seguente algoritmo:

$$\begin{cases} x_{k+1} = \mathbf{P}_X \left[x_k - \alpha L_x(x_k, \mu_k) \right] \\ \mu_{k+1} = \mathbf{P}_{D_k} \left[\mu_k + \alpha L_\mu(x_k, \mu_k) \right] \end{cases} \ , \tag{4.4}$$

con:

$$D_k \stackrel{\mathbf{def}}{=} \left\{ \mu \in M \;\middle|\; ||\mu|| \leq \frac{1}{\gamma} \left[f(\tilde{x}) - q' \right] + r_k^* \right\} \ ,$$

per ogni $k \geq 1$ (r_k^* è dato dalla formula (4.3)). Chiaramente, vale per ogni $k \geq 1$ $M^* \subseteq D_k$.

Con questa variante dell'algoritmo, la stima sulla violazione del vincolo migliora (in particolare tende a 0 per $k \to +\infty$). In particolare, usando la prima proprietà della proposizione (4.1.1) con $r = r_k^*$, e sostituendo a tale parametro la sua espressione esplicita, si ottiene:

$$||\mathbf{g}(\hat{x}_k)^+|| \leq \frac{8}{k\alpha}\left(\frac{f(\tilde{x}) - \tilde{q}}{\gamma}\right) + \frac{2||x_0 - x^*||}{k\alpha} + \frac{2h}{\sqrt{k}}$$

Capitolo 5

Un esempio numerico

Concludiamo questo seminario con un esempio esplicito, per evidenziare anche alcuni limiti del metodo presentato.

Siano $n, m \in \mathbb{N}^+$, e siano $c, a_1, \dots a_m \in \mathbb{R}^n$. Siano poi $b_1, \dots, b_m \in \mathbb{R}$. Vogliamo minimizzare la funzione:

$$f(x) \stackrel{\textbf{def}}{=} \langle c, x \rangle$$

al variare dei vettori $x \in \mathbb{R}^n$ tali che per ogni $j = 1, \dots, m$:

$$g_j(x) \stackrel{\textbf{def}}{=} \langle a_j, x \rangle \leq b_j$$

Sia $A \in M(m, n, \mathbb{R})$ tale che per ogni $j = 1, \dots, m$ la riga j-esima di A sia il vettore a_j^t. Allora la lagrangiana del problema è:

$$l(x, \mu) \stackrel{\textbf{def}}{=} \langle c, x \rangle + \langle \mu, Ax - b \rangle$$

È più comodo, però, lavorare con la cosiddetta **lagrangiana aumentata**:

$$L(x, \mu) \stackrel{\textbf{def}}{=} \langle c, x \rangle + \langle \mu, Ax - b \rangle + \frac{\rho}{2} ||(Ax - b)^+||_2^2 \ ,$$

relativa allo stesso problema di minimizzazione, con:

$$F(x) \stackrel{\textbf{def}}{=} \langle c, x \rangle + \frac{\rho}{2} ||(Ax - b)^+||_2^2 \ ,$$

la quale ha la stessa soluzione ottima, una volta posti i vincoli di prima (perchè il termine che è stato aggiunto risulta nullo). Nella formula, $\rho > 0$ è arbitrario.

Il sistema dell'algoritmo del sottogradiente primale-duale diventa allora:

$$\begin{cases} x_{k+1} = x_k - \alpha_k \left[c + A^t M^{(k)} \left(\mu_k + \rho(Ax_k - b)^+ \right) \right] \\ \mu_{k+1} = \mu_k + \alpha_k (Ax_k - b)^+ \end{cases} \ ,$$

dove per ogni $k \geq 0$ la matrice $M^{(k)} \in M(m, \mathbb{R})$ è diagonale, e per ogni $j = 1, \ldots, m$:

$$M_{jj}^{(k)} = \begin{cases} 1, & \langle a_j, x_k \rangle > b_j \\ 0, & \langle a_j, x_k \rangle \leq b_j \end{cases}$$

La scelta del passo variabile è dovuta al fatto che, in questo caso, si cerca di far convergere la soluzione generata al punto di sella, per $k \to +\infty$. In ogni caso, implementando al calcolatore tale metodo (con $\alpha_k = \frac{1}{k}$ per ogni $k \geq 1$: la scelta di una successione non sommabile ma a quadrato sommabile garantisce la convergenza), si scopre che (anche usando passi variabili) la violazione del vincolo si assesta intorno a un valore circa uguale a 10^{-3}, mentre il valore $f(x_k) - f^*$ cala molto lentamente, e anche dopo 2500 iterazioni non è comunque minore di 10^{-2}. Concludiamo quindi che, anche usando passi variabili e non costanti, i metodi del sottogradiente esposti in questo seminario sono, in generale, intrinsecamente lenti.

Negli anni sono stati sviluppati vari metodi, per sfruttare la struttura del problema e aumentare la velocità di convergenza del metodo: per un'analisi dettagliata di tali metodi, si veda [3].

Bibliografia

[1] Bertsekas, D.P. *Nonlinear Programming.* Athena Scientific, 1999.

[2] Bertsekas, D.P. - Nedíc, A. - Ozdaglar, A. *Convex Analysis and Optimization.* Athena Scientific, 2003.

[3] Boyd, S. *Subgradient Methods.* Notes for Stanford University, 2014.

[4] Kallio, M. - Ruszczyński, A. Approximate primal solutions and rate analysis for dual subgradient methods. *Working Paper WP-94-15, IIASA*, 1994.

[5] Nedíc, A. - Ozdaglar, A. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM J. Optim.*, 19:1757–1780, 2009.

[6] Nedíc, A. - Ozdaglar, A. Subgradient Methods for Saddle-Point Problems. *Journal of Optimization Theory and Applications*, 142(1):205–228, 2009.