

Ottimizzazione dei Sistemi Complessi

G. Liuzzi¹

Giovedì 23 Aprile 2020

¹Istituto di Analisi dei Sistemi ed Informatica IASI - CNR

Un bel passo indietro ...

Metodo del gradiente (rivisto)

Consideriamo il problema *NON vincolato*

$$\begin{aligned} \min & f(x) \\ \text{s.t.} & x \in \mathbb{R}^n \end{aligned}$$

Disponiamo di una stima x_k della soluzione, quindi scriviamo l'app. di Taylor (troncata al I ordine) di $f(x)$ in x_k

$$\bar{f}(x; x_k) = f(x_k) + \nabla f(x_k)^\top (x - x_k) = \nabla f(x_k)^\top d + f(x_k)$$

Vogliamo determinare una nuova stima x_{k+1} minimizzando $\bar{f}(x; x_k)$

Un bel passo indietro ...

Metodo del gradiente (rivisto)

Consideriamo il problema *NON vincolato*

$$\begin{aligned} \min f(x) \\ \text{s.t. } x \in \mathbb{R}^n \end{aligned}$$

Disponiamo di una stima x_k della soluzione, quindi scriviamo l'app. di Taylor (troncata al I ordine) di $f(x)$ in x_k

$$\bar{f}(x; x_k) = f(x_k) + \nabla f(x_k)^\top (x - x_k) = \nabla f(x_k)^\top d + f(x_k)$$

Vogliamo determinare una nuova stima x_{k+1} minimizzando $\bar{f}(x; x_k)$

Un bel passo indietro ...

Metodo del gradiente (rivisto)

Consideriamo il problema *NON vincolato*

$$\begin{aligned} \min f(x) \\ \text{s.t. } x \in \mathbb{R}^n \end{aligned}$$

Disponiamo di una stima x_k della soluzione, quindi scriviamo l'app. di Taylor (troncata al I ordine) di $f(x)$ in x_k

$$\bar{f}(x; x_k) = f(x_k) + \nabla f(x_k)^\top (x - x_k) = \nabla f(x_k)^\top d + f(x_k)$$

Vogliamo determinare una nuova stima x_{k+1} minimizzando $\bar{f}(x; x_k)$

Metodo del Gradiente

Purtroppo, il problema

$$\begin{aligned} \min \quad & \bar{f}(x; x_k) \\ \text{s.t.} \quad & x \in \mathbb{R}^n \end{aligned}$$

se $\nabla f(x_k) \neq 0$, non ammette soluzione!!

Metodo del Gradiente

Sia $q_k(d) = f(x_k) + \nabla f(x_k)^\top d + \|d\|^2/2$ e consideriamo il problema

$$\begin{aligned} \min \quad & q_k(d) \\ \text{s.t.} \quad & x \in \mathbb{R}^n \end{aligned}$$

La direzione $d^* = -\nabla f(x_k)$ è tale che

- $\nabla q_k(d^*) = 0$
- $\nabla^2 q_k(d^*) = \mathbb{I} \succ 0$

Metodo del Gradiente

Sia $q_k(d) = f(x_k) + \nabla f(x_k)^\top d + \|d\|^2/2$ e consideriamo il problema

$$\begin{aligned} \min \quad & q_k(d) \\ \text{s.t.} \quad & x \in \mathbb{R}^n \end{aligned}$$

La direzione $d^* = -\nabla f(x_k)$ è tale che

- $\nabla q_k(d^*) = 0$
- $\nabla^2 q_k(d^*) = \mathbb{I} \succ 0$

Metodo del Gradiente

Quindi, si definisce l'iterazione

$$x_{k+1} = x_k + d^* = x_k - \alpha_k \nabla f(x_k)$$

Un (altro) bel passo indietro ... o avanti!

Consideriamo il problema *NON vincolato*

$$\begin{array}{ll} \min & f(x) \\ \text{c.v.} & x \in \mathbb{R}^n \end{array}$$

Supponiamo di **non** conoscere il punto di minimo locale x^* ma di disporre di una **stima** x_k di x^*

Idea

- definire una approssimazione $q(x; x_k)$ di f vicino ad x_k
- minimizzare $q(x; x_k)$
- aggiornare la stima di x^* definendo x_{k+1}

Un (altro) bel passo indietro ... o avanti!

Consideriamo il problema *NON vincolato*

$$\begin{array}{ll} \min & f(x) \\ \text{c.v.} & x \in \mathbb{R}^n \end{array}$$

Supponiamo di **non** conoscere il punto di minimo locale x^* ma di disporre di una **stima** x_k di x^*

Idea

- definire una approssimazione $q(x; x_k)$ di f vicino ad x_k
- minimizzare $q(x; x_k)$
- aggiornare la stima di x^* definendo x_{k+1}

Il metodo di Newton (1)

Naturalmente, risulta:

$$q(x; x_k) = f(x_k) + \nabla f(x_k)^\top (x - x_k) + \frac{1}{2}(x - x_k)^\top \nabla^2 f(x_k)(x - x_k)$$

$$q_k(d) = f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2}d^\top \nabla^2 f(x_k)d$$

$$\nabla q_k(d) = \nabla f(x_k) + \nabla^2 f(x_k)d$$

$\nabla \bar{q}_k(d) = 0$ quando $\nabla f(x_k) + \nabla^2 f(x_k)d = 0$.

Se $\nabla^2 f(x_k)$ è non singolare, allora possiamo definire:

$$\begin{aligned} d_k &= -\nabla^2 f(x_k)^{-1} \nabla f(x_k) && \text{direzione di Newton} \\ x_{k+1} &= x_k + d_k && \text{iterazione di Newton} \end{aligned}$$

Il metodo di Newton (1)

Naturalmente, risulta:

$$q(x; x_k) = f(x_k) + \nabla f(x_k)^\top (x - x_k) + \frac{1}{2}(x - x_k)^\top \nabla^2 f(x_k)(x - x_k)$$

$$q_k(d) = f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2}d^\top \nabla^2 f(x_k)d$$

$$\nabla q_k(d) = \nabla f(x_k) + \nabla^2 f(x_k)d$$

$\nabla \bar{q}_k(d) = 0$ quando $\nabla f(x_k) + \nabla^2 f(x_k)d = 0$.

Se $\nabla^2 f(x_k)$ è non singolare, allora possiamo definire:

$$\begin{aligned} d_k &= -\nabla^2 f(x_k)^{-1} \nabla f(x_k) && \text{direzione di Newton} \\ x_{k+1} &= x_k + d_k && \text{iterazione di Newton} \end{aligned}$$

Introduzione

Consideriamo il problema

$$\begin{array}{ll} \min_x & f(x) \\ \text{c.v.} & h(x) = 0 \end{array}$$

Le condizioni di KKT per il problema danno luogo al sistema nonlineare

$$F(x, \mu) = \begin{bmatrix} \nabla f(x) + \nabla h(x)\mu \\ h(x) \end{bmatrix} = 0$$

Idea: Risolvere il sistema $F(x, \mu)$ usando il **metodo di Newton**

Introduzione

Consideriamo il problema

$$\begin{array}{ll} \min_x & f(x) \\ \text{c.v.} & h(x) = 0 \end{array}$$

Le condizioni di KKT per il problema danno luogo al sistema nonlineare

$$F(x, \mu) = \begin{bmatrix} \nabla f(x) + \nabla h(x)\mu \\ h(x) \end{bmatrix} = 0$$

Idea: Risolvere il sistema $F(x, \mu)$ usando il **metodo di Newton**

Metodo di Newton-Lagrange

Se vogliamo usare il metodo di Newton, la prima cosa da fare è calcolare lo Jacobiano di F

$$\begin{bmatrix} \nabla_{xx}^2 L(x, \mu) & \nabla h(x) \\ \nabla h(x)^\top & 0 \end{bmatrix}$$

Il metodo di Newton è definito dalla iterazione

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix} + \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix}$$

dove $(\delta_k^x, \delta_k^\mu)$ tali che

$$\begin{bmatrix} \nabla_{xx}^2 L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{bmatrix} \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix} = -F(x_k, \mu_k)$$

Metodo di Newton-Lagrange

Se vogliamo usare il metodo di Newton, la prima cosa da fare è calcolare lo Jacobiano di F

$$\begin{bmatrix} \nabla_{xx}^2 L(x, \mu) & \nabla h(x) \\ \nabla h(x)^\top & 0 \end{bmatrix}$$

Il metodo di Newton è definito dalla iterazione

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix} + \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix}$$

dove $(\delta_k^x, \delta_k^\mu)$ tali che

$$\begin{bmatrix} \nabla_{xx}^2 L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{bmatrix} \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix} = -F(x_k, \mu_k)$$

Metodo di Newton-Lagrange

Se vogliamo usare il metodo di Newton, la prima cosa da fare è calcolare lo Jacobiano di F

$$\begin{bmatrix} \nabla_{xx}^2 L(x, \mu) & \nabla h(x) \\ \nabla h(x)^\top & 0 \end{bmatrix}$$

Il metodo di Newton è definito dalla iterazione

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix} + \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix}$$

dove $(\delta_k^x, \delta_k^\mu)$ tali che

$$\begin{bmatrix} \nabla_{xx}^2 L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{bmatrix} \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix} = -F(x_k, \mu_k)$$

Metodo di Newton-Lagrange

Il passo k è ben definito quando la matrice

$$\begin{bmatrix} \nabla_{xx}^2 L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{bmatrix}$$

è **non singolare**. Questo accade sotto la seguente ipotesi:

Assunzione

- $rg(\nabla h(x_k)) = p$, i.e. i gradienti dei vincoli in x_k sono lin. indipendenti
- $d^\top \nabla_{xx}^2 L(x_k, \mu_k) d > 0$ per ogni $d \neq 0$ e tale che $\nabla h(x_k)^\top d = 0$

Metodo di Newton-Lagrange

Il passo k è ben definito quando la matrice

$$\begin{bmatrix} \nabla_{xx}^2 L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{bmatrix}$$

è **non singolare**. Questo accade sotto la seguente ipotesi:

Assunzione

- $rg(\nabla h(x_k)) = p$, i.e. i gradienti dei vincoli in x_k sono lin. indipendenti
- $d^\top \nabla_{xx}^2 L(x_k, \mu_k) d > 0$ per ogni $d \neq 0$ e tale che $\nabla h(x_k)^\top d = 0$

Metodo SQP

È possibile vedere il metodo di Newton-Lagrange in maniera alternativa. A tal fine, definiamo il problema

$$\begin{aligned} \min_d \quad & \frac{1}{2} d^\top \nabla_{xx}^2 L(x_k, \mu_k) d + \nabla f(x_k)^\top d \\ \text{c.v.} \quad & \nabla h(x_k)^\top d + h(x_k) = 0 \end{aligned}$$

Se vale l'assunzione di prima, questo problema ha una unica soluzione $(\bar{d}, \bar{\mu})$ che (ovviamente) soddisfa

$$\begin{aligned} \nabla_{xx}^2 L(x_k, \mu_k) \bar{d} + \nabla f(x_k) + \nabla h(x_k) \bar{\mu} &= 0 \\ \nabla h(x_k)^\top \bar{d} + h(x_k) &= 0 \end{aligned}$$

Metodo SQP

È possibile vedere il metodo di Newton-Lagrange in maniera alternativa. A tal fine, definiamo il problema

$$\begin{aligned} \min_d \quad & \frac{1}{2} d^\top \nabla_{xx}^2 L(x_k, \mu_k) d + \nabla f(x_k)^\top d \\ \text{c.v.} \quad & \nabla h(x_k)^\top d + h(x_k) = 0 \end{aligned}$$

Se vale l'assunzione di prima, questo problema ha una unica soluzione $(\bar{d}, \bar{\mu})$ che (ovviamente) soddisfa

$$\begin{aligned} \nabla_{xx}^2 L(x_k, \mu_k) \bar{d} + \nabla f(x_k) + \nabla h(x_k) \bar{\mu} &= 0 \\ \nabla h(x_k)^\top \bar{d} + h(x_k) &= 0 \end{aligned}$$

Metodo SQP

È possibile vedere il metodo di Newton-Lagrange in maniera alternativa. A tal fine, definiamo il problema

$$\begin{aligned} \min_d \quad & \frac{1}{2} d^\top \nabla_{xx}^2 L(x_k, \mu_k) d + \nabla f(x_k)^\top d \\ \text{c.v.} \quad & \nabla h(x_k)^\top d + h(x_k) = 0 \end{aligned}$$

Se vale l'assunzione di prima, questo problema ha una unica soluzione $(\bar{d}, \bar{\mu})$ che (ovviamente) soddisfa

$$\begin{aligned} \nabla_{xx}^2 L(x_k, \mu_k) \bar{d} + \nabla f(x_k) + \nabla h(x_k) \bar{\mu} &= 0 \\ \nabla h(x_k)^\top \bar{d} + h(x_k) &= 0 \end{aligned}$$

Metodo SQP

$$\begin{aligned}\nabla_{xx}^2 L(x_k, \mu_k) \bar{d} + \nabla f(x_k) + \nabla h(x_k) \bar{\mu} &= 0 \\ \nabla h(x_k)^\top \bar{d} + h(x_k) &= 0\end{aligned}$$

lo possiamo scrivere come

$$\begin{pmatrix} \nabla_{xx}^2 L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{pmatrix} \begin{pmatrix} \bar{d} \\ \bar{\mu} \end{pmatrix} = - \begin{pmatrix} \nabla f(x_k) \\ h(x_k) \end{pmatrix}$$

e, sottraendo $\nabla h(x_k) \mu_k$ dalla prima eq.

$$\begin{pmatrix} \nabla_{xx}^2 L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{pmatrix} \begin{pmatrix} \bar{d} \\ \bar{\mu} - \mu_k \end{pmatrix} = - \begin{pmatrix} \nabla f(x_k) + \nabla h(x_k) \mu_k \\ h(x_k) \end{pmatrix}$$

Metodo SQP

Ponendo $\bar{d}_\mu = \bar{\mu} - \mu_k$ otteniamo proprio l'iterazione del metodo di Newton-Lagrange

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} \bar{d} \\ \bar{d}_\mu \end{pmatrix} + \begin{pmatrix} x_k \\ \mu_k \end{pmatrix}$$

dove risulta $\mu_{k+1} = \bar{\mu}$

Metodo di soluzione

Algoritmo SQP

Dati: (x_0, μ_0) , maxit

for $k = 0, 1, \dots, \text{maxit}$

Calcola \bar{d}_k e $\bar{\mu}_k$

Poni $x_{k+1} = x_k + \bar{d}_k$ e $\mu_{k+1} = \bar{\mu}_k$

if (x_{k+1}, μ_{k+1}) è KKT then

$x^* \leftarrow x_{k+1}$, $\mu^* \leftarrow \mu_{k+1}$ e STOP

endif

endfor

Return: miglior coppia trovata (x^*, μ^*)

Metodo di soluzione

Algoritmo SQP

Dati: (x_0, μ_0) , maxit

for $k = 0, 1, \dots, \text{maxit}$

Calcola \bar{d}_k e $\bar{\mu}_k$

Poni $x_{k+1} = x_k + \bar{d}_k$ e $\mu_{k+1} = \bar{\mu}_k$

if (x_{k+1}, μ_{k+1}) è KKT **then**

$x^* \leftarrow x_{k+1}$, $\mu^* \leftarrow \mu_{k+1}$ e STOP

endif

endfor

Return: miglior coppia trovata (x^*, μ^*)

Metodo di soluzione

Algoritmo SQP

Dati: (x_0, μ_0) , maxit

for $k = 0, 1, \dots, \text{maxit}$

 Calcola \bar{d}_k e $\bar{\mu}_k$

 Poni $x_{k+1} = x_k + \bar{d}_k$ e $\mu_{k+1} = \bar{\mu}_k$

if (x_{k+1}, μ_{k+1}) è KKT **then**

$x^* \leftarrow x_{k+1}$, $\mu^* \leftarrow \mu_{k+1}$ e STOP

endif

endfor

Return: miglior coppia trovata (x^*, μ^*)

Proprietà di convergenza di SQP

Proposizione

Sia (x^, μ^*) soluzione del problema tale che siano soddisfatte*

- *LICQ^a*
- *SOSC^b*

Allora, se (x_0, μ_0) è suff. vicino a (x^, μ^*) , l'algoritmo SQP è ben definito e converge alla soluzione del problema*

^aLinear Independence Constraint Qualification

^bSecond Order Sufficient Condition

Vincoli di disuguaglianza

Come si gestiscono in questo contesto vincoli di disuguaglianza
 $g(x) \leq 0$?

- 1) Aggiunta di variabili slack: $g_i(x) + s_i = 0$, $s_i \geq 0$ e gestione “esplicita” dei vincoli di bound sulla variabili (SNOPT, KNITRO)
- 2) Approccio SiQP - SQP with inequalities
- 3) Approccio SeQP - SQP with equalities (SNOPT)

Vincoli di disuguaglianza

$$\begin{array}{ll} \min_x & f(x) \\ \text{c.v.} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

Allora, data una stima corrente (x_k, μ_k, λ_k) della soluzione, definisci il problema

$$\begin{array}{ll} \min_d & \nabla f(x_k)^\top d + \frac{1}{2} d^\top \nabla_{xx}^2 L(x_k, \mu_k, \lambda_k) d \\ \text{c.v.} & \nabla h(x_k)^\top d + h(x_k) = 0 \\ & \nabla g(x_k)^\top d + g(x_k) \leq 0 \end{array}$$

N.B. ogni sottoproblema ha esattamente lo stesso numero di vincoli del problema originario

Vincoli di disuguaglianza

$$\begin{array}{ll} \min_x & f(x) \\ \text{c.v.} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

Allora, data una stima corrente (x_k, μ_k, λ_k) della soluzione, definisci il problema

$$\begin{array}{ll} \min_d & \nabla f(x_k)^\top d + \frac{1}{2} d^\top \nabla_{xx}^2 L(x_k, \mu_k, \lambda_k) d \\ \text{c.v.} & \nabla h(x_k)^\top d + h(x_k) = 0 \\ & \nabla g(x_k)^\top d + g(x_k) \leq 0 \end{array}$$

N.B. ogni sottoproblema ha esattamente lo stesso numero di vincoli del problema originario

Vincoli di disuguaglianza

$$\begin{array}{ll} \min_x & f(x) \\ \text{c.v} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

Immaginiamo di poter conoscere quali vincoli di disuguaglianza sono attivi in una soluzione del problema x^* .

Indichiamo I^* l'insieme degli indici dei vincoli attivi in x^*

$$I^* = \{i : g_i(x^*) = 0\}$$

Quindi, $g_i(x^*) < 0$, per ogni $i \notin I^*$

Vincoli di disuguaglianza

$$\begin{array}{ll} \min_x & f(x) \\ \text{c.v} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

Immaginiamo di poter conoscere quali vincoli di disuguaglianza sono attivi in una soluzione del problema x^* .

Indichiamo I^* l'insieme degli indici dei vincoli attivi in x^*

$$I^* = \{i : g_i(x^*) = 0\}$$

Quindi, $g_i(x^*) < 0$, per ogni $i \notin I^*$

Vincoli di disuguaglianza

Sotto questa ipotesi, anche il problema

$$\begin{aligned} \min_x \quad & f(x) \\ \text{c.v.} \quad & h(x) = 0 \\ & g_i(x) = 0, \quad i \in I^* \end{aligned}$$

ammette x^* come soluzione locale.

N.B. questo problema ha (nuovamente) tutti e soli vincoli di uguaglianza. Pertanto, posso usare il metodo di soluzione che abbiamo visto per problemi con vincoli di uguaglianza

Vincoli di disuguaglianza

Sotto questa ipotesi, anche il problema

$$\begin{aligned} \min_x \quad & f(x) \\ \text{c.v.} \quad & h(x) = 0 \\ & g_i(x) = 0, \quad i \in I^* \end{aligned}$$

ammette x^* come soluzione locale.

N.B. questo problema ha (nuovamente) tutti e soli vincoli di uguaglianza. Pertanto, posso usare il metodo di soluzione che abbiamo visto per problemi con vincoli di uguaglianza

Vincoli di disuguaglianza

Cosa manca? Normalmente non sappiamo come è fatto I^* . Quindi:

- si dà una stima I_k di I^*
- si risolve il sotto problema
- si aggiorna la stima di I^* definendo I_{k+1}

Per questo motivo questa classe di metodi è anche nota come “active-set” SQP

Vincoli di disuguaglianza

Cosa manca? Normalmente non sappiamo come è fatto I^* . Quindi:

- si dà una stima I_k di I^*
- si risolve il sotto problema
- si aggiorna la stima di I^* definendo I_{k+1}

Per questo motivo questa classe di metodi è anche nota come “active-set” SQP

Vincoli di disuguaglianza

Cosa manca? Normalmente non sappiamo come è fatto I^* . Quindi:

- si dà una stima I_k di I^*
- si risolve il sotto problema
- si aggiorna la stima di I^* definendo I_{k+1}

Per questo motivo questa classe di metodi è anche nota come “active-set” SQP

Problema approssimante

Data una stima dei vincoli attivi I_k ed una stima della soluzione (x_k, μ_k, λ_k) , il problema “approssimante” è

$$\begin{aligned} \min_d \quad & \nabla f(x_k)^\top d + \frac{1}{2} d^\top \nabla_{xx}^2 L(x_k, \mu_k, \lambda_k) d \\ \text{c.v.} \quad & \nabla h(x_k)^\top d + h(x_k) = 0 \\ & \nabla g_{I_k}(x_k)^\top d + g_{I_k}(x_k) = 0 \end{aligned}$$

Soluzione del problema approssimante

In base a ragionamenti del tutto analoghi a quelli appena visti, bisogna risolvere il sistema lineare:

$$\begin{bmatrix} \nabla_{xx}^2 L_k & \nabla h(x_k) & \nabla g_{l_k}(x_k) \\ \nabla h(x_k)^\top & 0 & 0 \\ \nabla g_{l_k}(x_k)^\top & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{d} \\ \bar{\mu} \\ \bar{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) \\ h(x_k) \\ g_{l_k}(x_k) \end{bmatrix}$$

Metodo di soluzione (vincoli di uguaglianza)

Algoritmo SQP

Dati: (x_0, μ_0, λ_0) , maxit , $\gamma > 0$

for $k = 0, 1, \dots, \text{maxit}$

$$I_k = \{i : g_i(x_k) \geq -\gamma(\lambda_k)_i\}$$

Calcola $\bar{d}_k, \bar{\mu}_k, \bar{\lambda}_k$

Poni $x_{k+1} = x_k + \bar{d}_k, \mu_{k+1} = \bar{\mu}_k,$

$(\lambda_{k+1})_{I_k} = \bar{\lambda}_k, (\lambda_{k+1})_i = 0, i \notin I_k$

if $(x_{k+1}, \mu_{k+1}, \lambda_{k+1})$ è KKT **then**

$x^* \leftarrow x_{k+1}, \mu^* \leftarrow \mu_{k+1}, \lambda^* \leftarrow \lambda_{k+1}$ e **STOP**

endif

endfor

Return: miglior punto trovato (x^*, μ^*, λ^*)

Metodo di soluzione (vincoli di uguaglianza)

Algoritmo SQP

Dati: (x_0, μ_0, λ_0) , maxit , $\gamma > 0$

for $k = 0, 1, \dots, \text{maxit}$

$$I_k = \{i : g_i(x_k) \geq -\gamma(\lambda_k)_i\}$$

Calcola $\bar{d}_k, \bar{\mu}_k, \bar{\lambda}_k$

Poni $x_{k+1} = x_k + \bar{d}_k, \mu_{k+1} = \bar{\mu}_k,$
 $(\lambda_{k+1})_{I_k} = \bar{\lambda}_k, (\lambda_{k+1})_i = 0, i \notin I_k$

if $(x_{k+1}, \mu_{k+1}, \lambda_{k+1})$ è KKT **then**

$x^* \leftarrow x_{k+1}, \mu^* \leftarrow \mu_{k+1}, \lambda^* \leftarrow \lambda_{k+1}$ e STOP

endif

endfor

Return: miglior punto trovato (x^*, μ^*, λ^*)

Metodo di soluzione (vincoli di uguaglianza)

Algoritmo SQP

Dati: (x_0, μ_0, λ_0) , maxit , $\gamma > 0$

for $k = 0, 1, \dots, \text{maxit}$

$$I_k = \{i : g_i(x_k) \geq -\gamma(\lambda_k)_i\}$$

Calcola $\bar{d}_k, \bar{\mu}_k, \bar{\lambda}_k$

Poni $x_{k+1} = x_k + \bar{d}_k$, $\mu_{k+1} = \bar{\mu}_k$,
 $(\lambda_{k+1})_{I_k} = \bar{\lambda}_k$, $(\lambda_{k+1})_i = 0$, $i \notin I_k$

if $(x_{k+1}, \mu_{k+1}, \lambda_{k+1})$ è KKT **then**

$x^* \leftarrow x_{k+1}$, $\mu^* \leftarrow \mu_{k+1}$, $\lambda^* \leftarrow \lambda_{k+1}$ e STOP

endif

endfor

Return: miglior punto trovato (x^*, μ^*, λ^*)

Esercizio – SeQP

Consideriamo il problema:

$$\begin{aligned} \min & (x - 1)^2 + y^2 - y \\ \text{c.v.} & x^2 + y^2 - 1 \leq 0 & (\lambda_1) \\ & -x \leq 0 & (\lambda_2) \end{aligned}$$

Il problema è convesso e ha $(x^*, y^*)^\top = (2/\sqrt{5}, 1/\sqrt{5})^\top$ come soluzione globale che soddisfa le condizioni di KKT con moltiplicatori $\lambda_1^* = \sqrt{5}/2 - 1$ e $\lambda_2^* = 0$.

Notiamo che all'ottimo globale è attivo il solo primo vincolo cioè risulta $I_0^* = \{1\}$.

Esercizio – SeQP

Supponiamo ora di voler risolvere il problema usando il metodo SeQP partendo dal punto iniziale $(x_0, y_0)^T = (0, 0)^T$ con stima iniziale dei moltiplicatori di KKT $\lambda_1^0 = 0$, $\lambda_2^0 = 0$.

Usiamo $\gamma = 1$ per la stima dei vincoli attivi

Iterazione 0

$$x_0 = 0, y_0 = 0, \lambda_1^0 = 0, \lambda_2^0 = 0, KKT_{error} = 2$$

$$g_1^0 = g_1(x_0, y_0) = -1 \not\geq -\gamma\lambda_1^0 = 0$$

$$g_2^0 = g_2(x_0, y_0) = 0 \geq -\gamma\lambda_2^0 = 0$$

$$\text{quindi } I_0^0 = \{2\}$$

Il problema QP che dobbiamo risolvere è:

$$\begin{aligned} \min \quad & d_1^2 + d_2^2 - 2d_1 - d_2 \\ \text{c.v.} \quad & -d_1 = 0 \end{aligned} \quad (\mu_2)$$

Le condizioni di KKT per il problema QP forniscono la soluzione:

$$d_1 = 0, d_2 = 1/2, \mu_2 = -2$$

La nuova stima di KKT è pertanto

$$x_1 = 0, y_1 = 1/2, \lambda_1^1 = 0, \lambda_2^1 = -2$$

Iterazione 1

$$x_1 = 0, y_1 = 1/2, \lambda_1^1 = 0, \lambda_2^1 = -2, KKT_{error} = 2$$

$$g_1^1 = g_1(x_1, y_1) = -3/4 \not\geq -\gamma\lambda_1^1 = 0$$

$$g_2^1 = g_2(x_1, y_1) = 0 \not\geq -\gamma\lambda_2^1 = 2$$

quindi $I_0^1 = \emptyset$

Il problema QP che dobbiamo risolvere è:

$$\min d_1^2 + d_2^2 - 2d_1$$

Le condizioni di KKT per il problema QP forniscono la soluzione:

$$d_1 = 1, d_2 = 0$$

La nuova stima di KKT è pertanto

$$x_2 = 1, y_2 = 1/2, \lambda_1^2 = 0, \lambda_2^2 = 0$$

Iterazione 2

$$x_2 = 1, y_2 = 1/2, \lambda_1^2 = 0, \lambda_2^2 = 0, KKT_{error} = 0.25$$

$$g_1^2 = g_1(x_2, y_2) = 1/4 \geq -\gamma \lambda_1^2 = 0$$

$$g_2^2 = g_2(x_2, y_2) = -1 \not\geq -\gamma \lambda_2^2 = 0$$

$$\text{quindi } I_0^2 = \{1\}$$

Il problema QP che dobbiamo risolvere è:

$$\begin{aligned} \min \quad & d_1^2 + d_2^2 \\ \text{c.v.} \quad & 2d_1 + d_2 = -1/4 \quad (\mu_1) \end{aligned}$$

Le condizioni di KKT per il problema QP forniscono la soluzione:

$$d_1 = -1/10, d_2 = -1/20, \mu_1 = 1/10$$

La nuova stima di KKT è pertanto

$$x_3 = 9/10, y_3 = 9/20, \lambda_1^3 = 1/10, \lambda_2^3 = 0$$

Iterazione 3

$$x_3 = 0.9, y_3 = 0.45, \lambda_1^3 = 0.1, \lambda_2^3 = 0, KKT_{error} = 0.2$$

...

Ricapitolando e proseguendo

k	x	y	λ_1	λ_2	KKT_{error}
0	0	0	0	0	2.0
1	0	0.5	0	-2	2.0
2	1	0.5	0	0	0.25
3	0.9	0.45	0.1	0	0.2
4	0.894	0.447	0.118	0	0.0002
5	0.89443	0.44721	0.118034	0	4.5e-09

N.B. la soluzione ottima del problema è

$(x^*, y^*)^\top = (2/\sqrt{5}, 1/\sqrt{5})^\top \simeq (0.89443, 0.44721)$ con moltiplicatori
 $\lambda_1^* = \sqrt{5}/2 - 1 \simeq 0.118034$ e $\lambda_2^* = 0$!!