

UNIVERSITÀ DEGLI STUDI DI PISA
FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI
Corso di Laurea Specialistica in Informatica

Tesi di laurea

Metodi bundle con doppia stabilizzazione

Relatore
prof. **Antonio Frangioni**

Studente
Tamara Knežević

Anno Accademico
2008/2009

Prefazione

Esistono diversi algoritmi per la minimizzazione delle funzioni convesse nondifferenziabili; una delle classi più studiate è quella degli algoritmi di tipo *bundle*. Questa classe contiene molti algoritmi piuttosto diversi l'uno dall'altro, ma con una caratteristica comune: in un modo o nell'altro, sono tutti varianti dell'*algoritmo dei piani di taglio* [2]. Questo è un algoritmo semplice che costruisce un'approssimazione poliedrale della funzione convessa e, successivamente, minimizza quest'approssimazione al posto della funzione. È anche un algoritmo instabile, il cui comportamento numerico spesso non è soddisfacente. Gli algoritmi che noi considereremo implementano modifiche varie dell'algoritmo dei piani di taglio. Queste modifiche sono fatte per stabilizzare l'algoritmo (infatti, si chiamano stabilizzazioni), mantenendone la semplicità ed efficienza.

Benché gli algoritmi di tipo bundle seguano tutti uno stesso schema, e le loro dimostrazioni di convergenza siano spesso molto simili, essi sono stati proposti in momenti diversi da autori diversi, e le loro proprietà teoriche sono state di volta in volta studiate individualmente. In questa tesi cerchiamo di costruire uno schema unificante che permetta di studiare tutti, o quantomeno una buona parte di, questi algoritmi. Alcuni passi iniziali in questa direzione sono stati compiuti in [1], ma molti algoritmi di tipo bundle non sono compresi nello schema proposto in quel lavoro. L'idea di base è che questi algoritmi implementino in modo diverso uno stesso principio, che corrisponde a diversi modi in cui viene "stabilizzato" l'algoritmo dei piani di taglio. In particolare, scopo della tesi è stato quello di capire:

- quali sono le proprietà che caratterizzano ognuna di queste stabilizzazioni;
- se ci sono le proprietà che tutte le stabilizzazioni possiedono;
- se è possibile, a partire da queste ultime, definire una stabilizzazione unica dell'algoritmo dei piani di taglio, tale che ogni algoritmo sia considerabile come una particolare implementazione della stabilizzazione unica.

Questo testo è organizzato nel modo seguente:

- Il Capitolo 1 descrive il problema della minimizzazione delle funzioni convesse nondifferenziabili mediante metodi bundle, e l'idea per una stabilizzazione generale.

- Il Capitolo 2 descrive cinque diversi algoritmi di tipo bundle per la minimizzazione di funzioni convesse nondifferenziabili.
- Il Capitolo 3 presenta gli algoritmi del Capitolo 2 come diverse stabilizzazioni dell'algoritmo dei piani di taglio.
- Il Capitolo 4 propone una stabilizzazione unica che descrive contemporaneamente tutte le stabilizzazioni del Capitolo 3.
- Il Capitolo 5 sviluppa una teoria di convergenza generale di un algoritmo che usi la stabilizzazione unica sotto alcune condizioni sui termini di stabilizzazione.
- L'Appendice contiene i risultati di Analisi convessa e le notazioni utilizzati nel testo.

Indice

1	Introduzione	7
1.1	Introduzione ai metodi bundle	8
1.2	Il modello	10
2	Presentazione degli algoritmi	13
2.1	Algoritmo dei piani di taglio	13
2.2	Penalty e trust-region bundle	19
2.3	Metodi bundle generalizzati	21
2.4	Level bundle	23
2.5	Metodo dei piani di taglio con i centri di Chebychev	24
2.6	Dual bundle	27
2.7	Separating plane algorithms (SPA)	29
3	Metodi bundle come le stabilizzazioni dell'algoritmo dei piani di taglio	33
3.1	Algoritmo dei piani del taglio	33
3.2	Metodi bundle generalizzati	34
3.3	Level bundle	35
3.4	Metodo dei piani di taglio con i centri di Chebychev	38
3.5	Dual Bundle	44
3.6	Separating plane algorithms	48
3.7	Tabelle riassuntive delle forme astratte	50
3.8	Tabelle riassuntive delle forme concrete	52
3.9	Tabelle riassuntive delle forme implementabili	54
4	Proprietá dei termini di stabilizzazione	57
4.1	Termine di stabilizzazione D_t	57
4.2	Termine di stabilizzazione della funzione obiettivo	61
4.2.1	Caso "good" della stabilizzazione ν	62
4.2.2	Caso "good" della stabilizzazione ϕ	64
4.3	Stabilizzazione generalizzata della funzione obiettivo	65

4.3.1	Caso “good” della stabilizzazione (Un1)	66
4.3.2	Caso “good” della stabilizzazione (Un2)	68
5	Dimostrazione di convergenza	71
5.1	Introduzione	71
5.2	Formulazione dell’algoritmo	72
5.3	Convergenza	74
A	Definizioni e notazioni	81
A.1	Definizioni	81
A.2	Notazioni	88

Capitolo 1

Introduzione

Il problema che siamo interessati a risolvere è:

$$\min_x \{g(x) : x \in X\}$$

dove $g : X \rightarrow R$ è una funzione convessa e chiusa, non necessariamente differenziabile, e X è un insieme convesso. Possiamo includere il vincolo $x \in X$ nella funzione obiettivo settando

$$f(x) = \begin{cases} g(x) & \text{se } x \in X \\ +\infty & \text{altrimenti.} \end{cases}$$

La funzione $f : R^n \rightarrow R \cup \{\infty\}$ è convessa e chiusa su tutto R^n , e risolvere

$$(II) \quad \min_x \{ f(x) \} \tag{1.1}$$

è equivalente al problema originale. Per semplificare la notazione e gli argomenti assumeremo nel seguito che $X = R^n$, ossia f sia finita ovunque; usualmente, l'introduzione di vincoli ($X \neq R^n$, purché convesso) non comporta sostanziali modifiche alla teoria di convergenza degli algoritmi di tipo bundle.

Nelle applicazioni di nostro interesse, un metodo risolutivo per (II) ha due componenti: un oracolo e un algoritmo.

L'oracolo contiene le informazioni che caratterizzano la particolare funzione f che si desidera minimizzare. Negli algoritmi di tipo bundle si assume un oracolo "minimale" che, per ciascun $x \in R^n$, fornisce il valore $f(x)$ ed un *subgradiente* $g \in \partial f(x)$ (si veda l'Appendice per ulteriori dettagli).

L'algoritmo non conosce l'esplicita definizione di f , e deve determinare la soluzione ottima di (II) facendo solamente affidamento sulle informazioni (iterativamente) fornitegli dall'oracolo. L'idea alla base di tutti gli algoritmi che

vedremo è la costruzione di un “modello” che in qualche modo approssima f e la successiva minimizzazione di questo “modello” al posto di minimizzazione di f . Globalmente, il “modello” guida l’algoritmo verso il punto ottimo per (II) (oppure dimostra che tale punto non esiste). Come vedremo, il “modello” consiste di solito di una funzione che è una qualche approssimazione di f originale e in più può avere anche altre componenti, come le stabilizzazioni. L’algoritmo, è effettivamente indipendente da f , poiché la conoscenza di f è sostituita dalle chiamate all’oracolo che avvengono in punti ben precisi del codice implementato e sono eseguite per ogni oracolo e quindi per ogni f . In altre parole, l’oracolo è una “black box” per l’algoritmo. Ma è vero anche il contrario, ossia l’algoritmo è una “black box” per l’oracolo: per una f e un punto iniziale x_0 dati dall’oracolo, l’algoritmo restituisce il minimo di f se esso esiste.

1.1 Introduzione ai metodi bundle

Un metodo bundle è un metodo iterativo per la soluzione di (II). Si basa sul *bundle*, ossia una collezione di informazioni ottenute dall’oracolo nelle iterazioni precedenti. Ad ogni iterazione, le informazioni contenute nel bundle vengono usate per rendere il “modello” un’approssimazione migliore (rispetto all’iterazione precedente) di f .

Tutti gli algoritmi che consideriamo usano il bundle $\beta = \{g_i \in \partial f(x_i)\}$ per costruire il modello, dove $\{x_i\}$ è la sequenza di punti tentativi (o iterate) nei quali chiediamo all’oracolo di calcolare il subgradiente g_i e di valutare f . Le informazioni nel bundle ci permettono, ad ogni iterazione, di costruire un’approssimazione di f e di minimizzarla trovando così la prossima iterata; opportunamente gestito, questo processo assicura che $\{x_i\} \rightarrow x^*$ quando $i \rightarrow \infty$, dove x^* è una soluzione ottima di (1.1). Naturalmente, x^* potrebbe non esistere per qualche f ; esempi ovvio sono l’identità ($f(x) = x$) e $f(x) = 1/x$ (per $x > 0$). Anche in questo caso vorremo l’algoritmo convergente; per questo definiamo f_* come il valore ottimo di (1.1) (che può, quindi, essere $-\infty$) e vorremo almeno che risulti $\{f(x_i)\} \rightarrow f_*$.

Una proprietà desiderabile per gli algoritmi è quindi che il valore della funzione obiettivo deve decrescere:

$$f(x_{i+1}) < f(x_i) \quad \forall i. \tag{1.2}$$

In altri termini, ad ogni iterazione bisogna trovare il punto x_{i+1} , e questa ricerca consiste nell’individuare una direzione d_i di spostamento dall’iterata corrente x_i ($x_{i+1} = x_i + d_i$). Lo spostamento lungo questa direzione ci deve

portare “più vicino all’ottimo”, cioè la condizione (1.2) deve essere rispettata. Quindi d_i deve essere una direzione di decrescita. Purtroppo, nel caso dell’ottimizzazione nondifferenziabile non è facile ottenere una proprietà di questo tipo avendo a disposizione il solo oracolo. In particolare, non è in generale possibile garantire che la direzione d_i determinata sia di decrescita, a differenza di ciò che accade nel caso dell’ottimizzazione di funzioni differenziabili. Pertanto, nei metodi bundle si distinguono due casi. Se risulta $f(x_{i+1}) < f(x_i)$ effettivamente ci si sposta lungo d_i , e questo viene chiamato un Serious Step. Se invece risulta $f(x_{i+1}) \geq f(x_i)$, allora uno deve aggiungere ulteriori informazioni nel bundle e arricchire il modello per ottenere una miglior direzione all’iterazione successiva (il cosiddetto Null Ntep). Quando viene eseguito un serious step, x_{i+1} diventa la nuova iterata corrente. Nei null step, invece, cambia solo il bundle e l’iterata corrente rimane invariata. In altre parole, un NS consiste nel ricalcolare la direzione di spostamento, avendo a disposizione un bundle modificato e la solita iterata corrente.

Quindi, quando x_{i+1} diventa l’iterata corrente, può rimanere tale per una sola iterazione successiva, se avviene di nuovo un SS, o può rimanere l’iterata corrente per più iterate successive, tante quanti sono i NS che avvengono tra il SS nel quale x_{i+1} è diventato l’iterata corrente ed il SS successivo. Come vedremo, una sequenza di NS consecutivi può essere interpretata come un algoritmo di minimizzazione del modello usando un metodo dei piani di taglio, mentre la sequenza dei SS corrisponde alla “vera” minimizzazione della funzione. I metodi bundle sono quindi inerentemente “a due livelli”, una caratteristica che emergerà chiaramente nelle dimostrazioni di convergenza.

Un’altra caratteristica interessante degli algoritmi bundle è che hanno sia un’interpretazione primale che una duale. Nel primo caso il modello consiste nell’uso di un’approssimazione della funzione obiettivo del problema che viene minimizzata al posto di f per trovare la direzione di decrescita e l’iterata successiva. Un’approssimazione frequentemente usata è la funzione dei piani di taglio. Nel secondo caso invece si considera la funzione duale (o coniugata) f^* di f . Gli argomenti della coniugata sono i subgradienti di f (oggetti duali), che si usano come direzioni di decrescita. L’interpretazione duale del problema (II) come quello di determinare il valore di f^* in 0 (si veda l’Appendice) permette di chiarificare molte parti delle dimostrazioni di convergenza; per questo, di ognuno di metodi del Capitolo 2 scriveremo sia la forma primale che quella duale.

1.2 Il modello

Ogni algoritmo che considereremo nel Capitolo 2 implementa un diverso *modello*. L'idea del modello è la seguente: poiché f non si conosce, invece di cercare direttamente di trovare il minimo di f è necessario utilizzare una funzione nota che “rimpiazzi” f nei calcoli. Per semplicità, possiamo all'inizio immaginare che questa funzione abbia la proprietà di conservare i minimi di f (se ne esistono), ossia che risolvere (II) equivalga a risolvere

$$\min_x \{ \nu(x) \}. \quad (1.3)$$

dove ν è un'opportuna funzione; come vedremo, il concetto è più complesso. La domanda quindi è quali sono le proprietà che la funzione ν deve avere per essere un “buon” modello, ossia tale che la minimizzazione di (1.3) fornisca una valida alternativa a (II).

Sia $\bar{x} = x_i$ l'iterata corrente e chiamiamo $\bar{x}_+ = x_{i+1}$ l'iterata successiva. Se d è la direzione di decrescita, \bar{x}_+ sarà della forma

$$\bar{x}_+ = \bar{x} + d,$$

e definiamo la funzione traslata

$$f_{\bar{x}}(d) = f(\bar{x} + d) - f(\bar{x}). \quad (1.4)$$

L'idea fondamentale ed innovativa di questo lavoro è che il passo fondamentale di tutti gli algoritmi bundle sia la soluzione di un problema nella forma

$$(\Pi_{\bar{x}}) \quad \min_d \{ \nu(f_{\bar{x}}(d)) \} \quad (1.5)$$

in cui ν è la *stabilizzazione della funzione obiettivo*. In effetti, quasi tutti gli algoritmi del Capitolo 2 considerano come parte del loro modello un *termine di stabilizzazione delle iterate* che regola la distanza tra l'iterata corrente e quella successiva. Questo termine serve per focalizzare la ricerca della prossima iterata sui punti più promettenti (tipicamente quelli che diminuiscono di più il valore della funzione obiettivo in (1.1)). Vedremo nel Capitolo 2 perché questo termine è introdotto e che forma può avere. Per ora diciamo che si tratta di una funzione chiusa e convessa che chiamiamo $D_t(x - \bar{x}) = D_t(d)$, dove il parametro t serve per “settare” il peso che ha il termine di stabilizzazione nella funzione obiettivo del problema. Per questo, in generale conviene considerare invece di (1.5) un problema nella forma

$$(\Pi_{\bar{x},t}) \quad \min_d \{ \nu(f_{\bar{x}}(d)) + D_t(d) \} \quad (1.6)$$

in cui, è visibile separatamente il termine di stabilizzazione ν della funzione ed il termine di stabilizzazione delle iterate D . Intuitivamente, ν deve “guidare la ricerca verso punti in cui la funzione obiettivo è migliore”, mentre D deve “evitare che la prossima iterata sia eccessivamente distante dall’iterata corrente”. In effetti, in generale dovremo assumere che anche ν abbia un parametro ϵ , in qualche modo analogo a t , che influenzi le caratteristiche del modello; rimandiamo però la discussione di questo punto al Capitolo 3.

L’idea alla base di questo lavoro è quella di determinare la forma e le proprietà che ν deve avere affinché la soluzione di (1.6) conduca ad un algoritmo convergente per il problema originario (II). È ovvio che ciò richiede anche opportune ipotesi sul termine D , ma queste sono già state studiate in [1]. A prima vista, questo scopo potrebbe apparire paradossale: infatti, è ovvio che la scelta $\nu(f(x)) = f(x)$ e $D_t(d) = 0$ garantisce queste proprietà; è anche chiaro però che in questo caso (1.6) diviene identico a (II), per il quale dobbiamo sviluppare un opportuno algoritmo risolutivo, e quindi non è stato fatto alcun passo avanti nella soluzione del problema. In effetti l’idea del modello è che (1.6) debba essere “significativamente più facile” da risolvere di (II); inoltre, deve poter essere risolto con la sola conoscenza di f rappresentata dal bundle β . Per questo consideriamo (1.6) la versione “astratta” degli algoritmi, cioè una descrizione di ciò che l’algoritmo farebbe ad ogni iterazione se f fosse nota. Poiché però f non si conosce, è poi necessario passare dalla forma astratta ad una forma “concreta”

$$(\Pi_{\beta, \bar{x}, t}) \quad \min_d \{ \nu(f_{\beta, \bar{x}}(d)) + D_t(d) \}, \quad (1.7)$$

nella quale invece della conoscenza di f si assume la sola conoscenza del modello f_β della funzione costruito con le informazione contenute nel bundle. Per tale forma si può spesso evitare di descrivere esattamente la forma di f_β , enunciando solamente alcune proprietà di base che essa deve avere; intuitivamente, l’idea è che “con un bundle infinitamente grande, f_β deve essere equivalente ad f ”. Al fine di rendere applicabile l’approccio è comunque necessario discutere anche della forma “implementabile” degli algoritmi, dove il modello f_β è specificato in dettaglio. Come vedremo, tipicamente si tratta del *modello dei piani di taglio* \hat{f}_β (vedere Sezione 2.1). Ciò conduce alla versione implementabile

$$(\hat{\Pi}_{\beta, \bar{x}, t}) \quad \min_d \{ \nu(\hat{f}_{\beta, \bar{x}}(d)) + D_t(d) \}, \quad (1.8)$$

Considerando che per una funzione g indichiamo con g^* la sua coniugata (definita da (A.1.6)), possiamo scrivere i *duali di Fenchel* dei nostri problemi: la versione duale del problema (1.6) è

$$(\Delta_{\bar{x}, t}) \quad - \min_g \{ (\nu(f_{\bar{x}}(\cdot)))^*(g) + D_t^*(-g) \}, \quad (1.9)$$

quella del problema (1.7) è

$$(\Delta_{\beta, \bar{x}, t}) \quad - \min_g \{(\nu(f_{\beta, \bar{x}}(\cdot)))^*(g) + D_t^*(-g)\} \quad (1.10)$$

e quella del problema (1.8) è

$$(\hat{\Delta}_{\beta, \bar{x}, t}) \quad - \min_g \{(\nu(\hat{f}_{\beta, \bar{x}}(\cdot)))^*(g) + D_t^*(-g)\}. \quad (1.11)$$

Vedremo in seguito il senso di tali problemi.

Riassumendo, nei problemi generalizzati (1.6), (1.7), (1.8), (1.9), (1.10), e (1.11) ci sono due tipi di stabilizzazione:

1. la stabilizzazione D_t che regola la distanza tra l'iterata corrente \bar{x} e l'iterata successiva \bar{x}_+ ;
2. la stabilizzazione ν che regola la differenza tra il valore della funzione obiettivo nell'iterata corrente $f(\bar{x})$ e nell'iterata successiva $f(\bar{x}_+)$.

Il nostro scopo è di scoprire se è possibile, facendo ipotesi giuste su ν e D_t , scrivere i primali astratti di tutti gli algoritmi del Capitolo 2 nella forma (1.6), i primali concreti nella forma (1.7) e i primali implementabili nella forma (1.8) e scrivere i duali astratti di tutti gli algoritmi nella forma (1.9), i duali concreti nella forma (1.10) e i duali implementabili nella forma (1.11). Inoltre intendiamo capire se sia possibile costruire una teoria di convergenza per l'algoritmo bundle così costruito che sia indipendente da ν , ossia che richieda solamente le "ipotesi giuste" individuate. Tale teoria di convergenza unificherebbe quindi in un solo trattamento molti algoritmi bundle fino ad oggi considerati diversi ed indipendenti.

Capitolo 2

Presentazione degli algoritmi

In questo capitolo presentiamo alcuni algoritmi bundle per risolvere (II). Abbiamo già detto che essi sono le stabilizzazioni diverse dell'algoritmo dei piani di taglio. Per questo partiremo con la descrizione dell'algoritmo dei piani di taglio e dopo presenteremo gli algoritmi stabilizzati.

2.1 Algoritmo dei piani di taglio

In questa sezione presentiamo l'algoritmo "classico" per la risoluzione di (II): l'algoritmo dei piani di taglio [2]. Cominceremo con la forma primale implementabile di questo algoritmo e poi ne daremo anche la forma concreta e la forma astratta. In fine, scriveremo anche le versioni duali.

L'approssimazione piú semplice di f è la funzione dei piani di taglio, che nell'iterata i -esima è:

$$\hat{f}_\beta(x) = \max_{i \in \beta} \{f(x_i) + g_i(x - x_i)\}. \quad (2.1)$$

Si noti come tale approssimazione usa sia il valore $f(x_i)$ che $g_i \in \partial f(x_i)$ ottenuti per ogni iterata precedente. La funzione \hat{f}_β è detta *modello dei piani di taglio* di f , ed è un'approssimazione poliedrale esterna di f con le seguenti proprietà:

$$\hat{f}_\beta \leq f \quad e \quad \hat{f}_\beta(x_i) = f(x_i) \quad i \in \beta. \quad (2.2)$$

L'algoritmo dei piani di taglio implementa un modello molto semplice che minimizza \hat{f}_β ad ogni iterazione:

$$\bar{x}_+ = \arg \min_x \hat{f}_\beta(x). \quad (2.3)$$

Questa è la forma implementabile dell'algoritmo dei piani di taglio: infatti, l'approssimazione di f usata in questo modello è nota (è \hat{f}_β definita da (2.1)). In questo caso la forma astratta corrisponde al problema (II) originale, in quanto non si usa nessun modello (alternativamente, si pone ν pari all'identità).

Per usi futuri introduciamo anche la funzione dei piani di taglio traslata

$$\hat{f}_{\beta, \bar{x}}(d) = \hat{f}_\beta(\bar{x} + d) - f(\bar{x}). \quad (2.4)$$

Riscriviamo $\hat{f}_\beta(\bar{x} + d)$ come

$$\begin{aligned} \hat{f}_\beta(\bar{x} + d) &= \max_{i \in \beta} \{f(x_i) + g_i(\bar{x} + d - x_i)\} \\ &= \max_{i \in \beta} \{f(x_i) + g_i(\bar{x} - x_i) + g_i d\} \\ &= \max_{i \in \beta} \{f(\bar{x}) - \alpha_i + g_i d\}, \end{aligned}$$

dove α_i è l'errore di linearizzazione definito da (A.17). Allora,

$$\hat{f}_{\beta, \bar{x}}(d) = \max_{i \in \beta} \{g_i d - \alpha_i\},$$

e risolviamo ad ogni iterazione

$$d^* = \arg \min_d \hat{f}_{\beta, \bar{x}}(d), \quad (2.5)$$

con $\bar{x}_+ = \bar{x} + d^*$. Risolvere (2.5) è equivalente a risolvere (2.3).

Possiamo adesso introdurre la *versione duale* dell'algoritmo dei piani di taglio. Questa può essere vista come un algoritmo che si occupa di calcolare il valore della coniugata della funzione dei piani di taglio nell'origine. Questo perché, per ogni funzione convessa f

$$\min_x f(x) = -\max_x \{0x - f(x)\} = -f^*(0).$$

È chiaro che non possiamo calcolare direttamente il valore $f^*(0)$, perché abbiamo a disposizione l'oracolo per f e non per f^* . Ricordiamo che, dato un x , l'oracolo per f restituisce un subgradiente $g \in \partial f(x)$. Allora, l'algoritmo duale ha lo scopo di determinare un x tale che il subgradiente calcolato dall'oracolo sia $g = 0$. Un tale x minimizza f . Infatti, la proprietà (vi) della coniugata ci dice che

$$g \in \partial f(x) \Leftrightarrow gx = f(x) + f^*(g).$$

Se $g = 0$ si ha

$$0 \in \partial f(x) \Leftrightarrow 0x = f(x) + f^*(0),$$

o, equivalentemente,

$$0 \in \partial f(x) \Leftrightarrow f(x) = -f^*(0),$$

cioé x è un punto di minimo per f .

Si può facilmente dimostrare che la coniugata della funzione dei piani di taglio è

$$\hat{f}_\beta^*(g) = \inf_\lambda \left\{ \sum_{i \in \beta} f^*(g_i) \lambda_i : \sum_{i \in \beta} g_i \lambda_i = g, \lambda \in \Delta \right\} \quad (2.6)$$

dove Δ è il semplice unitario, definito da (A.1.11). \hat{f}_β^* è un'approssimazione poliedrale interna di f^* , e

$$\hat{f}_\beta^* \geq f^*. \quad (2.7)$$

Poiché $g_i \in \partial f(x_i)$ e per (2.2) abbiamo $g_i \in \partial \hat{f}_\beta(x_i)$. Per la proprietà (vi) della coniugata si ha che

$$g_i \in \partial f(x_i) \implies f(x_i) + f^*(g_i) = x_i g_i$$

e

$$g_i \in \partial \hat{f}_\beta(x_i) \implies \hat{f}_\beta(x_i) + \hat{f}_\beta^*(g_i) = f(x_i) + \hat{f}_\beta^*(g_i) = x_i g_i,$$

e questo implica

$$f^*(g_i) = \hat{f}_\beta^*(g_i). \quad (2.8)$$

Per quanto riguarda $\hat{f}_{\beta, \bar{x}}^*(g)$, dalla (2.4) e dalle proprietà (iii) e (viii) della coniugata si ha che

$$\hat{f}_{\beta, \bar{x}}^*(g) = \hat{f}_\beta^*(g) - g\bar{x} + f(\bar{x}) = \alpha \quad (2.9)$$

(l'ultima uguaglianza è spiegata nell'appendice nella relazione (A.18)). Scriviamo il duale di Fenchel (A.1.21) di (2.3): (2.3) si può scrivere come

$$\min_x \{ \hat{f}_\beta(x) + \sigma_{\{0\}}(x) \},$$

e, allora, il duale di Fenchel è

$$\min_g \{ \hat{f}_\beta^*(g) + I_{\{0\}}(-g) \} = \min_g \{ \hat{f}_\beta^*(g) : g = 0 \}. \quad (2.10)$$

Quindi (2.10) è equivalente a

$$\min_\lambda \left\{ \sum_{i \in \beta} f^*(g_i) \lambda_i : \sum_{i \in \beta} g_i \lambda_i = g = 0, \lambda \in \Delta \right\}. \quad (2.11)$$

Analogamente il duale di Fenchel di (2.5) è

$$\min_g \{ \hat{f}_{\beta, \bar{x}}^*(g) + I_{\{0\}}(-g) \} = \min_g \{ \hat{f}_{\beta, \bar{x}}^*(g) : g = 0 \}. \quad (2.12)$$

o, equivalentemente,

$$\min_{\lambda} \left\{ \sum_{i \in \beta} f_{\bar{x}}^*(g_i) \lambda_i : \sum_{i \in \beta} g_i \lambda_i = g = 0, \lambda \in \Delta \right\}. \quad (2.13)$$

Per la relazione (2.9), si ha che (2.13) e (2.12) sono equivalenti a

$$- \min_{\lambda} \{ \alpha \lambda : \lambda G = 0, \lambda \in \Delta \}. \quad (2.14)$$

Consideriamo la forma concreta di (2.5). Per come è stata definita la forma concreta di un problema nella Sezione 1.1, possiamo dire che quella dell'algoritmo dei piani di taglio è equivalente a (2.5)—cioè minimizziamo un'approssimazione di f —ma senza specificare come è fatta $f_{\beta, \bar{x}}$. Cioè la versione primale concreta dell'algoritmo dei piani di taglio è

$$\min_d \{ f_{\beta, \bar{x}}(d) \}. \quad (2.15)$$

Lo stesso ragionamento vale anche per il duale concreto che allora è

$$\min_g \{ f_{\beta, \bar{x}}^*(g) : g = 0 \}. \quad (2.16)$$

Si noti che si possono scrivere le forme concrete equivalenti a quelle di sopra anche a partire da (2.3) (visto che è un problema equivalente a (2.5)). Si ottiene il primale concreto

$$\min_x \{ f_{\beta}(x) \},$$

il duale concreto

$$\min_g \{ f_{\beta}^*(g) : g = 0 \}.$$

Consideriamo ora la forma astratta dell'algoritmo dei piani di taglio. Per come è stata definita la forma astratta di un problema nella Sezione 1.1, possiamo dire che se f fosse nota, avremo minimizzato f stessa e non qualche sua approssimazione. Allora, se partiamo da (2.3) il primale astratto è esattamente (II) e il duale astratto $\min_g \{ f^*(g) : g = 0 \}$, che come abbiamo notato è equivalente a (II). Se partiamo da (2.5) otteniamo la seguente versione primale astratta dell'algoritmo dei piani di taglio:

$$\min_d \{ f_{\bar{x}}(d) \} \quad (2.17)$$

(ed è ancora un problema equivalente a (Π) , perché è semplicemente la versione traslata di (Π)). Lo stesso ragionamento vale anche per il duale astratto che allora è

$$\min_g \{ f_{\bar{x}}^*(g) : g = 0 \}. \quad (2.18)$$

Si noti che, usando le proprietà (iii) e (viii) della coniugata, si ha

$$f_{\bar{x}}^*(g) = f^*(g) - g\bar{x} + f(\bar{x}). \quad (2.19)$$

Riassumendo, per l'algoritmo dei piani di taglio, la forma astratta del primale è

$$(\Pi_{\bar{x}}) \quad \min_d \{ f_{\bar{x}}(d) \},$$

la forma concreta del primale è

$$(\Pi_{\beta, \bar{x}}) \quad \min_d \{ f_{\beta, \bar{x}}(d) \},$$

la forma implementabile del primale è

$$(\hat{\Pi}_{\beta, \bar{x}}) \quad \min_d \{ \hat{f}_{\beta, \bar{x}}(d) \},$$

la forma astratta del duale è

$$(\Delta_{\bar{x}}) \quad \min_g \{ f_{\bar{x}}^*(g) : g = 0 \},$$

la forma concreta del duale è

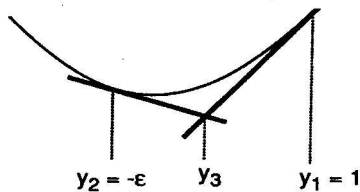
$$(\Delta_{\beta, \bar{x}}) \quad \min_g \{ f_{\beta, \bar{x}}^*(g) : g = 0 \},$$

e la forma implementabile del duale è

$$(\hat{\Delta}_{\beta, \bar{x}}) \quad \min_g \{ \hat{f}_{\beta, \bar{x}}^*(g) : g = 0 \}.$$

L'algoritmo dei piani di taglio è instabile.

Esempio 2.1.1 Questo esempio è tratto da [4, XV]. Consideriamo la seguente figura:



Qui, $n = 1$ e $f(x) = \frac{1}{2}x^2$. Osserviamo la terza iterazione, dopo che nelle prime due sono state generate due iterate: $y_1 = 1$ e $y_2 = -\epsilon < 0$. È chiaro che y_3 è la soluzione di

$$\hat{f}_{\beta=\{y_1\}}(y) = \hat{f}_{\beta=\{y_1, y_2\}}(y),$$

cioé

$$\frac{1}{2} + y - 1 = \frac{1}{2}(-\epsilon)^2 + (-\epsilon)(y + \epsilon),$$

e $y_3 = \frac{1}{2} - \frac{1}{2}\epsilon$ (i subgradienti g_1 e g_2 sono uguali rispettivamente a y_1 e y_2 perché f è ovunque differenziabile e $\partial f(y_i) = \nabla f(y_i) = \nabla \frac{1}{2}y_i^2 = y_i$). Al diminuire di ϵ , incrementano sia y_2 che y_3 , ma per un ϵ abbastanza piccolo, l'iterata corrente y_2 è migliore (piú vicina allo 0) della successiva y_3 - algoritmo è instabile".

L'instabilità, dunque, sta nel fatto che l'iterata successiva può essere arbitrariamente lontana dalla corrente. Il termine di stabilizzazione D_t viene introdotto per "tirla indietro". Ciò è utile poiché i punti lontani dall'iterata corrente non sono un'approssimazione migliore di quella che già abbiamo. Questo perché \hat{f}_β approssima la funzione obiettivo vicino all'iterata corrente e difficilmente sarà una buona approssimazione lontano da essa. Il termine di stabilizzazione D_t serve per limitare l'insieme dove si cerca l'iterata successiva.

Quindi, la costruzione di un'algoritmo dei piani di taglio stabilizzato inizia con la scelta di un centro di stabilizzazione che tipicamente è \bar{x} e di una norma, di solito la norma Euclidea. $D_t(d)$ deve permetterci di ottenere la stabilità dell'algoritmo in un intorno del centro di stabilizzazione, delimitato con la particolare norma scelta. Il prossimo punto tentativo viene allora calcolato cercando di ottenere un compromesso tra la condizione (1.2) e il mantenere piú piccola possibile la distanza (calcolata in norma scelta) tra il punto corrente e il punto tentativo che stiamo considerando. Questo procedimento, naturalmente, non è applicabile solo all'approssimazione dei piani di taglio, ma a ogni approssimazione f_β che è una sottostima di f , cioè tale che $f_\beta \leq f$ o, in termini di coniugata, tale che $f_\beta^* \geq f^*$ (proprietá (ii) della coniugata).

La scelta della stabilizzazione ν riflette la scelta di un modo di calcolare il minimo di f . Quindi, le nostre forme generali (1.6), (1.7), (1.8), (1.9), (1.10), (1.11) sono delle versioni dell'algoritmo dei piani di taglio stabilizzate. Osserviamo che l'algoritmo dei piani di taglio puro che abbiamo presentato in questa sezione, può essere pensato come caso particolare delle nostre forme generali con $\nu = Id$ e $D_t(d) = 0 \forall d, t$. Gli algoritmi presentati in seguito, implementano tutti delle stabilizzazioni ν e D_t dell'algoritmo dei piani di taglio diverse.

2.2 Penalty e trust-region bundle

Alcune delle versioni stabilizzate piú semplici dell'algoritmo dei piani di taglio sono trust-region e penalty bundle.

Gli algoritmi di tipo trust-region limitano la ricerca dei punti tentativi "vicino" al centro di stabilizzazione, dove per vicino si intende la palla associata con la data norma, centrata nel centro di stabilizzazione, di raggio t . Trovare l'iterata successiva $\bar{x}_+ = \bar{x} + d^*$ significa risolvere:

$$(\Pi_{\bar{x},t}) \quad d^* = \min_d \{f_{\bar{x}}(d) : \|d\|_2 \leq t\}.$$

La forma concreta di trust-region è:

$$(\Pi_{\beta,\bar{x},t}) \quad d^* = \min_d \{f_{\beta,\bar{x}}(d) : \|d\|_2 \leq t\}.$$

La forma implementabile di trust-region è:

$$(\hat{\Pi}_{\beta,\bar{x},t}) \quad d^* = \min_d \{\hat{f}_{\beta,\bar{x}}(d) : \|d\|_2 \leq t\}.$$

Il duale di Fenchel (A.1.21) del problema con la trust-region è

$$\begin{aligned} & \min_d \{f_{\bar{x}}(d) : \|d\|_2 \leq t\} \\ &= \min_d \{f_{\bar{x}}(d) : d \in B(0, t)\} \\ &= \min_d \{f_{\bar{x}}(d) + I_{B(0,t)}(d)\}. \end{aligned} \tag{2.20}$$

Per (A.6) e per (A.14) abbiamo il duale di Fenchel di (2.20):

$$\min_g \{f_{\bar{x}}^*(g) + \sigma_{B(0,t)}(-g)\} = \min_g \{f_{\bar{x}}^*(g) + t\| -g\|_2\}$$

(l'ultima equazione deriva da $\sigma_{B(0,t)}(-g) = \sup_{d \in B(0,t)} \{-gd\} = \sup_d \{-gd : \|d\|_2 \leq t\}$. Dato che $t\|\frac{-g}{\| -g\|_2}\|_2 = t$, possiamo dire che $d = t\frac{-g}{\| -g\|_2}$ è ammissibile per $\sup_{d \in B(0,t)} \{-gd\}$. Allora, $\sup_d \{-gd : \|d\|_2 \leq t\} \geq t(-g)\frac{-g}{\| -g\|_2} = t\| -g\|_2$).

Riassumendo, il duale astratto di trust-region è

$$(\Delta_{\bar{x},t}) \quad \min_g \{f_{\bar{x}}^*(g) + t\| -g\|_2\},$$

il duale concreto è

$$(\Delta_{\beta,\bar{x},t}) \quad \min_g \{f_{\beta,\bar{x}}^*(g) + t\| -g\|_2\},$$

e il duale implementabile è

$$(\hat{\Delta}_{\beta,\bar{x},t}) \quad \min_g \{ \hat{f}_{\beta,\bar{x}}^*(g) + t \| -g \|_2 \} = \\ - \min_{\lambda} \{ \alpha \lambda + t \| -\lambda G \|_2 : \lambda \in \Delta \}.$$

Gli algoritmi di tipo penalty bundle assegnano un peso ad ogni punto tentativo. I pesi si basano sulla distanza del punto tentativo dal centro di stabilizzazione. Dato il parametro della penalit  t , scriviamo i sei modelli. Il modello astratto  :

$$(\Pi_{\bar{x},t}) \quad d^* = \min_d \{ f_{\bar{x}}(d) + \frac{1}{2} t \| d \|_2^2 \},$$

il modello concreto di penalty bundle  :

$$(\Pi_{\beta,\bar{x},t}) \quad d^* = \min_d \{ f_{\beta,\bar{x}}(d) + \frac{1}{2} t \| d \|_2^2 \},$$

e il modello implementabile di penalty bundle  :

$$(\hat{\Pi}_{\beta,\bar{x},t}) \quad d^* = \min_d \{ \hat{f}_{\beta,\bar{x}}(d) + \frac{1}{2} t \| d \|_2^2 \}.$$

Qui, il peso   dato dalla norma Euclidea, mentre t serve per ‘‘settare’’ l’impatto che ha questo termine di stabilizzazione alla soluzione. Infatti,   facile vedere che, ad esempio, aumentare t , significa avvicinare sempre di pi  il punto tentativo al centro di stabilizzazione e viceversa, diminuire t , significa allontanare sempre di pi  il punto tentativo dal centro di stabilizzazione.

Scriviamo il duale astratto di Fenchel (A.1.21) del penalty bundle (assumiamo $t > 0$):

$$(\Delta_{\bar{x},t}) = \min_g \{ f_{\bar{x}}^*(g) + (\frac{1}{2} t \| -g \|_2^2)^* \} \\ = \min_g \{ f_{\bar{x}}^*(g) + \frac{1}{2} t \| -\frac{2g}{t} \|_2^2 \} \\ = \min_g \{ f_{\bar{x}}^*(g) + \frac{2}{t} \| -g \|_2^2 \}$$

Nello stesso modo, il duale concreto  

$$(\Delta_{\beta,\bar{x},t}) \quad \min_g \{ f_{\beta,\bar{x}}^*(g) + \frac{2}{t} \| -g \|_2^2 \},$$

e il duale implementabile  

$$(\hat{\Delta}_{\beta,\bar{x},t}) \quad \min_g \{ \hat{f}_{\beta,\bar{x}}^*(g) + \frac{2}{t} \| -g \|_2^2 \} = \\ \min_{\lambda} \{ \alpha \lambda + \frac{2}{t} \| -\lambda G \|_2^2 : \lambda \in \Delta \}. \quad (2.21)$$

La stabilizzazione ν   semplicemente l’identit  Id .

2.3 Metodi bundle generalizzati

Un primo tentativo di unificare la teoria di convergenza di diversi algoritmi bundle, in particolare proprio trust-region e proximal, è stato fatto in [1] con la proposta di una generalizzazione dei metodi del tipo penalty bundle nella quale è stata mantenuta la stabilizzazione $\nu = Id$, mentre il termine di stabilizzazione $\frac{1}{2}t\|d\|_2^2$ è stato sostituito con una funzione generica $D_t(d)$, della quale non è specificata nessuna particolare forma, ma solo alcune proprietà necessarie a garantire la convergenza dell'algoritmo.

Il problema primale astratto stabilizzato è

$$(\Pi_{\bar{x},t}) \quad \inf_d \{f_{\bar{x}}(d) + D_t(d)\} \quad (2.22)$$

a cui corrisponde il problema primale concreto stabilizzato

$$(\Pi_{\beta,\bar{x},t}) \quad \inf_d \{f_{\beta,\bar{x}}(d) + D_t(d)\}, \quad (2.23)$$

In [1] si propone per la prima volta che l'approssimazione f_β , usata nell'algoritmo per risolvere (Π) , non sia necessariamente specificata (tipicamente posta uguale a \hat{f}_β); l'unico vincolo sulla sua scelta è che deve essere una sottostima di f , cioè $f_\beta \leq f$. Ancora, in [1] si propone per la prima volta un generico duale astratto

$$(\Delta) \quad \inf_g \{f^*(g) : g = 0\}. \quad (2.24)$$

che abbiamo già visto essere equivalente a (Π) . Una proprietà interessante è che per il rilassamento Lagrangiano di (Δ) rispetto ai vincoli $g = 0$, usando \bar{x} come vettore di moltiplicatori Lagrangiani,

$$(\Delta_{\bar{x}}) \quad \inf_g \{f^*(g) - g\bar{x}\}.$$

dalla definizione (A.1.6) e dalla proprietà (i) della coniugata si ha

$$-v(\Delta_{\bar{x}}) = \sup_g \{g\bar{x} - f^*(g)\} = (f^*)^*(\bar{x}) = f(\bar{x}),$$

dove $v(\cdot)$ indica il valore ottimo del problema. In altri termini, il rilassamento Lagrangiano di (Δ) rispetto ai vincoli $g = 0$ è il problema di ottimizzazione che l'oracolo deve risolvere per calcolare $f(x)$ ed il subgradiente. Possiamo allora scrivere il problema duale astratto stabilizzato come

$$(\Delta_{\bar{x},t}) \quad \inf_g \{f_{\bar{x}}^*(g) + D_t^*(-g)\}. \quad (2.25)$$

Dato che $f_{\bar{x}}^*(g) = f^*(g) - \bar{x}g$, da questo risulta che l'algoritmo bundle generalizzato è, dal punto di vista duale, un *Lagrangiano aumentato generalizzato* di (Δ) in cui i vincoli $g = 0$ sono portati in funzione obiettivo con un "termine al primo ordine" $\bar{x}g$ ed un "termine al secondo ordine" $D_t^*(g)$. Si capisce quindi come sia possibile risolvere (Δ) risolvendo ripetutamente (2.25) e aggiornando \bar{x} e t . Poiché in generale (2.25) è difficile da risolvere quanto (Δ) , si passa quindi alla versione concreta in cui f è rimpiazzata dall'approssimazione f_β

$$(\Delta_{\beta, \bar{x}, t}) \quad \inf_g \{f_{\beta, \bar{x}}^*(g) + D_t^*(-g)\}. \quad (2.26)$$

Per la proprietà (ii) della coniugata si ha $f_\beta^* \geq f^*$, e pertanto (2.26) fornisce un'approssimazione superiore di (2.25), che può diventare arbitrariamente accurata continuando ad inserire elementi in β . In particolare, con la versione implementabile che usa il modello dei piani di taglio \hat{f}_β l'algoritmo implementa uno schema di approssimazione interna per la soluzione iterativa di (2.25).

Proprietá di D_t

Il termine di stabilizzazione D_t è una funzione convessa e chiusa e parametrizzata in t , ma la sua esatta forma non è specificata; per esempio, D_t può essere un fattore come $\frac{1}{2}t\| \cdot \|_2^2$ nel penalty bundle, oppure il raggio di una trust-region. Il parametro di penalitá t determina la "forza" di D_t . Affinché l'algoritmo sia convergente, D_t deve rispettare alcune proprietà:

- (P1) $\forall t > 0, D_t(0) = 0$ e $0 \in \partial D_t(0)$ (D_t è nonnegativa).
- (P2) $\forall t > 0$ e $\epsilon > 0, S_\epsilon(D_t)$ è compatto e $0 \in \text{int } S_\epsilon(D_t)$.
- (P3) $\forall t > 0, \lim_{\|d\|_2 \rightarrow \infty} D_t(d)/\|d\|_2 = +\infty$ (D_t è fortemente coerciva).
- (P4) $\forall t > 0, D_t \geq D_\tau$ per ogni $\tau \geq t$ (D_t è noncrescente in t).
- (P5) $\lim_{t \rightarrow \infty} D_t(d) = 0 \forall d$ (D_t converge puntualmente alla funzione identicamente nulla).

Queste condizioni su D_t sono equivalenti alle seguenti condizioni su D_t^*

- (P*1) $\forall t > 0, D_t^*(0) = 0$ e $0 \in \partial D_t^*(0)$ (D_t^* è nonnegativa).
- (P*2) $\forall t > 0$ e $\epsilon > 0, S_\epsilon(D_t^*)$ è compatto e $0 \in \text{int } S_\epsilon(D_t^*)$.
- (P*3) $\forall t > 0, D_t^*$ è finita ovunque.
- (P*4) $\forall t > 0, D_t^* \leq D_\tau^*$ per ogni $\tau \geq t$ (D_t^* è nondecreciente in t).

(P*5) $\forall \epsilon > 0, \lim_{t \rightarrow \infty} \inf_g \{D_t^*(g) : \|g\|_2 \geq \epsilon\} = +\infty$ ($\{D_t^*\}$ converge “uniformemente” a $I_{\{0\}}$).

Si possono dare anche versioni diverse della proprietà; ad esempio, è possibile evitare l’assunzione sulla forte coercività (P3) se sappiamo che $(\Pi_{\beta, \bar{x}, t})$ è limitata dal basso

(P3’) f è limitata dal basso, un f_* finito, tale che $f_* \leq v(\Pi)$ è noto e $f_\beta \geq f_*$ $\forall \beta$.

(P*3’) Δ è non vuoto, un f_* finito, tale che $f^*(0) \leq -f_*$ è noto e $f_\beta^*(0) \leq -f_*$ $\forall \beta$.

Alternativamente, si ottengono proprietà di convergenza più forti utilizzando

(P3’’) $\forall t$ D_t è fortemente coerciva e strettamente convessa.

(P*3’’) $\forall t$ D_t^* è finito ovunque e differenziabile.

(P5’) $\forall t$ $\partial D_t(0) = \{0\}$ (D_t è differenziabile in 0, cioè $\nabla D_t(0) = 0$).

(P*5’) $\forall t$ $S_0(D_t^*) = \{0\}$ (D_t^* è strettamente convesso in 0, cioè 0 è l’unico minimo di D_t^*).

Sotto queste ipotesi si può dimostrare la convergenza dell’algoritmo; si noti che questo permette di unificare gli algoritmi proximal bundle con quelli di tipo trust region, perché sia $D_t(\cdot) = 1/(2t)\|\cdot\|^2$ che $D_t(\cdot) = I_{B(0,t)}(\cdot)$ hanno le proprietà necessarie.

2.4 Level bundle

In [5], viene presentato l’algoritmo Level bundle nel suo aspetto primale. Qui si sceglie, ad ogni iterazione, il livello t che sarà il limite al valore della funzione obiettivo e il problema da risolvere diventa

$$(\Pi_{\bar{x}, t}) \quad \min_d \left\{ \frac{1}{2} \|d\|_2^2 : f_{\bar{x}}(d) \leq t \right\}. \quad (2.27)$$

oppure nella forma concreta

$$(\Pi_{\beta, \bar{x}, t}) \quad \min_d \left\{ \frac{1}{2} \|d\|_2^2 : f_{\beta, \bar{x}}(d) \leq t \right\}, \quad (2.28)$$

oppure nella forma implementabile

$$(\hat{\Pi}_{\beta, \bar{x}, t}) \quad \min_d \left\{ \frac{1}{2} \|d\|_2^2 : \hat{f}_{\beta, \bar{x}}(d) \leq t \right\}. \quad (2.29)$$

Nella forma concreta e nella forma implementabile, il livello t si può scegliere nel modo seguente: ad ogni iterazione, scegliamo un livello t che soddisfa

$$\inf f_{\beta, \bar{x}} \leq t < f(\bar{x}).$$

Se non vale la prima disuguaglianza, l'insieme delle soluzioni ammissibili del Level bundle è vuoto, mentre la seconda disuguaglianza guida la sequenza $\{f_{\beta, \bar{x}}\}$ verso il minimo valore di f . Questo perché l'iterata corrente \bar{x} viene aggiornata solo se è stato ottenuto un decremento della funzione obiettivo da $f(\bar{x})$ verso t . Allora, il nuovo t sarà più piccolo di $f(\bar{x}_+)$ e, di conseguenza, anche il valore della funzione obiettivo all'iterazione successiva. Nella forma astratta, t dovrebbe soddisfare

$$\inf f \leq t < f(\bar{x}),$$

ma questo presuppone che $\inf f$ esista e che sia noto.

Riscrivendo (2.27) come

$$\min_d \{I_{\{f_{\bar{x}}(\cdot) \leq t\}}(d) + \frac{1}{2} \|d\|_2^2\}$$

si nota che nel Level bundle il valore dell'iterata successiva è controllato sia con un termine di stabilizzazione che limita la distanza tra essa e l'iterata corrente (D_t), che da un termine di stabilizzazione sul valore della funzione obiettivo. In altre parole, il Level bundle può essere visto come una variante stabilizzata dell'algoritmo dei piani di taglio dove $D_t(d) = \frac{1}{2} \|d\|_2^2$ e $\nu(f_{\bar{x}}(d)) = I_{\{d: f_{\bar{x}}(d) \leq t\}}$. Lo stesso vale per (2.28) e (2.29), con rispettivamente $f_{\beta, \bar{x}}$ e $\hat{f}_{\beta, \bar{x}}$ al posto di $f_{\bar{x}}$.

2.5 Metodo dei piani di taglio con i centri di Chebychev

L'idea dell'approccio proposto in [7, 8] è quella di scegliere la prossima iterata "al centro" del possibile insieme di valori.

È noto che f può essere scritta in termini dei suoi iperpiani di supporto come

$$f(x) = \sup \{f(x') + g(x - x') : x' \in R^n, g \in \partial f(x')\}. \quad (2.30)$$

L'epigrafo di f è allora

$$\begin{aligned} \text{epi } f &= \{(x, r) \in R^n \times R : f(x) \leq r\} \\ &= \{(x, r) \in R^n \times R : f(x') + g(x - x') \leq r, x' \in R^n, g \in \partial f(x')\}. \end{aligned}$$

Dato un punto \bar{x} , $f(\bar{x})$ è il limite superiore al valore ottimo di (Π) e la parte dell'epigrafo che sta sotto $f(\bar{x})$ definisce una curva di livello

$$X_{\bar{x}} = \{(x, r) \in R^n \times R : r \leq f(\bar{x}), f(x') + g(x - x') \leq r, \\ x' \in R^n, g \in \partial f(x')\} \quad (2.31)$$

che contiene l'insieme ottimale $(X^*, v(\Pi))$. $X_{\bar{x}}$ è detto *insieme di localizzazione*. L'idea è di scegliere, ad ogni iterazione, l'iterata successiva \bar{x}_+ in modo da “restringere il più possibile” l'insieme di localizzazione. Per fare questo, in alcuni algoritmi si usano i centri legati all'insieme di localizzazione; in particolare in questo approccio si usa il *centro di Chebychev*, ossia il centro del cerchio più grande iscrivibile nell'insieme.

Per trovare il centro di Chebychev dell'insieme $X_{\bar{x}}$ occorre risolvere il problema lineare semi-infinito

$$\begin{aligned} \max \sigma : \\ r + \sigma &\leq f(\bar{x}), \\ f(x') + g(x - x') - r + \sigma \sqrt{1 + \|g\|_2^2} &\leq 0, \\ \sigma, r \in R, x \in R^n, x' \in R^n, g \in \partial f(x'). \end{aligned} \quad (2.32)$$

La Proposizione 2.1 di [8] ci dice che vicino all'ottimo il vincolo $r + \sigma \leq f(\bar{x})$ è stretto, cioè vale uguaglianza e quindi possiamo eliminare questo vincolo e la variabile r e denotando con v l'opposto di σ , riscriviamo (2.32) come

$$\begin{aligned} \min v : \frac{g(x - x') + f(x') - f(\bar{x})}{1 + \sqrt{1 + \|g\|_2^2}} &\leq v \\ v \in R, x \in R^n, x' \in R^n, g \in \partial f(x'). \end{aligned} \quad (2.33)$$

La funzione

$$\psi_{\bar{x}}(x) = \sup \left\{ \frac{g(x - x') + f(x') - f(\bar{x})}{1 + \sqrt{1 + \|g\|_2^2}} : x' \in R^n, g \in \partial f(x') \right\}. \quad (2.34)$$

$\psi_{\bar{x}}$ è convessa poiché è il massimo puntuale di altre funzioni convesse [[3], Proposizione IV.2.1.2]; il problema (2.33) è equivalente al problema

$$\min_{x \in R^n} \psi_{\bar{x}}(x). \quad (2.35)$$

la cui soluzione ottima \bar{x}_+ è il centro di Chebychev, ed il cui valore ottimo è l'opposto del raggio del più grande cerchio iscrivibile in $X_{\bar{x}}$ ed avente centro \bar{x}_+ . È quindi possibile strutturare una variante dell'algoritmo dei piani di

taglio, detta di *Elzinga-Moore*, in cui si usano i centri di Chebychev. Si noti che solo il taglio $r \leq f(\bar{x})$ guida la discesa verso la soluzione di (II). Ogni volta il punto corrente \bar{x} è spostato a \bar{x}_+ e la funzione obiettivo diminuisce di almeno il diametro del cerchio piú grande iscrivibile in $X_{\bar{x}}$; quindi $X_{\bar{x}}$ tende all'insieme ottimo e la sequenza di centri di Chebychev degli $X_{\bar{x}}$ tende ad una soluzione ottima.

Determinare il centro di Chebychev dell'insieme $X_{\bar{x}}$ richiede che esso sia limitato. In piú, le approssimazioni di $X_{\bar{x}}$ e $\psi_{\bar{x}}$ (che saranno date sotto) sono "locali", nel senso che approssimano $X_{\bar{x}}$ e $\psi_{\bar{x}}$ vicino al punto corrente e difficilmente saranno buone approssimazioni lontano da esso. Quindi si rafforza la ricerca del prossimo punto "vicino" al punto corrente aggiungendo il termine quadratico di stabilizzazione D_t in (2.35). Il termine quadratico garantisce compattezza e quindi si può eliminare la condizione su $X_{\bar{x}}$ limitato (vedere la definizione (A.1.3)).

Dato il parametro $t > 0$, scriviamo (2.35) con la stabilizzazione:

$$\Phi_{\bar{x}}(x) = \min_{y \in R^n} \{ \psi_{\bar{x}}(y) + \frac{t}{2} \|y - x\|_2^2 \}. \quad (2.36)$$

Il problema

$$\min_{x \in R^n} \Phi_{\bar{x}}(x) \quad (2.37)$$

è equivalente a (2.35). Ovviamente, minimizzare $\psi_{\bar{x}}$ avendo a disposizione un oracolo per f non è un problema facile, e quindi occorre passare alla versione concreta/implementabile. In particolare, usando il modello dei piani di taglio \hat{f}_β si ottiene l'approssimazione esterna dell'insieme di localizzazione $X_{\bar{x}}$

$$X_{\beta, \bar{x}} = \{ (x, r) \in R^n \times R : r \leq f(\bar{x}), f(x_i) + g_i(x - x_i) \leq r, \\ i \in \beta, x_i \in R^n, g_i \in \partial f(x_i) \} \quad (2.38)$$

che, dunque, contiene le soluzioni ottime di (II). Ragionando nello stesso modo come per $\psi_{\bar{x}}$ costruiamo

$$\psi_{\beta, \bar{x}}(x) = \max_{i \in \beta} \left\{ \frac{g_i(x - x_i) + f(x_i) - f(\bar{x})}{1 + \sqrt{1 + \|g_i\|_2^2}} \right\}. \quad (2.39)$$

che è un'approssimazione dal basso di $\psi_{\bar{x}}$, cioè $\psi_{\beta, \bar{x}}(x) \leq \psi_{\bar{x}}(x)$. Il problema

$$\min_{x \in R^n} \psi_{\beta, \bar{x}}(x) \quad (2.40)$$

è allora un'approssimazione di (2.35). La versione stabilizzata di (2.40) è

$$\Phi_{\beta, \bar{x}}(x) = \min_{x \in R^n} \{ \psi_{\beta, \bar{x}}(x) + \frac{t}{2} \|x - \bar{x}\|_2^2 \}. \quad (2.41)$$

o, equivalentemente,

$$\begin{aligned} \min \quad & v + \frac{t}{2} \|x - \bar{x}\|_2^2 : \\ & \tilde{g}_i(x - \bar{x}) - \tilde{\alpha}_i \leq v, \\ & i \in \beta, v \in R, x \in R^n, \end{aligned} \quad (2.42)$$

dove α_i è l'errore di linearizzazione tra x_i e \bar{x} definito dalla (A.17), $\tilde{g}_i = \gamma_i g_i$, $\tilde{\alpha}_i = \gamma_i \alpha_i$ e $\gamma_i = (1 + \sqrt{1 + \|g_i\|_2^2})^{-1}$. In [8] è dato anche il seguente duale di (2.42)

$$\min \left\{ \frac{1}{2t} \left\| \sum_{i=1}^k \lambda_i \tilde{g}_i \right\|_2^2 + \sum_{i=1}^k \lambda_i \tilde{\alpha}_i : \lambda \in \Delta_k \right\}, \quad (2.43)$$

dove Δ_k è simplesso unitario nello spazio R^k .

Come vedremo nel Capitolo 3, il problema (2.37) può essere scritto come

$$\min_d \left\{ \sup_{g, \epsilon} \left\{ \frac{f_{\bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_\epsilon f_{\bar{x}}(d) \right\} + \frac{t}{2} \|d\|_2^2 \right\},$$

il problema (2.41) come

$$\min_d \left\{ \sup_{g, \epsilon} \left\{ \frac{f_{\beta, \bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_\epsilon f_{\beta, \bar{x}}(d) \right\} + \frac{t}{2} \|d\|_2^2 \right\},$$

e il problema (2.42) come

$$\min_d \left\{ \sup_{g, \epsilon} \left\{ \frac{\hat{f}_{\beta, \bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_\epsilon \hat{f}_{\beta, \bar{x}}(d) \right\} + \frac{t}{2} \|d\|_2^2 \right\}.$$

In altre parole, l'algoritmo dei piani di taglio con centri di Chebychev può essere visto come una variante stabilizzata dell'algoritmo dei piani di taglio puro dove $D_t(d) = \frac{t}{2} \|d\|_2^2$ e $\nu(f_{\bar{x}}(d)) = \min_d \left\{ \sup_{g, \epsilon} \left\{ \frac{f_{\bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_\epsilon f_{\bar{x}}(d) \right\} \right\}$.

2.6 Dual bundle

In [4] viene presentato l'algoritmo Dual Bundle. Questo algoritmo nasce per migliorare gli algoritmi del subgradiente coniugato. In questi metodi, l'idea

era di trovare la miglior direzione di decrescita d^* , che è noto essere (l'opposto del)la soluzione ottima di

$$\min\{\|g\|_2 : g \in \partial f(x)\},$$

per poi effettuare un opportuno passo di discesa lungo d^* . Poiché in generale non si conosce $\partial f(x)$ è necessario costruirne un'approssimazione utilizzando i subgradienti g_i noti, ossia quelli ottenuti alle precedenti iterate. In particolare, è noto che se $\beta(x)$ è il sottoinsieme di β che contiene i soli vettori che sono subgradienti in x , allora $\text{conv}\{g_i : i \in \beta(x)\} \subseteq \partial f(x)$; si può quindi prendere questo insieme come approssimazione interna del subdifferenziale (in pratica si calcolano gli errori di linearizzazione definiti dalla (A.17) e si cancellano quei g_i per i quali α_i è “troppo grande”).

In pratica, questi metodi si comportano come il metodo del gradiente, che è noto essere lento e può convergere a un punto non ottimale quando f non è differenziabile. Intuitivamente, il problema è che $\partial f(x)$ è “troppo piccolo”. Quindi l'idea del metodo Dual bundle è quella di allargare la ricerca di d^* a $\partial_\epsilon f(x)$ per un opportuno valore $\epsilon > 0$, ossia di risolvere

$$\min\{\|g\|_2 : g \in \partial_\epsilon f(x)\} ; \quad (2.44)$$

è noto che se $g^* = 0$ allora vale (A.5) e x è la soluzione ϵ -ottima di (II), altrimenti g^* è una direzione di ϵ -decrescita, ossia esiste $p > 0$ tale che $f(x - pg^*) \leq f(x) - \epsilon$. Poiché anche $\partial_\epsilon f(x)$ non si conosce, deve essere approssimato; analogamente al caso precedente, l'insieme

$$G(\epsilon) = \left\{ g = \sum_{i \in \beta} \lambda_i g_i : \lambda_i \geq 0 \quad \forall i \in \beta, \sum_{i \in \beta} \lambda_i = 1, \sum_{i \in \beta} \lambda_i \alpha_i \leq \epsilon \right\} \quad (2.45)$$

è un'approssimazione interna di $\partial_\epsilon f(\bar{x})$. Quindi, per calcolare d^* dobbiamo risolvere il problema dei minimi quadrati vincolato

$$\min_g \left\{ \frac{1}{2} \left\| \sum_{i \in \beta} \lambda_i g_i \right\|_2^2 : \lambda_i \geq 0 \quad \forall i \in \beta, \sum_{i \in \beta} \lambda_i = 1, \sum_{i \in \beta} \lambda_i \alpha_i \leq \epsilon \right\}. \quad (2.46)$$

L'osservazione fondamentale è che $G(\epsilon)$ è l'approssimazione di $\partial_\epsilon f(\bar{x})$ che corrisponde alla funzione dei piani di taglio, ossia $\partial_\epsilon \hat{f}(\bar{x}) = G(\epsilon)$. Notoriamente, la funzione coniugata è strettamente legata all' ϵ -subdifferenziale, infatti, per le proprietà (iii) e (v) si ha

$$g \in \partial_\epsilon f(\bar{x}) \Leftrightarrow f^*(g) - g\bar{x} + f(\bar{x}) \leq \epsilon \Leftrightarrow f_{\bar{x}}^*(g) \leq \epsilon .$$

pertanto il Dual bundle astratto (2.44) può essere riscritto come

$$\min_g \{ \|g\|_2 + I_{S_\epsilon(f_{\bar{x}}^*)}(g) \}, \quad (2.47)$$

il Dual bundle concreto come

$$\min_g \{ \|g\|_2 + I_{S_\epsilon(f_{\beta, \bar{x}}^*)}(g) \},$$

e il Dual bundle implementabile come

$$\min_g \{ \|g\|_2 + I_{S_\epsilon(\hat{f}_{\beta, \bar{x}}^*)}(g) \}.$$

Ricordando che l'algoritmo dei piani di taglio può essere visto come il problema di minimizzare $f_{\bar{x}}^*$ col vincolo $g = 0$, l'algoritmo Dual bundle può essere visto come una variante stabilizzata del duale dell'algoritmo dei piani di taglio dove $D_t^*(-g) = \|-g\|_2$ e $(\nu(f_{\bar{x}}(\cdot)))^*(g) = I_{\{f_{\bar{x}}^*(\cdot) \leq \epsilon\}}(g)$.

2.7 Separating plane algorithms (SPA)

L'algoritmo di [6] è "inerentemente duale", in quanto risolve (II) calcolando il valore di $f^*(0)$ (come abbiamo visto, i due problemi sono equivalenti). Intuitivamente, dato che il punto $(f^*(0), 0)$ appartiene al epi f^* , basterebbe conoscere epi f^* "attorno all'origine" per conoscere $f^*(0)$. Poiché però f non è nota (ossia lo è solamente attraverso l'oracolo), allora non lo sono neanche f^* e epi f^* ; pertanto occorre approssimare epi f^* usando il bundle β .

Per evitare alcune complicazioni nelle dimostrazioni di convergenza, in [6] sono fatte le seguenti assunzioni su f : il minimo di f esiste, $0 \in \text{intdom } f^*$ e $f(0) = 0$. Per quanto riguarda l'ultima, nei fatti questa significa semplicemente che si lavora con la funzione traslata $f_{\bar{x}}$ invece che con la f originale. Si noti che, (2.19) con $g = 0$ ci da

$$f_{\bar{x}}^*(0) = f^*(0) + f(\bar{x}).$$

In altri termini, calcolare $f_{\bar{x}}^*(0)$ è equivalente a calcolare $f^*(0)$ dato che $f(\bar{x})$ è noto.

L'approccio proposto in [6] consiste nel costruire sequenze di approssimazioni esterne e interne, rispettivamente $\{A_O^k\}$ e $\{A_i^k\}$, di epi $f_{\bar{x}}^*$, che ad ogni iterazione sono migliorate per approssimare sempre più accuratamente epi $f_{\bar{x}}^*$ vicino al punto $(f_{\bar{x}}^*(0), 0)$. In particolare, l'approssimazione interna utilizzata è $\hat{f}_{\beta, \bar{x}}^*$, che abbiamo già visto essere "l'oggetto duale" nei Paragrafi 2.1, 2.3 e 2.6. L'approssimazione esterna invece è calcolata attraverso il modello dei piani di taglio duale; poichè

$$f(x_i) + f^*(g_i) = g_i x_i$$

e si ha

$$f_{\bar{x}}(d) = f(\bar{x} + d) - f(\bar{x}) \quad f_{\bar{x}}^*(g) = f^*(g) - \bar{x}g + f(\bar{x})$$

si ha anche

$$f_{\bar{x}}(x_i - \bar{x}) + f_{\bar{x}}^*(g_i) = g_i(x_i - \bar{x})$$

per tutti i punti in β , ossia $(x_i - \bar{x}) \in \partial f_{\bar{x}}^*(g_i)$. Pertanto

$$\check{f}_{\bar{x},\beta}^*(g) = \max_{i \in \beta} \{ f_{\bar{x}}^*(g_i) + (x_i - \bar{x})(g - g_i) \}$$

è tale che $\check{f}_{\bar{x},\beta}^* \leq f_{\bar{x}}^*$; in altri termini, epi $\check{f}_{\bar{x},\beta}^*$ è un'approssimazione esterna di epi $f_{\bar{x}}^*$. Si noti che la funzione può essere equivalentemente riscritta come

$$\check{f}_{\bar{x},\beta}^*(g) = \max_{i \in \beta} \{ f(\bar{x}) - f(x_i) + (x_i - \bar{x})g \} .$$

Di conseguenza,

$$\check{f}_{\bar{x},\beta}^*(0) = \max_{i \in \beta} \{ f(\bar{x}) - f(x_i) \} \leq f_{\bar{x}}^*(0)$$

ossia si ha una valutazione inferiore su $f_{\bar{x}}^*(0)$; infatti il valore è equivalente a $f(\bar{x}) - \min_{i \in \beta} \{ f(x_i) \} = - \min_{i \in \beta} \{ f_{\bar{x}}(x_i - \bar{x}) \}$, che è evidentemente una valutazione inferiore sul valore minimo di f (si noti che se \bar{x} è la migliore delle iterate questo valore è 0).

Ovviamente, se si ha che $\hat{f}_{\beta,\bar{x}}^*(0) = \check{f}_{\bar{x},\beta}^*(0)$ allora si è determinato il valore di $f^*(0)$ (e di $f_{\bar{x}}^*(0)$). Altrimenti si deve determinare un iperpiano che separa il punto $(\check{f}_{\bar{x},\beta}^*(0), 0)$ da epi f^* , in modo da raffinare l'approssimazione interna (ed eventualmente esterna). Dato che si ha a disposizione un oracolo per f e non per f^* , l'unica operazione possibile in pratica su epi f^* è il calcolo del sup di forme lineari su esso:

$$\begin{aligned} (\text{epi } f^*)_{(x,-1)} &= \sigma_{\text{epi } f^*}(x, -1) = \\ \sup_{(g,r) \in \text{epi } f^*} \{gx - r\} &= \sup_g \{gx - f^*(g)\} = f(x). \end{aligned} \quad (2.48)$$

Per questo bisogna individuare un iperpiano $(x, -1)$ che separa l'approssimazione interna epi $\hat{f}_{\beta,\bar{x}}^*$ dal punto $(\check{f}_{\bar{x},\beta}^*(0), 0)$; assumendo che \bar{x} sia effettivamente la migliore delle iterate, questo significa separare $\hat{f}_{\beta,\bar{x}}^*$ dall'origine $(0, 0)$. Il modo standard di risolvere un problema di separazione tra un punto ed un insieme è quello di scriverlo come *problema di minima distanza tra di essi*. Si potrebbe cioè risolvere

$$\min_g \{ \|\hat{f}_{\beta,\bar{x}}^*(g), h\| \}$$

(di cui è utile rimarcare la notevole somiglianza con (2.47). In [6] ciò viene generalizzato considerando la distanza

$$\|g\|_2^2 + \psi(f^*(g)),$$

oppure

$$\|g\|_2^2 + \psi(f_{\bar{x}}^*(g)),$$

dove

$$\begin{aligned} \psi(\omega) \text{ è funzione monotona, crescente, non negativa di} \\ \omega \geq 0, \psi(0) = 0, \psi(\omega) > 0 \text{ per } \omega > 0. \end{aligned} \quad (2.49)$$

Ciò porta ad un problema astratto

$$\min_g \{ \psi(f_{\bar{x}}^*(g)) + \|g\|_2^2 \}.$$

a cui corrisponde il problema concreto

$$\min_g \{ \psi(f_{\bar{x},\beta}^*(g)) + \|g\|_2^2 \}.$$

e quello implementabile

$$\min_g \{ \psi(\hat{f}_{\bar{x},\beta}^*(g)) + \|g\|_2^2 \}.$$

che può essere posto in forma esplicita, usando (2.6),

$$\min \left\{ \left| \sum_{i \in \beta} \lambda_i g_i \right|_2^2 + \psi \left(\sum_{i \in \beta} \lambda_i f^*(g_i) \right) : \lambda \in \Delta \right\}. \quad (2.50)$$

Il vettore separatore $(x, -1)$ si ottiene considerando la soluzione del duale di (2.50)

$$\min_{x,\tau,v} \left\{ \|x\|_2^2 + v + \max_{\omega} \{ \tau\omega - \psi(\omega) \} : v \geq xg_i - \tau f^*(g_i) \quad i \in \beta \right\} \quad (2.51)$$

che, usando le proprietà della coniugata (cf. Teorema A.1.2), può essere riscritto come

$$\min_d \{ (\psi \circ f_{\bar{x}}^*)^*(d) + \|d\|_2^2 \}.$$

In altre parole, l'algoritmo SPA può essere visto dal punto di vista duale come una variante stabilizzata dell'algoritmo dei piani di taglio dove $D_t^*(-g) = \| -g \|_2^2$ e $(\nu(f_{\bar{x}}(\cdot)))^*(g) = \psi(f_{\bar{x}}^*(g))$; dal punto di vista primale è quindi una variante stabilizzata dell'algoritmo dei piani di taglio dove $D_t(d) = \|d\|_2^2$ e $\nu(f_{\bar{x}}(d)) = (\psi \circ f_{\bar{x}}^*)^*(d)$.

Capitolo 3

Metodi bundle come le stabilizzazioni dell'algoritmo dei piani di taglio

In questo capitolo spieghiamo in che modo, tutti gli algoritmi, presentati nel Capitolo 2, possono essere visti come le stabilizzazioni diverse dell'algoritmo dei piani di taglio. Scriviamo, per tutti, le sei versioni del problema di minimizzazione di f : $(\Pi_{\bar{x},t})$, $(\Pi_{\beta,\bar{x},t})$, $(\hat{\Pi}_{\beta,\bar{x},t})$, $(\Delta_{\bar{x},t})$, $(\Delta_{\beta,\bar{x},t})$ e $(\hat{\Delta}_{\beta,\bar{x},t})$. Omettiamo penalty e trust-region bundle, in quanto generalizzati dai Metodi bundle generalizzati.

3.1 Algoritmo dei piani del taglio

Le sei versioni dell'algoritmo dei piani di taglio sono:

$$(\Pi_{\bar{x}}) \quad \min_d \{ f_{\bar{x}}(d) \},$$

la forma concreta del primale è

$$(\Pi_{\beta,\bar{x}}) \quad \min_d \{ f_{\beta,\bar{x}}(d) \},$$

la forma implementabile del primale è

$$(\hat{\Pi}_{\beta,\bar{x}}) \quad \min_d \{ \hat{f}_{\beta,\bar{x}}(d) \},$$

la forma astratta del duale è

$$(\Delta_{\bar{x}}) \quad \min_g \{ f_{\bar{x}}^*(g) : g = 0 \},$$

la forma concreta del duale è

$$(\Delta_{\beta, \bar{x}}) \quad \min_g \{f_{\beta, \bar{x}}^*(g) : g = 0\},$$

e la forma implementabile del duale è

$$(\hat{\Delta}_{\beta, \bar{x}}) \quad \min_g \{\hat{f}_{\beta, \bar{x}}^*(g) : g = 0\}.$$

3.2 Metodi bundle generalizzati

Scriviamo la forma implementabile di $(\Pi_{\beta, \bar{x}, t})$, cioè nel caso in cui f_β è l'approssimazione dei piani di taglio:

$$\begin{aligned} \min_d \{\hat{f}_{\beta, \bar{x}}(d) + D_t(d)\} &= \min_d \{\max_{i \in \beta} \{g_i d - \alpha_i\} + D_t(d)\} \\ &= \min_{d, v} \{v + D_t(d) : v \geq g_i d - \alpha_i \quad \forall i \in \beta\} \\ &= \min_{d, v} \{v + D_t(d) : v \geq Gd - \alpha\}. \end{aligned} \quad (3.1)$$

Scriviamo la forma implementabile di $(\Delta_{\beta, \bar{x}, t})$, cioè nel caso in cui f_β è l'approssimazione dei piani di taglio:

$$\begin{aligned} \inf_g \{\hat{f}_{\beta, \bar{x}}^*(g) + D_t^*(-g)\} &= \min_\lambda \left\{ \sum_{i \in \beta} f^*(g_i) \lambda_i - \left(\sum_{i \in \beta} g_i \lambda_i \right) \bar{x} + f(\bar{x}) + D_t^* \left(- \sum_{i \in \beta} g_i \lambda_i \right) \right. \\ &\quad \left. : \sum_{i \in \beta} g_i \lambda_i = g, \lambda \in \Delta \right\} \\ &= \min_\lambda \{\alpha \lambda + D_t^*(-\lambda G) : \lambda \in \Delta\}. \end{aligned} \quad (3.2)$$

Per i metodi bundle generalizzati, la forma astratta del primale è

$$(\Pi_{\bar{x}, t}) \quad \inf_d \{f_{\bar{x}}(d) + D_t(d)\},$$

la forma concreta del primale è

$$(\Pi_{\beta, \bar{x}, t}) \quad \inf_d \{f_{\beta, \bar{x}}(d) + D_t(d)\},$$

e la forma implementabile del primale è:

$$(\hat{\Pi}_{\beta, \bar{x}, t}) \quad \min_d \{v + D_t(d) : v \geq Gd - \alpha\},$$

oppure

$$(\hat{\Pi}_{\beta, \bar{x}, t}) \quad \inf_d \{\hat{f}_{\beta, \bar{x}}(d) + D_t(d)\}.$$

La forma astratta del duale è

$$(\Delta_{\bar{x},t}) \quad \inf_g \{f_{\bar{x}}^*(g) + D_t^*(-g)\},$$

la forma concreta del duale è

$$(\Delta_{\beta,\bar{x},t}) \quad \inf_g \{f_{\beta,\bar{x}}^*(g) + D_t^*(-g)\},$$

e la forma implementabile del duale è:

$$(\hat{\Delta}_{\beta,\bar{x},t}) \quad \min_{\lambda} \{\alpha\lambda + D_t^*(-\lambda G) : \lambda \in \Delta\},$$

oppure

$$(\hat{\Delta}_{\beta,\bar{x},t}) \quad \inf_g \{\hat{f}_{\beta,\bar{x}}^*(g) + D_t^*(-g)\}.$$

3.3 Level bundle

Per poter scrivere le sei versioni dell'algoritmo, dobbiamo dare anche un'interpretazione duale del Level bundle.

Per scrivere il duale di (2.27), usiamo il duale di Fenchel (A.1.21).

Riscriviamo (2.27) in una forma piú appropriata:

$$\min_d \left\{ \frac{1}{2} \|d\|_2^2 : f_{\bar{x}}(d) \leq t \right\} \quad \equiv \quad \min_d \left\{ \frac{1}{2} \|d\|_2^2 + I_{S_t(f_{\bar{x}})}(d) \right\}. \quad (3.3)$$

oppure

$$\min_d \left\{ \frac{1}{2} \|d\|_2^2 + \nu_t(f_{\bar{x}}(d)) \right\}, \quad (3.4)$$

dove $\nu_t(f_{\bar{x}}(d)) = I_{S_t(f_{\bar{x}})}(d)$.

Il duale di Fenchel di (3.3) è

$$\min_g \left\{ \frac{1}{2} \| -g \|_2^2 + (I_{S_t(f_{\bar{x}})})^*(g) \right\}. \quad (3.5)$$

oppure

$$\min_g \left\{ \frac{1}{2} \| -g \|_2^2 + (\nu_t(f_{\bar{x}}))^*(g) \right\}, \quad (3.6)$$

dove $(\nu_t(f_{\bar{x}}))^*(g) = (I_{S_t(f_{\bar{x}})})^*(g)$.

$(I_{S_t(f_{\bar{x}})})^*$ è la coniugata di una composizione di due funzioni. Possiamo, quindi, usare il Teorema A.1.2 per calcolare $(I_{S_t(f_{\bar{x}})})^*$. Poniamo

$$\nu_t(v) = I_{\{v \leq t < 0\}} = \begin{cases} 0 & \text{se } v \leq t \\ +\infty & \text{altrimenti,} \end{cases}$$

e $f = f_{\bar{x}}$. Allora, entrambe l'ipotesi del teorema sono soddisfatte.

Dato che

$$\begin{aligned}\nu_t^*(\gamma) &= \sigma_{\{v \leq t < 0\}}(\gamma) \\ &= \sup_{v \leq t < 0} \{\gamma v\} \\ &= \gamma t + I_{R_+}(\gamma),\end{aligned}\tag{3.7}$$

$(I_{S_t(f_{\bar{x}})})^*$ è

$$\min_{\gamma \geq 0} \left\{ \gamma f_{\bar{x}}^* \left(\frac{1}{\gamma} g \right) + \gamma t \right\}.\tag{3.8}$$

Allora il duale astratto è

$$(\Delta_{\bar{x}, t}) \quad \min_{\gamma \geq 0, g} \left\{ \gamma f_{\bar{x}}^* \left(\frac{1}{\gamma} g \right) + \gamma t + D_t^*(-g) \right\},\tag{3.9}$$

dove $D_t^*(-g) = \frac{1}{2} \| -g \|_2$. Il duale concreto è

$$(\Delta_{\beta, \bar{x}, t}) \quad \min_{\gamma \geq 0, g} \left\{ \gamma f_{\beta, \bar{x}}^* \left(\frac{1}{\gamma} g \right) + \gamma t + D_t^*(-g) \right\}.\tag{3.10}$$

Scriviamo la forma implementabile del duale.

$$(\hat{\Delta}_{\beta, \bar{x}, t}) \quad \min_{\gamma \geq 0, g} \left\{ \gamma \hat{f}_{\beta, \bar{x}}^* \left(\frac{1}{\gamma} g \right) + \gamma t + D_t^*(-g) \right\},\tag{3.11}$$

il che è equivalente a

$$\min_{\gamma \geq 0, g, \lambda'} \left\{ \gamma \alpha \lambda' + \gamma t + D_t^*(-g) : \lambda' \in \Delta, G\lambda' = \frac{1}{\gamma} g \right\}.\tag{3.12}$$

Poniamo $\lambda = \gamma \lambda'$. Si ha

$$\min_{\gamma \geq 0, \lambda, g} \left\{ \alpha \lambda + \gamma t + D_t^*(-g) : \lambda \geq 0, e\lambda = \gamma, G\lambda = g \right\},\tag{3.13}$$

cioé

$$(\hat{\Delta}_{\beta, \bar{x}, t}) \quad \min_{\lambda} \left\{ \alpha \lambda + t(e\lambda) + D_t^*(-G\lambda) : \lambda \geq 0 \right\}\tag{3.14}$$

Lo stesso risultato si ottiene scrivendo il duale quadratico (A.1.20) di (2.29). Infatti, (2.29) è equivalente a:

$$\min_d \left\{ \frac{1}{2} \|d\|_2^2 : g_i d - \alpha_i \leq t \quad \forall i \in \beta \right\},$$

dove α_i è l'errore di linearizzazione tra \bar{x} e x_i definito da (A.17). Questo si può scrivere anche come

$$\min_d \left\{ \frac{1}{2} \|d\|_2^2 : Gd - \alpha \leq te \right\}.$$

Poniamo $Q = I, q = c = 0$ e

$$A = \begin{bmatrix} -G \end{bmatrix}$$

e

$$b = \begin{bmatrix} -te - \alpha \\ \lambda = \begin{bmatrix} \lambda \end{bmatrix} \end{bmatrix}$$

Quindi il duale quadratico è:

$$\max_{\lambda, \gamma} \left\{ -\frac{1}{2} \|\gamma\|_2^2 + \lambda(-te - \alpha) : -\lambda G + \gamma = 0, \lambda \geq 0 \right\}, \quad (3.15)$$

il che è equivalente a

$$-\min_{\lambda \geq 0} \left\{ \frac{1}{2} \|\lambda G\|_2^2 + (\alpha + te)\lambda \right\}, \quad (3.16)$$

il che è equivalente a (3.14).

Riassumendo, per il Level bundle, la forma astratta del primale è

$$(\Pi_{\bar{x}, t}) \quad \min_d \left\{ \frac{1}{2} \|d\|_2^2 : f_{\bar{x}}(d) \leq t \right\},$$

oppure

$$(\Pi_{\bar{x}, t}) \quad \min_d \{ I_{S_t(f_{\bar{x}})}(d) + D_t(d) \},$$

la forma concreta del primale è

$$(\Pi_{\beta, \bar{x}, t}) \quad \min_d \left\{ \frac{1}{2} \|d\|_2^2 : f_{\beta, \bar{x}}(d) \leq t \right\},$$

oppure

$$(\Pi_{\beta, \bar{x}, t}) \quad \min_d \{ I_{S_t(f_{\beta, \bar{x}})}(d) + D_t(d) \},$$

e la forma implementabile del primale è

$$(\hat{\Pi}_{\beta, \bar{x}, t}) \quad \min_d \left\{ \frac{1}{2} \|d\|_2^2 : \hat{f}_{\beta, \bar{x}}(d) \leq t \right\},$$

oppure

$$(\hat{\Pi}_{\beta, \bar{x}, t}) \quad \min_d \{ I_{S_t(\hat{f}_{\beta, \bar{x}})}(d) + D_t(d) \},$$

dove $D_t(d) = \frac{1}{2}\|d\|_2^2$.

La forma astratta del duale è

$$(\Delta_{\bar{x},t}) \quad \min_{\gamma \geq 0, g} \left\{ \gamma f_{\bar{x}}^*\left(\frac{1}{\gamma}g\right) + \gamma t + D_t^*(-g) \right\},$$

oppure

$$(\Delta_{\bar{x},t}) \quad \min_g \left\{ (I_{S_t(f_{\bar{x}})})^*(g) + D_t^*(-g) \right\},$$

la forma concreta del duale è

$$(\Delta_{\beta, \bar{x}, t}) \quad \min_{\gamma \geq 0, g} \left\{ \gamma f_{\beta, \bar{x}}^*\left(\frac{1}{\gamma}g\right) + \gamma t + D_t^*(-g) \right\},$$

oppure

$$(\Delta_{\beta, \bar{x}, t}) \quad \min_g \left\{ (I_{S_t(f_{\beta, \bar{x}})})^*(g) + D_t^*(-g) \right\},$$

e la forma implementabile del duale è

$$(\hat{\Delta}_{\beta, \bar{x}, t}) \quad - \min_{\lambda \geq 0} \left\{ \frac{1}{2} \|\lambda G\|_2^2 + (\alpha + t\epsilon)\lambda \right\},$$

oppure

$$(\hat{\Delta}_{\beta, \bar{x}, t}) \quad \min_{\gamma \geq 0, g} \left\{ \gamma \hat{f}_{\beta, \bar{x}}^*\left(\frac{1}{\gamma}g\right) + \gamma t + D_t^*(-g) \right\},$$

oppure

$$(\hat{\Delta}_{\beta, \bar{x}, t}) \quad \min_g \left\{ (I_{S_t(\hat{f}_{\beta, \bar{x}})})^*(g) + D_t^*(-g) \right\},$$

dove $D_t^*(-g) = \frac{1}{2}\| -g\|_2^2$.

3.4 Metodo dei piani di taglio con i centri di Chebychev

Algoritmo dei piani di taglio con i centri di Chebychev, nella versione primale astratta, risolve ad ogni iterazione:

$$\min_x \left\{ \psi_{\bar{x}}(x) + \frac{t}{2} \|x - z\|_2^2 \right\},$$

dove, $\psi_{\bar{x}}(x)$ è definita da (2.34) e z è il corrente centro di stabilizzazione.

Consideriamo un valore ammissibile s tale che $s \geq f_*$ ($\exists \tilde{x} : s = f(\tilde{x})$) e definiamo

$$\psi_s(x) = \sup_{x', g} \left\{ \frac{g(x - x') + f(x') - s}{1 + \sqrt{1 + \|g\|_2^2}} : x' \in R^n, g \in \partial f(x') \right\}. \quad (3.17)$$

Usando la definizione (A.1.6) della coniugata, riscriviamo (3.17) come

$$\begin{aligned}
\psi_s(x) &= \sup_{x',g} \left\{ \frac{gx - f^*(g) - s}{1 + \sqrt{1 + \|g\|_2^2}} : x' \in R^n, g \in \partial f(x') \right\} \\
&= \sup_g \left\{ \frac{gx - f^*(g) - s}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \text{dom } f^* \right\} \\
&= \sup_g \left\{ \frac{gx - f^*(g) - s}{1 + \sqrt{1 + \|g\|_2^2}} \right\}.
\end{aligned} \tag{3.18}$$

Prendiamo $s = f(\bar{x})$. Allora (3.18) diventa

$$\psi_{\bar{x}}(x) = \sup_g \left\{ \frac{gx - f^*(g) - f(\bar{x})}{1 + \sqrt{1 + \|g\|_2^2}} \right\}. \tag{3.19}$$

Da (2.19) si ha che $-f^*(g) - f(\bar{x}) = -g\bar{x} - f_{\bar{x}}^*(g)$ e (3.19) diventa

$$\begin{aligned}
\psi_{\bar{x}}(d) &= \sup_g \left\{ \frac{g(x - \bar{x}) - f_{\bar{x}}^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} \right\} \\
&= \sup_g \left\{ \frac{gd - f_{\bar{x}}^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} \right\}.
\end{aligned} \tag{3.20}$$

Dalla proprietà (v) della coniugata sappiamo che $f_{\bar{x}}(d) + f_{\bar{x}}^*(g) = dg + \epsilon \forall g \in \partial_{\epsilon} f_{\bar{x}}(d)$.

Allora (3.20) è equivalente a

$$\psi_{\bar{x}}(d) = \sup_{g,\epsilon} \left\{ \frac{f_{\bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_{\epsilon} f_{\bar{x}}(d) \right\}. \tag{3.21}$$

Prendiamo \bar{x} come centro di stabilità. Otteniamo la seguente forma astratta del primale

$$(\Pi_{\bar{x},t}) \quad \min_d \left\{ \sup_{g,\epsilon} \left\{ \frac{f_{\bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_{\epsilon} f_{\bar{x}}(d) \right\} + \frac{t}{2} \|d\|_2^2 \right\}.$$

La forma concreta del primale è

$$(\Pi_{\beta,\bar{x},t}) \quad \min_d \left\{ \sup_{g,\epsilon} \left\{ \frac{f_{\beta,\bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_{\epsilon} f_{\beta,\bar{x}}(d) \right\} + \frac{t}{2} \|d\|_2^2 \right\}.$$

La forma implementabile del primale è

$$(\hat{\Pi}_{\beta,\bar{x},t}) \quad \min_d \left\{ \sup_{g,\epsilon} \left\{ \frac{\hat{f}_{\beta,\bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_{\epsilon} \hat{f}_{\beta,\bar{x}}(d) \right\} + \frac{t}{2} \|d\|_2^2 \right\}.$$

Questo è equivalente al primale implementabile (2.42). Infatti, (2.42) è

$$\begin{aligned}
(\hat{\Pi}_{\beta, \bar{x}, t}) \quad \min \quad & v + \frac{t}{2} \|x - \bar{x}\|_2^2 : \\
& \tilde{g}_i(x - \bar{x}) - \tilde{\alpha}_i \leq v, i \in \beta, \\
& v \in R, x \in R^n, \tilde{g}_i = \gamma_i g_i, \tilde{\alpha}_i = \gamma_i \alpha_i, \gamma_i = (1 + \sqrt{1 + \|g_i\|_2^2})^{-1}, i \in \beta.
\end{aligned}$$

Questo problema si può riscrivere come

$$\begin{aligned}
\min_{v, d} \quad & v + \frac{t}{2} \|d\|_2^2 : \\
& \gamma_i g_i d - \gamma_i \alpha_i \leq v, i \in \beta, \\
& v \in R, d \in R^n, \gamma_i = (1 + \sqrt{1 + \|g_i\|_2^2})^{-1}, i \in \beta,
\end{aligned}$$

o, equivalentemente

$$\min_d \left\{ \max_{g \in \beta} \left\{ \frac{gd - \alpha}{1 + \sqrt{1 + \|g\|_2^2}} \right\} + \frac{t}{2} \|d\|_2^2 \right\},$$

il che è equivalente, usando la definizione dell'errore di linearizzazione, a

$$\min_d \left\{ \max_{g \in \beta} \left\{ \frac{gd - \hat{f}_{\beta, \bar{x}}^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} \right\} + \frac{t}{2} \|d\|_2^2 \right\},$$

dove il primo membro della somma è esattamente (3.20) con $\hat{f}_{\beta, \bar{x}}$ al posto di $f_{\bar{x}}$. Quindi si può rifare un ragionamento analogo al primale astratto per arrivare ad avere che (2.42) è equivalente a

$$\min_d \left\{ \sup_{g, \epsilon} \left\{ \frac{\hat{f}_{\beta, \bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_\epsilon \hat{f}_{\beta, \bar{x}}(d) \right\} + \frac{t}{2} \|d\|_2^2 \right\}.$$

Per poter scrivere le sei versioni dell'algorithm dei piani di taglio con i centri di Chebychev, dobbiamo scriverne anche le versioni duali.

Consideriamo la forma astratta del primale con $\psi_{\bar{x}}(d)$ definita da (3.20) e, a partire da essa, calcoliamo il duale astratto. Come prima cosa usiamo [4], Ch.X, Th.2.4.4. che dice

Teorema 3.4.1 *Sia $\{f_j\}_{j \in J}$ una collezione di funzioni chiuse e convesse da R^n in R . Il loro supremo $f = \sup_{j \in J} f_j$, a patto che sia diverso da $+\infty$, è chiuso e convesso da R^n in R e la sua coniugata è closed-convex hull delle f_j^* :*

$$\left(\sup_{j \in J} f_j \right)^* = \bar{co} \left(\inf_{j \in J} f_j^* \right).$$

Nel nostro caso, $J = R^n$. Invece, la coniugata di $\frac{gd - f_x^*(g)}{1 + \sqrt{1 + \|g\|_2^2}}$ è ottenuta applicando le proprietà (viii) e (ix) della coniugata e tenendo presente che $(gd)^* = I_g(\cdot)$. Si ottiene:

$$\begin{aligned} (\sup_g \{ \frac{gd - f_x^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} \})^*(z) &= \bar{co}(\inf_{g \in R^n} \{ \frac{1}{1 + \sqrt{1 + \|g\|_2^2}} (f_x^*(g) + I_g(\cdot(1 + \sqrt{1 + \|g\|_2^2}))) \})(z) \\ &= \bar{co}(\inf_{g \in R^n} \left\{ \begin{array}{ll} \frac{1}{1 + \sqrt{1 + \|g\|_2^2}} f_x^*(g) & \text{se } g = z(1 + \sqrt{1 + \|g\|_2^2}) \\ +\infty & \text{altrimenti.} \end{array} \right\}) \\ &= \bar{co}(\inf_{g \in R^n} \{ \frac{1}{1 + \sqrt{1 + \|g\|_2^2}} f_x^*(g) : z = \frac{g}{1 + \sqrt{1 + \|g\|_2^2}} \}) \end{aligned}$$

Quindi il duale è

$$\min_z \{ \bar{co}(\inf_{g \in R^n} \{ \frac{1}{1 + \sqrt{1 + \|g\|_2^2}} f_x^*(g) : z = \frac{g}{1 + \sqrt{1 + \|g\|_2^2}} \}) + \frac{1}{2t} \|z\|_2^2 \}. \quad (3.22)$$

Consideriamo “l’interno” del nostro closed convex hull:

$$\inf_{g \in R^n} \{ \frac{1}{1 + \sqrt{1 + \|g\|_2^2}} f_x^*(g) : z = \frac{g}{1 + \sqrt{1 + \|g\|_2^2}} \}. \quad (3.23)$$

Fare il \bar{co} di (3.23) significa cercare le combinazioni convesse di un numero infinito dei g_i ($g_i \in R^n$).

Per [3], Ch. III, Proposizione 1.3.4., il co di un insieme S può essere descritto anche come l’insieme di tutte le combinazioni convesse dei suoi elementi:

$$\begin{aligned} co(S) &= \{ g \in R^n : \text{per qualche } k \in N, \exists g_1, \dots, g_k \in S, \\ &\quad \exists \lambda_1, \dots, \lambda_k \in \Delta_k : g = \sum_{i=1}^k \lambda_i g_i \}. \end{aligned} \quad (3.24)$$

Se S è infinito, o ha un numero infinito di elementi g_i , allora k in (3.24) può essere $+\infty$. Significa che alcuni g sono combinazioni convesse di un numero infinito dei g_i . Questo è precisamente quello che succede in (3.23).

Anche se S è infinito, alcuni dei g generati in (3.24) sono combinazioni convesse di un numero finito dei g_i (basta pensare alle combinazioni convesse con alcuni $\lambda_i = 0$). Infatti, $\forall k \neq \infty$ possiamo definire S_k - l’insieme di tutte le combinazioni convesse di k elementi $g_i \in S$ (sono infinite perché i g_i sono

infiniti). Chiaramente, con k g_i possiamo generare solo un sottoinsieme di tutte le combinazioni convesse generabili dagli elementi di S . Infatti, si ha che

$$S_1 \subset S_2 \subset \dots \subset S_k \subset \dots$$

e

$$\lim_{k \rightarrow \infty} \cup S_k = co(S).$$

In altre parole, le combinazioni convesse di un numero infinito dei g_i si possono calcolare trovando, per ogni k , tutte le combinazioni convesse di k g_i (dunque le infinite combinazioni convesse di un numero finito di elementi) con $k \rightarrow \infty$. Questo non sembra migliorare molto il nostro problema di calcolo di combinazioni convesse di un numero infinito dei g_i . Ma il Teorema di Carathéodory ci dice che k non deve andare all'infinito. Bastano $n + 1$ elementi di S per generare qualunque combinazione convessa in $co(S)$. In altre parole, $co(S) = S_{n+1}$.

Teorema 3.4.2 (Teorema di Carathéodory) *Ogni $x \in co(S) \subset R^n$ può essere rappresentato come una combinazione convessa di $n + 1$ elementi di S .*

Noi usiamo questo risultato per fare l'operazione di \bar{co} su (3.23). Qui, $S = \text{dom } f_{\bar{x}}^* = R^n$. Definiamo

$$V_{n+1} = \{W \subseteq \text{dom } f_{\bar{x}}^* : |W| = n + 1\}, \quad (3.25)$$

l'insieme degli insiemi costituiti da $n + 1$ elementi di $\text{dom } f_{\bar{x}}^*$. Allora il \bar{co} di (3.23) è

$$\inf_{W \in V_{n+1}} \left\{ \inf_{\lambda \in \Delta_{|W|}} \left\{ \sum_{g \in W} \lambda_g \frac{f_{\bar{x}}^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} : \sum_{g \in v} \lambda_g \frac{g}{1 + \sqrt{1 + \|g\|_2^2}} = z \right\} \right\}. \quad (3.26)$$

In fine, il duale astratto è

$$(\Delta_{\bar{x}, t}) \quad \min_z \left\{ \inf_{W \in V_{n+1}} \left\{ \inf_{\lambda \in \Delta_{|W|}} \left\{ \sum_{g \in W} \lambda_g \frac{f_{\bar{x}}^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} : \sum_{g \in R^n} \lambda_g \frac{g}{1 + \sqrt{1 + \|g\|_2^2}} = z \right\} \right\} + \frac{1}{2t} \|z\|_2^2 \right\}, \quad (3.27)$$

Applichiamo (3.27) all'approssimazione $f_{\beta, \bar{x}}$ e alla funzione dei piani del taglio. In questo caso, S di (3.24) è uguale a $\text{dom } f_{\beta, \bar{x}}^* = \text{dom } \hat{f}_{\beta, \bar{x}}^* = \beta$ ed è, dunque, un insieme finito. Si ha

$$co(\beta) = \{g \in R^n : \text{per qualche } k \in N, \exists g_1, \dots, g_k \in \beta, \\ \exists \lambda_1, \dots, \lambda_k \in \Delta_k : g = \sum_{i=1}^k \lambda_i g_i\}.$$

[3], Ch. III, Esempio 1.3.5. ci dice che per ottenere il $co(\beta)$ non è necessario generare tutte le combinazioni convesse ottenute per ogni k , dove $k = 1, \dots, |\beta|$, ma possiamo restringerci al caso $k = |\beta|$. Allora non ci serve più l'insieme V_{n+1} definito in (3.25), ma piuttosto l'insieme

$$V_{|\beta|} = \{W \subseteq \beta : |W| = |\beta|\} = \{\beta\}.$$

A questo punto, il duale concreto ((3.27) applicata alla funzione $f_{\beta, \bar{x}}$) è

$$\min_z \left\{ \inf_{\lambda \in \Delta_{|\beta|}} \left\{ \sum_{g \in \beta} \lambda_g \frac{f_{\beta, \bar{x}}^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} : \sum_{g \in \beta} \lambda_g \frac{g}{1 + \sqrt{1 + \|g\|_2^2}} = z \right\} + \frac{1}{2t} \|z\|_2^2 \right\},$$

e il duale implementabile ((3.27) applicata alla funzione dei piani del taglio) è

$$\min_z \left\{ \inf_{\lambda \in \Delta_{|\beta|}} \left\{ \sum_{g \in \beta} \lambda_g \frac{\hat{f}_{\beta, \bar{x}}^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} : \sum_{g \in \beta} \lambda_g \frac{g}{1 + \sqrt{1 + \|g\|_2^2}} = z \right\} + \frac{1}{2t} \|z\|_2^2 \right\}.$$

Quest'ultimo è equivalente a

$$\min_{\lambda \in \Delta_{|\beta|}} \left\{ \sum_{i \in \beta} \lambda_i \frac{\hat{f}_{\beta, \bar{x}}^*(g_i)}{1 + \sqrt{1 + \|g_i\|_2^2}} + \frac{1}{2t} \|z\|_2^2 : z = \sum_{i \in \beta} \lambda_i \frac{g_i}{1 + \sqrt{1 + \|g_i\|_2^2}} \right\},$$

cioè:

$$\min_{\lambda \in \Delta_{|\beta|}} \left\{ \sum_{i \in \beta} \lambda_i \frac{\hat{f}_{\beta, \bar{x}}^*(g_i)}{1 + \sqrt{1 + \|g_i\|_2^2}} + \frac{1}{2t} \left\| \sum_{i \in \beta} \lambda_i \frac{g_i}{1 + \sqrt{1 + \|g_i\|_2^2}} \right\|_2^2 \right\}.$$

Questo è esattamente il duale proposto in [7] ($\alpha_i = \hat{f}_{\beta, \bar{x}}^*(g_i)$).

Riassumendo, per l'algoritmo dei piani di taglio con i centri di Chebychev, la forma astratta del primale è

$$(\Pi_{\bar{x}, t}) \quad \min \left\{ \sup_d \left\{ \frac{f_{\bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_\epsilon f_{\bar{x}}(d) \right\} + \frac{t}{2} \|d\|_2^2 \right\},$$

la forma concreta del primale è

$$(\Pi_{\beta, \bar{x}, t}) \quad \min \left\{ \sup_d \left\{ \frac{f_{\beta, \bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_\epsilon f_{\beta, \bar{x}}(d) \right\} + \frac{t}{2} \|d\|_2^2 \right\},$$

e la forma implementabile del primale è

$$(\hat{\Pi}_{\beta, \bar{x}, t}) \quad \min \left\{ \sup_d \left\{ \frac{\hat{f}_{\beta, \bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \|g\|_2^2}} : g \in \partial_\epsilon \hat{f}_{\beta, \bar{x}}(d) \right\} + \frac{t}{2} \|d\|_2^2 \right\},$$

la forma astratta del duale è

$$(\Delta_{\bar{x},t}) \quad \min_z \left\{ \inf_{W \in V_{n+1}} \left\{ \inf_{\lambda \in \Delta_{|W|}} \left\{ \sum_{g \in W} \lambda_g \frac{f_{\bar{x}}^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} : \right. \right. \right. \\ \left. \left. \left. \sum_{g \in R^n} \lambda_g \frac{g}{1 + \sqrt{1 + \|g\|_2^2}} = z \right\} \right\} + \frac{1}{2t} \|z\|_2^2 \right\},$$

la forma concreta del duale è

$$(\Delta_{\beta, \bar{x}, t}) \quad \min_z \left\{ \inf_{\lambda \in \Delta_{|\beta|}} \left\{ \sum_{g \in \beta} \lambda_g \frac{f_{\beta, \bar{x}}^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} : \right. \right. \\ \left. \left. \sum_{g \in \beta} \lambda_g \frac{g}{1 + \sqrt{1 + \|g\|_2^2}} = z \right\} + \frac{1}{2t} \|z\|_2^2 \right\},$$

e la forma implementabile del duale è

$$(\hat{\Delta}_{\beta, \bar{x}, t}) \quad \min_z \left\{ \inf_{\lambda \in \Delta_{|\beta|}} \left\{ \sum_{g \in \beta} \lambda_g \frac{f_{\beta, \bar{x}}^*(g)}{1 + \sqrt{1 + \|g\|_2^2}} : \right. \right. \\ \left. \left. \sum_{g \in \beta} \lambda_g \frac{g}{1 + \sqrt{1 + \|g\|_2^2}} = z \right\} + \frac{1}{2t} \|z\|_2^2 \right\}.$$

3.5 Dual Bundle

Per poter scrivere le sei versioni dell'algoritmo, dobbiamo dare anche un *interpretazione primale* del Dual bundle.

Prendiamo di nuovo la forma astratta del metodo Dual bundle (2.44) con $x = \bar{x}$.

Naturalmente, per poter usare (A.14) dobbiamo riscrivere il dual bundle in una forma piú appropriata. Una possibilitá è di esprimere $\partial_\epsilon f(\bar{x})$ via la coniugata di f .

Usando la proprietá (v) della coniugata abbiamo:

$$\min_g \{ \|g\|_2 : g \in \partial_\epsilon f(\bar{x}) \} \\ \equiv \min_g \{ \|g\|_2 : f^*(g) - g\bar{x} \leq \epsilon - f(\bar{x}) \}$$

oppure, semplicemente

$$\min_g \{ \|g\|_2 : g \in \partial_\epsilon f(\bar{x}) \} = \min_g \{ \|g\|_2 + I_{\partial_\epsilon f(\bar{x})}(g) \}.$$

Dalla definizione (A.1.16) e dalla relazione (A.6) si ha che

$$\min_g \{ \|g\|_2 + I_{\partial_\epsilon f(\bar{x})}(g) \} = \min_g \{ \|g\|_2 + (f'_\epsilon(\bar{x}, \cdot))^*(g) \}. \quad (3.28)$$

Il duale di Fenchel di (3.28) è

$$\min_d \{ \|d\|_2 + f'_\epsilon(\bar{x}, \cdot)(d) \}. \quad (3.29)$$

(3.29) è equivalente a

$$\min_{d, \gamma \geq 0} \{ \|d\|_2 + \sigma_{\text{epi } f_{\bar{x}}^*}(d, -\gamma) + \gamma\epsilon \}, \quad (3.30)$$

([4], Ch. XI, Sec. 2.1), dove $f_{\bar{x}}^*(g)$ è data da (2.19).

(3.30) è equivalente a

$$\min_{\gamma \geq 0, d} \{ \gamma [f(\frac{1}{\gamma}d + \bar{x}) - f(\bar{x})] + \gamma\epsilon + \|d\|_2 \} \quad (3.31)$$

([4], Ch. XI, Sec. 1.3). Quindi, il primale astratto del Dual bundle è

$$(\Pi_{\bar{x}, \epsilon}) \quad \min_{\gamma \geq 0, d} \{ \gamma f_{\bar{x}}(\frac{1}{\gamma}d) + \gamma\epsilon + \|d\|_2 \}$$

il primale concreto è

$$(\Pi_{\beta, \bar{x}, \epsilon}) \quad \min_{\gamma \geq 0, d} \{ \gamma f_{\beta, \bar{x}}(\frac{1}{\gamma}d) + \gamma\epsilon + \|d\|_2 \}$$

e il primale implementabile è

$$(\hat{\Pi}_{\beta, \bar{x}, \epsilon}) \quad \min_{\gamma \geq 0, d} \{ \gamma \hat{f}_{\beta, \bar{x}}(\frac{1}{\gamma}d) + \gamma\epsilon + \|d\|_2 \}$$

Questo è equivalente al duale quadratico di dual bundle implementabile. Infatti, consideriamo la forma implementabile (2.46) del Dual bundle. (2.46) si può scrivere, equivalentemente, come:

$$\min_{\lambda, d} \{ \frac{1}{2} \| -d \|^2 : -\alpha\lambda \geq -\epsilon, e\lambda = 1, d = -G\lambda, \lambda \geq 0 \} \quad (3.32)$$

(3.32) è un problema di programmazione quadratica (QP).

Scriviamo il duale quadratico di (3.32):

-poniamo $Q = I, d = x, \lambda = y, q = c = 0$ e

$$A = \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix}$$

e

$$B = \begin{bmatrix} -\alpha \\ e \\ G \end{bmatrix}$$

e

$$b = \begin{bmatrix} -\epsilon \\ 1 \\ 0 \end{bmatrix}$$

e

$$\lambda = \begin{bmatrix} \sigma \\ v \\ z \end{bmatrix}.$$

Allora il duale quadratico di (3.32) è:

$$\max_{\sigma, v, z, \gamma} \left\{ -\frac{1}{2} \| -\gamma \|^2 - \sigma\epsilon + v : -\sigma\alpha + ve + zG \leq 0, z + \gamma = 0, \sigma \geq 0 \right\}, \quad (3.33)$$

o, equivalentemente

$$\begin{aligned} & \max_{\sigma, v, z} \left\{ -\frac{1}{2} \|z\|_2^2 - \sigma\epsilon + v : -\sigma\alpha + ve + zG \leq 0, \sigma \geq 0 \right\} \\ &= \min_{\sigma, v, z} \left\{ \frac{1}{2} \|z\|_2^2 + \sigma\epsilon - v : -\sigma\alpha + zG \leq -ve, \sigma \geq 0 \right\} \\ &= \min_{\sigma, z} \left\{ \frac{1}{2} \|z\|_2^2 + \sigma\epsilon + \max_{i \in \beta} \{g_i z - \sigma\alpha_i\} : \sigma \geq 0 \right\} \\ &= \min_{\sigma, z} \left\{ \frac{1}{2} \|z\|_2^2 + \sigma\epsilon + \sigma \max_{i \in \beta} \left\{ g_i \frac{z}{\sigma} - \alpha_i \right\} : \sigma \geq 0 \right\} \\ &= \min_{\sigma, z} \left\{ \frac{1}{2} \|z\|_2^2 + \sigma\epsilon + \sigma \hat{f}_{\beta, \bar{x}} \left(\frac{z}{\sigma} \right) \right\}, \end{aligned}$$

il che è equivalente a $(\hat{\Pi}_{\beta, \bar{x}, \epsilon})$.

Infine, osserviamo che

$$\begin{aligned} & \min_g \{ \|g\|_2 : g \in \partial_\epsilon f(\bar{x}) \} \\ &\equiv \min_g \{ \|g\|_2 : f^*(g) - g\bar{x} \leq \epsilon - f(\bar{x}) \} \\ &\equiv \min_g \{ \|g\|_2 : f_{\bar{x}}^*(g) \leq \epsilon \}. \end{aligned} \quad (3.34)$$

È quindi possibile scrivere Dual bundle come

$$\min_g \{ \| -g \|_2 + I_{S_\epsilon(f_{\bar{x}}^*)}(g) \}, \quad (3.35)$$

cioé come

$$\min_g \{ \| -g \|_2 + \nu_\epsilon(f_{\bar{x}}^*(g)) \}, \quad (3.36)$$

dove $\nu_\epsilon(f_{\bar{x}}^*(g)) = I_{S_\epsilon(f_{\bar{x}}^*)}(g)$.

Il suo duale è allora

$$\min_d \{ \|d\|_2 + (I_{S_\epsilon(f_{\bar{x}}^*)})^*(d) \}, \quad (3.37)$$

Riassumendo, per il Dual bundle, la forma astratta del primale è

$$(\Pi_{\bar{x},\epsilon}) \quad \min_{\gamma \geq 0, d} \{ \gamma f_{\bar{x}}(\frac{1}{\gamma}d) + \gamma\epsilon + \|d\|_2 \},$$

oppure

$$(\Pi_{\bar{x},\epsilon}) \quad \min_d \{ \|d\|_2 + (I_{S_\epsilon(f_{\bar{x}}^*)})^*(d) \},$$

la forma concreta del primale è

$$(\Pi_{\beta,\bar{x},\epsilon}) \quad \min_{\gamma \geq 0, d} \{ \gamma f_{\beta,\bar{x}}(\frac{1}{\gamma}d) + \gamma\epsilon + \|d\|_2 \},$$

oppure

$$(\Pi_{\beta,\bar{x},\epsilon}) \quad \min_d \{ \|d\|_2 + (I_{S_\epsilon(\hat{f}_{\beta,\bar{x}}^*)})^*(d) \},$$

e la forma implementabile del primale è

$$(\hat{\Pi}_{\beta,\bar{x},\epsilon}) \quad \min_{\gamma \geq 0, d} \{ \gamma \hat{f}_{\beta,\bar{x}}(\frac{1}{\gamma}d) + \gamma\epsilon + \|d\|_2 \},$$

oppure

$$(\hat{\Pi}_{\beta,\bar{x},\epsilon}) \quad \min_d \{ \|d\|_2 + (I_{S_\epsilon(\hat{f}_{\beta,\bar{x}}^*)})^*(d) \},$$

la forma astratta del duale è

$$(\Delta_{\bar{x},\epsilon}) \quad \min \{ \|g\|_2 + I_{S_\epsilon(f_{\bar{x}}^*)}(g) \},$$

la forma concreta del duale è

$$(\Delta_{\beta,\bar{x},\epsilon}) \quad \min \{ \|g\|_2 + I_{S_\epsilon(\hat{f}_{\beta,\bar{x}}^*)}(g) \},$$

e la forma implementabile è

$$(\hat{\Delta}_{\beta,\bar{x},\epsilon}) \quad \min_{\lambda} \{ \frac{1}{2} \| \sum g_i \lambda_i \|_2^2 : \sum \alpha_i \lambda_i \leq \epsilon, \lambda \in \Delta \}.$$

Data la relazione (2.9), la forma implementabile del Dual bundle si può riscrivere come

$$(\hat{\Delta}_{\beta,\bar{x},\epsilon}) \quad \min_g \{ \frac{1}{2} \|g\|_2^2 : \hat{f}_{\beta,\bar{x}}^*(g) \leq \epsilon \}, \quad (3.38)$$

oppure come

$$(\hat{\Delta}_{\beta,\bar{x},\epsilon}) \quad \min_g \{ \frac{1}{2} \|g\|_2^2 + I_{S_\epsilon(\hat{f}_{\beta,\bar{x}}^*)}(g) \}. \quad (3.39)$$

3.6 Separating plane algorithms

Abbiamo visto, in [6], che il problema duale (2.50) può essere riscritto come

$$\min_g \{\psi(f^*(g)) + \|g\|_2^2\}, \quad (3.40)$$

oppure come

$$\min_g \{\psi(f_{\bar{x}}^*(g)) + \|g\|_2^2\}. \quad (3.41)$$

Il problema primale (2.51) (duale di (2.50)) è equivalente al duale di Fenchel di (3.40) e di (3.41).

Infatti, il duale di Fenchel di (3.41) è

$$\min_d \{(\psi \circ f_{\bar{x}}^*)^*(d) + \|d\|_2^2\}, \quad (3.42)$$

il che, per il Teorema A.1.2, è equivalente a

$$\min_{d, \gamma \geq 0} \left\{ \gamma f_{\bar{x}}\left(\frac{d}{\gamma}\right) + \psi^*(\gamma) + \|d\|_2^2 \right\}.$$

Usando la definizione (A.1.6) della coniugata, questo si può riscrivere come

$$\begin{aligned} & \min_{d, \gamma \geq 0} \left\{ \sup_g \left\{ \gamma \frac{d}{\gamma} g - \gamma f_{\bar{x}}^*(g) \right\} + \psi^*(\gamma) + \|d\|_2^2 \right\} \\ &= \min_{d, v, \gamma} \left\{ v + \sup_{\omega} \{ \omega \gamma - \psi(\omega) \} + \|d\|_2^2 : v \geq dg - \gamma f_{\bar{x}}^*(g) \right\}. \end{aligned} \quad (3.43)$$

(3.43) è equivalente a (2.51). Infatti, se avessimo calcolato il duale direttamente da (3.40), sarebbe venuto esattamente (2.51). Ma (3.40) è equivalente a (3.41) e allora anche (3.43) è equivalente a (2.51) (vedere la dualità di Fenchel in (A.1.21)).

Riassumendo, per i Separating plane algorithms, la forma astratta del primale è

$$(\Pi_{\bar{x}, t}) \quad \min_d \{(\nu(f_{\bar{x}}^*))^*(d) + \|d\|_2^2\},$$

la forma concreta del primale è

$$(\Pi_{\beta, \bar{x}, t}) \quad \min_d \{(\nu(f_{\beta, \bar{x}}^*))^*(d) + \|d\|_2^2\},$$

la forma implementabile del primale è

$$(\hat{\Pi}_{\beta, \bar{x}, t}) \quad \min_d \{(\nu(\hat{f}_{\beta, \bar{x}}))^*(d) + \|d\|_2^2\},$$

oppure la forma astratta del duale è

$$(\Delta_{\bar{x},t}) \quad \min_g \{\nu(f_{\bar{x}}^*(g)) + \|g\|_2^2\},$$

la forma concreta del duale è

$$(\Delta_{\beta,\bar{x},t}) \quad \min_g \{\nu(f_{\beta,\bar{x}}^*(g)) + \|g\|_2^2\},$$

e la forma implementabile del duale è

$$(\hat{\Delta}_{\beta,\bar{x},t}) \quad \min_g \{\nu(\hat{f}_{\beta,\bar{x}}^*(g)) + \|g\|_2^2\},$$

dove $\nu = \psi$ di (2.49).

3.7 Tabelle riassuntive delle forme astratte

forme astratte	$(\Pi_{\bar{x},t})$
Piani di taglio	$\min_d \{f_{\bar{x}}(d)\}$
Bundle generalizzato	$\inf_d \{\nu(f_{\bar{x}}(d)) + D_t(d)\}$, con $\nu(f_{\bar{x}}(d)) = Id(f_{\bar{x}}(d))$
Level bundle	$\min_d \{\nu(f_{\bar{x}}(d)) + \frac{1}{2}\ d\ _2^2\}$, con $\nu(f_{\bar{x}}(d)) = I_{S_t(f_{\bar{x}})}(d)$
Chebychev	$\min_d \{ \psi_{\bar{x}}(d) + \frac{t}{2}\ d\ _2^2 \}$, con $\psi_{\bar{x}}(d) = \sup_{g,\epsilon} \left\{ \frac{f_{\bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \ g\ _2^2}} : g \in \partial_\epsilon f_{\bar{x}}(d) \right\}$
Dual bundle	$\min_d \{(\nu(f_{\bar{x}}^*))^*(d) + \ d\ _2\}$, con $(\nu(f_{\bar{x}}^*))^*(d) = (I_{S_\epsilon(f_{\bar{x}}^*)})^*(d)$
Separating plane	$\min_d \{(\nu(f_{\bar{x}}^*))^*(d) + \ d\ _2^2\}$, con $\nu = \psi$ definita da (2.49)

forme astratte	$(\Delta_{\bar{x},t})$
Piani di taglio	$\min_g \{f_{\bar{x}}^*(g) : g = 0\}$
Bundle generalizzato	$\inf_g \{(\nu(f_{\bar{x}}))^*(g) + D_t^*(-g)\}$, con $(\nu(f_{\bar{x}}))^*(g) = (Id(f_{\bar{x}}))^*(g) = f_{\bar{x}}^*(g)$
Level bundle	$\min_g \{(\nu(f_{\bar{x}}))^*(g) + \frac{1}{2}\ g\ _2^2\}$, con $(\nu(f_{\bar{x}}))^*(g) = (I_{S_t(f_{\bar{x}})})^*(g)$
Chebychev	$\min_z \{ \inf_{W \in V_{n+1}} \{ \inf_{\lambda \in \Delta_{ W }} \{ \sum_{g \in W} \lambda_g \frac{f_{\bar{x}}^*(g)}{1 + \sqrt{1 + \ g\ _2^2}} : \sum_{g \in R^n} \lambda_g \frac{g}{1 + \sqrt{1 + \ g\ _2^2}} = z \} \} + \frac{1}{2t} \ z\ _2^2 \}$
Dual bundle	$\min_g \{ \nu(f_{\bar{x}}^*(g)) + \ g\ _2 \}$, con $\nu(f_{\bar{x}}^*(g)) = I_{S_\epsilon(f_{\bar{x}})}(g)$
Separating plane	$\min_g \{ \nu(f_{\bar{x}}^*(g)) + \ g\ _2^2 \}$, con $\nu = \psi$ definita da (2.49)

3.8 Tabelle riassuntive delle forme concrete

forme concrete	$(\Pi_{\beta, \bar{x}, t})$
Piani di taglio	$\min_d \{f_{\beta, \bar{x}}(d)\}$
Bundle generalizzato	$\inf_d \{\nu(f_{\beta, \bar{x}}(d)) + D_t(d)\}, \text{ con } \nu(f_{\beta, \bar{x}}(d)) = Id(f_{\beta, \bar{x}}(d))$
Level bundle	$\min_d \{\nu(f_{\beta, \bar{x}}(d)) + \frac{1}{2}\ d\ _2^2\}, \text{ con } \nu(f_{\beta, \bar{x}}(d)) = I_{S_t(f_{\beta, \bar{x}})}(d)$
Chebychev	$\min_d \{ \psi_{\beta, \bar{x}}(d) + \frac{t}{2}\ d\ _2^2 \}$ con $\psi_{\beta, \bar{x}}(d) = \sup_{g, \epsilon} \left\{ \frac{f_{\beta, \bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \ g\ _2^2}} : g \in \partial_\epsilon f_{\beta, \bar{x}}(d) \right\}$
Dual bundle	$\min_d \{(\nu(f_{\beta, \bar{x}}^*))^*(d) + \ d\ _2\}, \text{ con } (\nu(f_{\beta, \bar{x}}^*))^*(d) = (I_{S_\epsilon(f_{\beta, \bar{x}}^*)})^*(d)\}$
Separating plane	$\min_d \{(\nu(f_{\beta, \bar{x}}^*))^*(d) + \ d\ _2^2\}, \text{ con } \nu = \psi \text{ definita da (2.49)}$

forme concrete	$(\Delta_{\beta, \bar{x}, t})$
Piani di taglio	$\min_g \{f_{\beta, \bar{x}}^*(g) : g = 0\}$
Bundle generalizzato	$\inf_g \{(\nu(f_{\beta, \bar{x}}))^*(g) + D_t^*(-g)\}, \text{ con } (\nu(f_{\beta, \bar{x}}))^*(g) = (Id(f_{\beta, \bar{x}}))^*(g) = f_{\beta, \bar{x}}^*(g)$
Level bundle	$\min_g \{(\nu(f_{\beta, \bar{x}}))^*(g) + \frac{1}{2}\ g\ _2^2\}, \text{ con } (\nu(f_{\beta, \bar{x}}))^*(g) = (I_{S_t(f_{\beta, \bar{x}})})^*(g)$
Chebychev	$\min_z \{ \inf_{\lambda \in \Delta_{ \beta }} \{ \sum_{g \in \beta} \lambda_g \frac{f_{\beta, \bar{x}}^*(g)}{1 + \sqrt{1 + \ g\ _2^2}} : \sum_{g \in \beta} \lambda_g \frac{g}{1 + \sqrt{1 + \ g\ _2^2}} = z \} + \frac{1}{2t} \ z\ _2^2 \}$
Dual bundle	$\min_g \{ \nu(f_{\beta, \bar{x}}^*(g)) + \ g\ _2 \}, \text{ con } \nu(f_{\beta, \bar{x}}^*(g)) = I_{S_\epsilon(f_{\beta, \bar{x}})}(g)$
Separating plane	$\min_g \{ \nu(f_{\beta, \bar{x}}^*(g)) + \ g\ _2^2 \}, \text{ con } \nu = \psi \text{ definita da (2.49)}$

3.9 Tabelle riassuntive delle forme implementabili

forme implementabili	$(\hat{\Pi}_{\beta, \bar{x}, t})$
Piani di taglio	$\min_d \{ \hat{f}_{\beta, \bar{x}}(d) \}$
Bundle generalizzato	$\inf_d \{ \nu(\hat{f}_{\beta, \bar{x}}(d)) + D_t(d) \}, \text{ con } \nu(\hat{f}_{\beta, \bar{x}}(d)) = Id(\hat{f}_{\beta, \bar{x}}(d))$
Level bundle	$\min_d \{ \nu(\hat{f}_{\beta, \bar{x}}(d)) + \frac{1}{2} \ d\ _2^2 \}, \text{ con } \nu(\hat{f}_{\beta, \bar{x}}(d)) = I_{S_t(\hat{f}_{\beta, \bar{x}})}(d)$
Chebychev	$\min_d \{ \hat{\psi}_{\beta, \bar{x}}(d) + \frac{t}{2} \ d\ _2^2 \}$ con $\hat{\psi}_{\beta, \bar{x}}(d) = \sup_{g, \epsilon} \left\{ \frac{\hat{f}_{\beta, \bar{x}}(d) - \epsilon}{1 + \sqrt{1 + \ g\ _2^2}} : g \in \partial_\epsilon \hat{f}_{\beta, \bar{x}}(d) \right\}$
Dual bundle	$\min_d \{ (\nu(\hat{f}_{\beta, \bar{x}}^*))^*(d) + \ d\ _2 \}, \text{ con } (\nu(\hat{f}_{\beta, \bar{x}}^*))^*(d) = (I_{S_\epsilon(\hat{f}_{\beta, \bar{x}}^*)})^*(d)$
Separating plane	$\min_d \{ (\nu(\hat{f}_{\beta, \bar{x}}^*))^*(d) + \ d\ _2^2 \}, \text{ con } \nu = \psi \text{ definita da (2.49)}$

forme implementabili	$(\hat{\Delta}_{\beta, \bar{x}, t})$
Piani di taglio	$\min_g \{ \hat{f}_{\beta}^*(g) : g = 0 \}$
Bundle generalizzato	$\inf_g \{ (\nu(\hat{f}_{\beta, \bar{x}}))^*(g) + D_t^*(-g) \}$, con $(\nu(\hat{f}_{\beta, \bar{x}}))^*(g) = (Id(\hat{f}_{\beta, \bar{x}}))^*(g) = \hat{f}_{\beta, \bar{x}}^*(g)$
Level bundle	$\min_g \{ (\nu(\hat{f}_{\beta, \bar{x}}))^*(g) + \frac{1}{2} \ g\ _2^2 \}$ con $(\nu(\hat{f}_{\beta, \bar{x}}))^*(g) = (I_{S_t(\hat{f}_{\beta, \bar{x}})})^*(g)$
Chebychev	$\min_z \{ \inf_{\lambda \in \Delta_{ \beta }} \{ \sum_{g \in \beta} \lambda_g \frac{\hat{f}_{\beta, \bar{x}}^*(g)}{1 + \sqrt{1 + \ g\ _2^2}} : \sum_{g \in \beta} \lambda_g \frac{g}{1 + \sqrt{1 + \ g\ _2^2}} = z \} + \frac{1}{2t} \ z\ _2^2 \}$
Dual bundle	$\min_g \{ \nu(\hat{f}_{\beta, \bar{x}}^*(g)) + \ g\ _2 \}$, con $\nu(\hat{f}_{\beta, \bar{x}}^*(g)) = I_{S_{\epsilon}(\hat{f}_{\beta, \bar{x}}^*)}(g)$
Separating plane	$\min_g \{ \nu(\hat{f}_{\beta, \bar{x}}^*(g)) + \ g\ _2^2 \}$, con $\nu = \psi$ definita da (2.49)

Capitolo 4

Proprietá dei termini di stabilizzazione

In questo capitolo cerchiamo di capire quale insieme di proprietá descrive contemporaneamente tutti i termini di stabilizzazione D_t e quale insieme descrive tutti i termini di stabilizzazione ν . Questo perché un D_t [risp. ν] che gode di tale insieme di proprietá è una generalizzazione di tutti D_t [risp. ν] che consideriamo. Dopo cercheremo di dimostrare che, sotto alcune di queste proprietá, il problema generalizzato è equivalente al problema originale (II), nel senso che hanno la stessa soluzione ottima e lo stesso valore ottimo.

4.1 Termine di stabilizzazione D_t

In questa sezione parleremo solo del termine di stabilizzazione D_t e ometteremo la stabilizzazione ν della funzione obiettivo.

Quando noi aggiungiamo al nostro modello, che approssima f , il termine di stabilizzazione D_t , il problema che risolviamo è chiaramente cambiato. Quindi quando cerchiamo di capire chi è la soluzione ottima del problema stabilizzato, dobbiamo prendere in considerazione anche l'influenza della stabilizzazione D_t . Faremo vedere sotto quali ipotesi su D_t , $f_{\bar{x}} + D_t$ ha il minimo uguale a f_* , dove $f_* = \inf f$.

Abbiamo già detto, nel Capitolo 1 che risolvere il problema (II)

$$\min_x \{ f(x) \}$$

è equivalente a risolvere

$$\min_d \{ f_{\bar{x}}(d) \}.$$

Se $d^* = 0$, allora \bar{x} è la soluzione ottima di (II) (infatti, in questo caso, qualunque spostamento lungo la miglior direzione di decrescita d^* non

diminuirebbe il valore della funzione obiettivo perché $f(\bar{x}_+) = f(\bar{x})$, cioè il minimo è stato raggiunto su $\bar{x}_+ = \bar{x}$. Oppure, detto in termini di subdifferenziale, $d^* = 0$ significa $0 \in \partial f_{\bar{x}}(0) \Rightarrow 0 \in \partial f(\bar{x})$ (per la proprietà (I) dell' ϵ -subdifferenziale) $\Leftrightarrow_{(A.4)} \bar{x}$ minimizza f .

Aggiungendo D_t si ha il problema

$$\min_d \{ f_{\bar{x}}(d) + D_t(d) \},$$

di cui la soluzione ottima è

$$d^* = \arg \min_d \{ f_{\bar{x}}(d) + D_t(d) \}.$$

Anche in questo caso dovrebbe essere $d^* = 0 \Rightarrow \bar{x}$ ottima per (II), cioè $0 \in \partial[f_{\bar{x}}(0) + D_t(0)]$. Il [[4], Ch. XI, Corollario 3.1.2] ci dice che, se $\text{ridom} f_{\bar{x}} \cap \text{ridom} D_t \neq \emptyset$, allora $0 \in \partial[f_{\bar{x}}(0) + D_t(0)] \equiv 0 \in [\partial f_{\bar{x}}(0) + \partial D_t(0)]$.

La condizione $\text{ridom} f_{\bar{x}} \cap \text{ridom} D_t \neq \emptyset$ è vera. Infatti, [3], Ch. III, Proposizione 2.1.10. ci dice che

$$\text{ridom} f_{\bar{x}} \cap \text{ridom} D_t = \text{ri}(\text{dom} f_{\bar{x}} \cap \text{dom} D_t).$$

R^n è il dominio, sia di f , che di D_t e quindi

$$\text{dom} f_{\bar{x}} \cap \text{dom} D_t \neq \emptyset,$$

e questo implica, per [3], Ch. III, Teorema 2.1.3., che

$$\text{ri}(\text{dom} f_{\bar{x}} \cap \text{dom} D_t) \neq \emptyset.$$

Naturalmente $0 \in [\partial f_{\bar{x}}(0) + \partial D_t(0)]$ non implica necessariamente $0 \in \partial f(\bar{x})$. Tutto dipende da come è fatto $\partial D_t(0)$. In questo senso, possiamo distinguere tra il caso “good” e il caso “bad”.

Il caso “good” è quando

$$\partial f_{\bar{x}}(0) + \partial D_t(0) = \partial f_{\bar{x}}(0), \quad (4.1)$$

perché allora \bar{x} è ottima sia per il problema stabilizzato che per il problema non stabilizzato.

Significa che una prima proprietà di D_t è la seguente proprietà che rende valida la relazione (4.1):

$$(P1_{D_t}) \quad \forall t \quad \partial D_t(0) = \{0\} \quad (D_t \text{ è differenziabile in } 0, \text{ cioè } \nabla D_t(0) = 0).$$

Questo è esattamente la proprietà (P5') di [1].

Il discorso analogo si può fare per il duale di (II) - (Δ).
 (Δ) è equivalente a

$$\begin{aligned} & \min_g \{f^*(g) : g = 0\} \\ \equiv & \min_g \{f_{\bar{x}}^*(g) : g = 0\} \end{aligned}$$

(è il duale astratto dell'algoritmo dei piani di taglio; vedere Sezione 2.1).
 È chiaro che

$$g^* = \arg \min_g \{f_{\bar{x}}^*(g) : g = 0\}$$

non può che essere 0 e si ha, per la relazione (A.3) che se $g^* = 0 \Rightarrow \bar{x}$ ottima per (II).

Aggiungendo la stabilizzazione,

$$g^* = \arg \min_g \{f_{\bar{x}}^*(g) + D_t^*(-g)\},$$

e risolvere all'ottimo questo problema significa non avere la penalità cioè che $D_t^*(g^*) = 0$ e ottenere la stessa soluzione ottima del duale non stabilizzato $g^* = 0 (\Rightarrow \bar{x}$ ottima per (II)). Ma non sempre succedono entrambe cose. Si ha che, allora, il caso "good" è quando

$$g^* = 0 \Leftrightarrow D_t^*(g^*) = 0. \quad (4.2)$$

Significa che una prima proprietà di D_t^* è la seguente proprietà che rende valida la relazione (4.2):

($P1_{D_t^*}$) 0 è l'unico minimo di D_t^* .

Questo è equivalente alla proprietà (P^*5') di [1].

Proposizione 4.1.1 (4.1) nel primale ha la proprietà equivalente nel duale che è (4.2) (proprio come ($P5'$) è equivalente a (P^*5')) e se vale una, vale anche l'altra. Cioè un caso è "good" sia per il primale che per il duale stabilizzati, oppure è per entrambi un caso "bad".

DIMOSTRAZIONE. Da [3], Ch. I, Sezione 6.2. abbiamo il seguente risultato:

$$\begin{aligned} & D_t \text{ è differenziabile in } 0 \\ \equiv & \text{ esiste una soluzione unica (in } g) \text{ di } 0 \in \partial D_t^*(g) \\ \equiv & \text{ esiste un unico minimo di } D_t^* \text{ ed è } g \end{aligned}$$

Dire che D_t è tale che $0 \in \partial D_t(0)$ equivale a dire che $0 \in \partial D_t^*(0)$ (proprietá (iv) della coniugata), dove il secondo 0 è proprio la unica soluzione di $0 \in \partial D_t^*(g)$ di cui sopra. Dunque, 0 è l'unico minimo di D_t^* .

Partiamo ora dal caso “good” nel duale. Se 0 è l'unico minimo di D_t^* , possiamo concludere che D_t è differenziabile in 0 e che $0 \in \partial D_t(0)$ il che è equivalente a $0 \in \partial D_t(0)$. Abbiamo finito perché D_t differenziabile in 0, significa che $\partial D_t(0)$ è singleton. Questo, insieme al fatto che $0 \in \partial D_t(0)$, significa che $\partial D_t(0) = \{0\} \diamond$.

La relazione (4.2) diventa ancora piú chiara se consideriamo cosa succede quando occorre un caso “bad” (cioé quando non vale). Allora il duale stabilizzato ha ancora il minimo quando $D_t^*(g^*) = 0$. Poiché (4.2) non vale, questo succede in uno o piú subgradienti g^* , che possono benissimo essere diversi da 0. Un tale g^* non può essere la soluzione di (Δ) .

Tutti gli algoritmi descritti nelle tabelle del capitolo precedente, a parte gli algoritmi di [7] e di [1], hanno il D_t e il D_t^* della forma $c\|\cdot\|_2$, oppure $c\|\cdot\|_2^2$, dove c è una costante. In [7], D_t e D_t^* sono della forma $c(t)\|\cdot\|_2^2$, con $c(t) : R_+ \rightarrow R_+$.

$c\|g\|_2$, $c\|g\|_2^2$, $c(t)\|g\|_2^2$ rispettano la $(P1_{D_t^*}) \forall t, c$.

$c\|d\|_2$, $c\|d\|_2^2$, $c(t)\|d\|_2^2$ rispettano la $(P1_{D_t}) \forall t, c$.

Quindi siamo sempre nel caso “good” per quanto riguarda D_t e D_t^* .

In [1] per avere l'algoritmo Bundle generalizzato è stata fatta una generalizzazione di D_t e D_t^* Sono elencate in ([1], Sezione 3.) (e riportate qui, nella Sezione 2.3) le proprietá che deve avere questo termine di stabilizzazione (cioé il termine di penalitá nel duale), ed è stata proposta una dimostrazione di convergenza sotto diverse tra queste proprietá. A differenza degli altri algoritmi descritti nelle tabelle del capitolo precedente, in [1] sono stati trattati sia il caso “good” che il caso “bad” di D_t e D_t^* . È stato fatto uno studio piú esauriente rispetto agli altri sulla questione di D_t e D_t^* , e lo si vede dal fatto che il Bundle generalizzato è una generalizzazione anche di penalty e trust-region bundle (Sezione 2.2), dove D_t e D_t^* sono $c(t)\|\cdot\|_2^2$ per il penalty e $c\|\cdot\|_2$ per il trust-region bundle.

Allora possiamo anche qui generalizzare i termini di stabilizzazione (penalitá) $c(t)\|\cdot\|_2^2$, $c\|\cdot\|_2^2$, $c\|\cdot\|_2$ con i D_t e D_t^* descritti dalle proprietá di ([1], Sezione 3.).

4.2 Termine di stabilizzazione della funzione obiettivo

Dalle tabelle del capitolo precedente, vediamo che gli algoritmi, a parte l'algoritmo dei piani di taglio con i centri di Chebychev, hanno due tipi di stabilizzazione della funzione obiettivo.

(St1) Un primo tipo ha una stabilizzazione nel primale del tipo

$$\min_d \{ \nu(f_{\bar{x}}(d)) + D_t(d) \}, \quad (4.3)$$

cioé, una stabilizzazione nel duale del tipo

$$\min_g \{ (\nu(f_{\bar{x}}))^*(g) + D_t^*(-g) \}, \quad (4.4)$$

oppure, usando la forma estesa della coniugata della composizione di funzioni (Teorema A.1.2)

$$\min_{g, \gamma \geq 0} \{ \gamma f_{\bar{x}}^*\left(\frac{g}{\gamma}\right) + \nu^*(\gamma) + D_t^*(-g) \}. \quad (4.5)$$

Questa stabilizzazione è presente nell'algoritmo Level bundle, con $\nu = I_{\{v \leq t\}}$, nel Bundle generalizzato e, a parte D_t , anche nell'algoritmo dei piani di taglio. In questi ultimi due, $\nu = Id$.

(St2) Un secondo tipo ha una stabilizzazione nel primale del tipo

$$\min_d \{ (\nu(f_{\bar{x}}^*))^*(d) + D_t(d) \}, \quad (4.6)$$

oppure, usando la forma estesa della coniugata della composizione di funzioni (Teorema A.1.2)

$$\min_{d, \gamma \geq 0} \{ f_{\bar{x}}\left(\frac{d}{\gamma}\right) + \nu^*(\gamma) + D_t(d) \}. \quad (4.7)$$

Il duale di (4.6) e (4.7) è

$$\min_g \{ \nu(f_{\bar{x}}^*(g)) + D_t^*(-g) \}. \quad (4.8)$$

Questa stabilizzazione è presente nell'algoritmo Bundle generalizzato, con $\nu = Id$, nell'algoritmo Dual bundle, con $\nu = I_{\{z \leq \epsilon\}}$ e nell'algoritmo Separating plane, con $\nu = \psi$, dove ψ ha le proprietà (2.49).

Per evitare la notazione confusa, chiamiamo ϕ la stabilizzazione ν nel tipo (St2) e riscriviamo (4.6) e (4.8) come

$$\min_d \{ (\phi(f_{\bar{x}}^*))^*(d) + D_t(d) \}, \quad (4.9)$$

e

$$\min_g \{ \phi(f_{\bar{x}}^*(g)) + D_t^*(-g) \}. \quad (4.10)$$

Quindi, anziché parlare dell'algoritmo generalizzato, meglio parlare degli algoritmi generalizzati - infatti, ci sono due generalizzazioni - una corrispondente alla stabilizzazione ν di tipo (St1) e l'altra corrispondente alla stabilizzazione ν di tipo (St2) (cioè alla stabilizzazione ϕ).

4.2.1 Caso "good" della stabilizzazione ν

Analogamente a D_t , possiamo distinguere il caso "good" e il caso "bad" anche per ν . Il caso "good" di D_t significava che all'ottimo questo termine di stabilizzazione non influenzava la soluzione e questo significa che all'ottimo $\partial D_t(0) = \{0\}$ ($P1_{D_t}$), mentre il caso "bad" occorre quando questa proprietà non vale. Anche per ν si può fare lo stesso ragionamento. Se vale la seguente proprietà:

$$(P1_\nu) \quad 0 \in \partial[\nu(f_{\bar{x}})](0) \iff 0 \in \partial f(\bar{x}),$$

allora il problema originale e il problema stabilizzato sono equivalenti, nel senso che hanno la stessa soluzione ottima (per ([4], Ch. XV, Th. 4.1.7)), cioè siamo nel caso "good" di ν . Se $(P1_\nu)$ non vale, siamo nel caso "bad" di ν .

Il seguente insieme di proprietà di ν ci garantisce che vale la $(P1_\nu)$:

$$(P2_\nu) \quad \nu : R \rightarrow R \text{ è una funzione chiusa, convessa e strettamente monotona crescente in } 0 \text{ e tale che } \nu(0) = 0.$$

È utile qui far vedere come è fatto il subdifferenziale delle funzioni strettamente monotoni e crescenti:

Proposizione 4.2.1 *Sia $\nu : R \rightarrow R$ una funzione convessa. Allora,*

- i) ν è strettamente monotona crescente in un punto $y \in \text{intdom } \nu \iff \exists \delta \in R$ tale che $\alpha \geq \delta > 0, \forall \alpha \in \partial \nu(y)$.
- ii) Se in più ν è strettamente monotona crescente ovunque, allora $\forall \alpha, y \exists \delta_* \in R$ tale che $\alpha \geq \delta_* > 0$, dove $\alpha \in \partial \nu(y)$.

DIMOSTRAZIONE.

- i) Dalla ([3], Ch.I, Def.4.1.5) sappiamo che $\partial\nu(y)$ è l'insieme di tutte le subderivate di ν su y . Se la derivata sinistra e derivata destra di ν su y sono finite, allora $\partial\nu(y)$ è la line-segment $[D_-\nu(y), D_+\nu(y)]$, dove $D_-\nu(y)$ è la derivata sinistra e $D_+\nu(y)$ è la derivata destra di ν su y . Ora consideriamo ([3], Ch.I, Th.4.1.1). Questo teorema ci dice che $D_-\nu(y)$ e $D_+\nu(y)$ sono finite se $y \in \text{intdom } \nu$ perche ν è convessa. Inoltre ci dice che $D_-\nu(y) \leq D_+\nu(y)$. Quindi ci basta dimostrare che $D_-\nu(y) > 0$. Sempre dal ([3], Ch.I, Th.4.1.1) prendiamo la definizione della derivata sinistra:

$$D_-\nu(y) = \sup_{x < y} \frac{\nu(x) - \nu(y)}{x - y}. \quad (4.11)$$

Se ν è strettamente monotona crescente in y , allora $x < y \Rightarrow \nu(x) < \nu(y)$ e quindi $D_-\nu(y)$ è la frazione di due quantità strettamente negative ed è, dunque, strettamente positiva. Di conseguenza, tutta $[D_-\nu(y), D_+\nu(y)]$ è strettamente positiva e ogni subderivata nel $\partial\nu(y)$ è strettamente positiva.

- ii) Se ν è strettamente monotona crescente ovunque, allora $D_-\nu(y) > 0$, $\forall y \in \text{intdom } \nu$. Sia

$$\delta_* = \min_y \{D_-\nu(y)\} > 0.$$

Allora, $\forall y' \in \text{intdom } \nu$ e $\forall \alpha' \in \partial\nu(y')$,

$$\alpha' \geq D_-\nu(y') \geq D_-\nu(y) = \delta_* > 0. \diamond$$

A questo punto possiamo dimostrare il seguente lemma:

Lemma 4.2.1 $(P2_\nu) \implies (P1_\nu)$.

DIMOSTRAZIONE. Ricordiamo che

- (1) $f_{\bar{x}}(0) = 0$ (per definizione),
- (2) $\partial f_{\bar{x}}(0) \equiv \partial f(\bar{x})$ (per la definizione di $\partial f_{\bar{x}}(0)$ e la proprietà (I) del subdifferenziale)
- (3) $0 \in \partial\nu(f_{\bar{x}}(0)) \iff 0 \in \partial(\alpha f_{\bar{x}})(0)$, $\alpha \in \partial\nu(f_{\bar{x}}(0))$, $\alpha \geq 0$ (per il Teorema A.1.18).

L'ultima proprietà è l' ϵ -subdifferenziale di una composizione di funzioni quando $\epsilon = 0$ che si ottiene ponendo a 0 $\epsilon, \epsilon_1, \epsilon_2$ nel Teorema A.1.18.

$$\Leftrightarrow 0 \in \partial f(\bar{x}) \Rightarrow_{(2)} 0 \in \partial f_{\bar{x}}(0) \Rightarrow_{(A.1)} \exists \alpha > 0 : 0 \in \partial \alpha f_{\bar{x}}(0).$$

In particolare, possiamo scegliere un $\alpha \in \partial \nu(f_{\bar{x}}(0))$ (perché $\alpha \in \partial \nu(f_{\bar{x}}(0)) \Rightarrow_{(1)} \alpha \in \partial \nu(0)$, e questo, per la Proposizione 4.2.1, implica che $\alpha > 0$).

Quindi abbiamo $0 \in \partial \alpha f_{\bar{x}}(0), \alpha \in \partial \nu(f_{\bar{x}}(0)), \alpha > 0 \Rightarrow_{(3)} 0 \in \partial \nu(f_{\bar{x}}(0))$.

$$\Rightarrow 0 \in \partial \nu(f_{\bar{x}}(0)) \Leftrightarrow_{(3)} \exists \alpha : 0 \in \partial \alpha f_{\bar{x}}(0), \alpha \in \partial \nu(f_{\bar{x}}(0)) \equiv_{(1)} 0 \in \partial \alpha f_{\bar{x}}(0), \alpha \in \partial \nu(0), \text{ e questo, usando ancora una volta la Proposizione 4.2.1, implica che } \alpha > 0.$$

Il fatto che $\alpha > 0$ ci permette di usare la relazione (A.1) per ottenere $0 \in \partial f_{\bar{x}}(0) \Rightarrow_{(2)} 0 \in \partial f(\bar{x})$. \diamond

4.2.2 Caso "good" della stabilizzazione ϕ

Per ([4], Ch. XV, Th. 4.1.7), il problema originale e quello stabilizzato possono essere considerati equivalenti anche nel senso che

$$\inf_d \{ (\phi(f_{\bar{x}}^*))^*(d) \} = \inf_d \{ f_{\bar{x}}(d) \}.$$

Consideriamo questa relazione dal punto di vista duale. In ottica della relazione (A.3), possiamo riscrivere la condizione come:

$$\phi(f_{\bar{x}}^*)(0) = f_{\bar{x}}^*(0). \quad (4.12)$$

Quando \bar{x} è ottima, $-f_{\bar{x}}^*(0) = \inf_d f_{\bar{x}}(d) = f_{\bar{x}}(0) = 0$. Inserendo questo risultato in (4.12), otteniamo la condizione per il caso "good" di ϕ :

$$(P1_{\phi}) \quad \phi(f_{\bar{x}}^*)(0) = 0 \iff \bar{x} \text{ è ottima.}$$

Un caso "bad" occorre quando $(P1_{\phi})$ non vale. Allora possiamo avere un \bar{x} tale che $f_{\bar{x}}^*(0) \neq 0$ e $\phi(f_{\bar{x}}^*)(0) = 0$. Un tale \bar{x} minimizza $\phi(f_{\bar{x}}^*)(0)$, ma non minimizza $f_{\bar{x}}^*(0)$. Questo significa che il problema stabilizzato e quello non stabilizzato non sono equivalenti, perché la condizione (4.12) non è più valida.

Il seguente insieme di proprietà di ϕ ci garantisce che vale la $(P1_{\phi})$:

$$(P2_{\phi}) \quad \phi : R \rightarrow R \text{ è una funzione chiusa, convessa e strettamente monotona crescente in } 0 \text{ e tale che } \phi(0) = 0.$$

A questo punto possiamo dimostrare il seguente lemma:

Lemma 4.2.1 $(P2_\phi) \implies (P1_\phi)$.

DIMOSTRAZIONE. Ricordiamo che

(1) $f_{\bar{x}}(0) = 0$ (per definizione).

\Leftarrow) Come già accenato, quando \bar{x} è ottima, $f_{\bar{x}}^*(0) = 0$. Insieme a $(P2_\phi)$, o meglio, al fatto che $\phi(0) = 0$, si ha $\phi(f_{\bar{x}}^*)(0) = 0$.

\Rightarrow) Come prima cosa, applichiamo la relazione (A.3) a $\phi(f_{\bar{x}}^*)(0)$ per ottenere che

$$\phi(f_{\bar{x}}^*)(0) = 0 \equiv \phi(-\min_d f_{\bar{x}}(d)) = 0.$$

La $(P2_\phi)$ ci dice che $\phi(0) = 0$ e, grazie alla stretta monotonia di ϕ in 0, ogni $y \neq 0$ ci da un valore di $\phi(y) \neq 0$.

Quindi $-\min_d f_{\bar{x}}(d) = 0 \Rightarrow_{(1)} d = 0$. Di conseguenza, $f(\bar{x}_+) = f(\bar{x})$, cioè \bar{x} è ottima.

4.3 Stabilizzazione generalizzata della funzione obiettivo

I tipi di stabilizzazioni (St1) e (St2) possono essere scritte in una forma unificata - possiamo pensare che sia (St1), che (St2) hanno entrambe le funzioni “stabilizzatrici” ν e ϕ , e che nel caso di (St1) $\phi = Id$, mentre nel caso di (St2) $\nu = Id$. In questa sezione spieghiamo esattamente come si arriva alla forma unificata.

Consideriamo il duale del tipo (St1). Per il Teorema A.1.2 lo possiamo riscrivere come

$$\min_{g, \gamma \geq 0} \{ \gamma f_{\bar{x}}^*\left(\frac{g}{\gamma}\right) + \nu^*(\gamma) \}.$$

Dato che consideriamo (St1), $\phi = Id$ e nulla cambia se riscriviamo il problema di sopra come

$$\min_{g, \gamma \geq 0} \{ \gamma \phi\left(f_{\bar{x}}^*\left(\frac{g}{\gamma}\right)\right) + \nu^*(\gamma) \},$$

e questo è equivalente a

$$\min_g [\nu([\phi(f_{\bar{x}}^*(\cdot))]^*)]^*(g) \tag{4.13}$$

(abbiamo di nuovo usato il Teorema A.1.2). Il primale di (4.13) è

$$\min_d \nu([\phi(f_{\bar{x}}^*)]^*)(d), \tag{4.14}$$

ed è equivalente al primale di (St1) quando $\phi = Id$.

Affinché (4.13) e (4.14) siano la forma unificante sia per (St1), che per (St2), bisogna considerare il caso $\nu = Id$. Con $\nu = Id$ in (4.13) otteniamo il duale e in (4.14) il primale di (St2).

In altre parole, (4.13) e (4.14) rappresentano una stabilizzazione sul valore della funzione obiettivo, che chiamiamo (Un1) e che è una generalizzazione di stabilizzazioni di tipo (St1) e di tipo (St2) (cioé (St1) e (St2) sono casi particolari di (Un1) dove, rispettivamente $\phi = Id$ e $\nu = Id$).

Indice 1 in (Un1) suggerisce che la forma generalizzata (Un1) non è unica. Infatti, ora ne scriviamo un'altra (Un2) a partire dal primale di (St2).

Consideriamo allora il primale del tipo (St2). Per il Teorema A.1.2 lo possiamo riscrivere come

$$\min_{d, \gamma \geq 0} \{ \gamma f_{\bar{x}}\left(\frac{d}{\gamma}\right) + \phi^*(\gamma) \}.$$

Dato che consideriamo (St2), $\nu = Id$ e nulla cambia se riscriviamo il problema di sopra come

$$\min_{d, \gamma \geq 0} \{ \gamma \nu\left(f_{\bar{x}}\left(\frac{d}{\gamma}\right)\right) + \phi^*(\gamma) \},$$

e questo è equivalente a

$$\min_d [\phi([\nu(f_{\bar{x}}(\cdot))]^*)]^*(d). \quad (4.15)$$

Il duale di (4.15) è

$$\min_g \phi([\nu(f_{\bar{x}})]^*)(g), \quad (4.16)$$

ed è equivalente al duale di (St2) quando $\nu = Id$.

Affinché (4.15) e (4.16) siano la forma unificante sia per (St1), che per (St2), bisogna considerare il caso $\phi = Id$. Con $\phi = Id$ in (4.15) otteniamo il primale e in (4.16) il duale di (St1).

Analogamente a (St1) e (St2), possiamo distinguere il caso “good” e il caso “bad” anche per (Un1) e (Un2).

4.3.1 Caso “good” della stabilizzazione (Un1)

Se vale la seguente proprietà:

$$(P1_{\nu, \phi}) \quad 0 \in \partial \nu([\phi(f_{\bar{x}}^*)]^*(0)) \iff 0 \in \partial f(\bar{x}),$$

allora il problema originale e il problema con la stabilizzazione di tipo (Un1) sono equivalenti, nel senso che \bar{x} è la soluzione ottima per entrambi ([4], Ch. XV, Th. 4.1.7).

Il seguente lemma ci da l'insieme di proprietà di ν e ϕ che garantiscono che vale la $(P1_{\nu, \phi})$:

Lemma 4.3.1 *Assumiamo che valgono $(P2_\nu)$ e $(P2_\phi)$. Allora vale $(P1_{\nu,\phi})$.*

DIMOSTRAZIONE. Ricordiamo che

(1) $f_{\bar{x}}(0) = 0$ e che

(2) $\partial f_{\bar{x}}(0) \equiv \partial f(\bar{x})$.

$\Leftrightarrow 0 \in \partial f(\bar{x}) \Rightarrow_{(2)} 0 \in \partial f_{\bar{x}}(0)$.

Per la proprietà (iv) della coniugata, questo è equivalente a $0 \in \partial f_{\bar{x}}^*(0)$. Dato che vale $(P2_\phi)$, vale anche $(P1_\phi)$ e, considerando l'ottimalità di \bar{x} , si ha

$$0 \in \partial f_{\bar{x}}^*(0) \equiv 0 \in \partial \phi(f_{\bar{x}}^*(0)).$$

Applicando un'altra volta la proprietà (iv) della coniugata, si ha che $0 \in \partial[\phi(f_{\bar{x}}^*)]^*(0)$

Usando la relazione (A.1), si ha che $0 \in \partial\alpha[\phi(f_{\bar{x}}^*)]^*(0)$, $\forall \alpha > 0$. In particolare, possiamo scegliere un $\alpha \in \partial\nu([\phi(f_{\bar{x}}^*)]^*(0))$. Tale α è strettamente positivo:

$\alpha \in \partial\nu([\phi(f_{\bar{x}}^*)]^*(0))$ in forma estesa è

$$\alpha \in \partial\nu(\min_{\gamma \geq 0} \{\gamma f_{\bar{x}}(0) + \phi^*(\gamma)\}) \equiv_{(1)} \alpha \in \partial\nu(\min_{\gamma \geq 0} \phi^*(\gamma)).$$

Usando ancora la relazione (A.3) si ha

$$\alpha \in \partial\nu(-\phi(0)),$$

e questo, dato che vale la proprietà $P1_\phi$ è equivalente ad $\alpha \in \partial\nu(0)$. Un tale α è maggiore di 0 perché ν è strettamente monotona crescente in 0 (Proposizione 4.2.1).

Quindi abbiamo $0 \in \partial\alpha[\phi(f_{\bar{x}}^*)]^*(0)$, $\alpha \in \partial\nu([\phi(f_{\bar{x}}^*)]^*(0))$, $\alpha > 0$ e per il Teorema A.1.18, questo è equivalente a $0 \in \partial\nu([\phi(f_{\bar{x}}^*)]^*(0))$.

\Rightarrow) Per il Teorema A.1.18

$$0 \in \partial\nu([\phi(f_{\bar{x}}^*)]^*(0)) \Leftrightarrow \exists \alpha \geq 0 : 0 \in \partial\alpha[\phi(f_{\bar{x}}^*)]^*(0), \alpha \in \partial\nu([\phi(f_{\bar{x}}^*)]^*(0))$$

Per le Proprietà (iv) e (ix) della coniugata, questo è equivalente a

$$0 \in \partial\alpha\phi(f_{\bar{x}}^*(0)), \alpha \in \partial\nu([\phi(f_{\bar{x}}^*)]^*(0)), \alpha > 0$$

e, per la relazione (A.1), quest'ultima è equivalente a

$$0 \in \partial\phi(f_{\bar{x}}^*(0)), \alpha \in \partial\nu([\phi(f_{\bar{x}}^*)]^*(0)), \alpha > 0.$$

Per la relazione (A.4), questo significa che $f_{\bar{x}}^*(0)$ minimizza ϕ . Ricordiamo che

$$f_{\bar{x}}^*(0) = f^*(0) - 0\bar{x} + f(\bar{x}).$$

Per la relazione (A.3), questo è equivalente a

$$f_{\bar{x}}^*(0) = f(\bar{x}) - \min_x \{f(x)\}.$$

Naturalmente, $f(\bar{x}) \geq \min_x \{f(x)\} \forall \bar{x}$ e allora

$$f_{\bar{x}}^*(0) = f(\bar{x}) - \min_x \{f(x)\} \geq 0.$$

Per la proprietà (P2 $_{\phi}$) abbiamo che $\phi(0) = 0$ e $\forall y > 0, \phi(y) > 0$. Quindi, se $y \geq 0$, allora y che minimizza ϕ è $y = 0$. Dunque, $f_{\bar{x}}^*(0) = 0$. Usando ancora (A.3) si ha

$$0 = f_{\bar{x}}^*(0) = -\min_d \{f_{\bar{x}}(d)\} \Rightarrow d = 0 \Rightarrow \bar{x} \text{ è ottima} \Rightarrow 0 \in \partial f(\bar{x}).$$

Rimane da far vedere che $\alpha > 0$:

Usando il Teorema A.1.2 e (A.3)

$$\alpha \in \partial \nu([\phi(f_{\bar{x}}^*)]^*)(0) \Leftrightarrow \alpha \in \partial \nu(\min_{\beta \geq 0} \beta f_{\bar{x}}(0) + \phi^*(\beta)),$$

e quest'ultima è, per (1), equivalente ad $\alpha \in \partial \nu(\min_{\beta \geq 0} \phi^*(\beta))$. Usando, ancora una volta la relazione (A.3), questo è equivalente ad $\alpha \in \partial \nu(-\phi(0))$. Dato che $\phi(0) = 0$ e che ν è strettamente monotona crescente in 0, si ha $\alpha \in \partial \nu(0)$ e $\alpha > 0$. \diamond

4.3.2 Caso “good” della stabilizzazione (Un2)

Se vale la seguente proprietà:

$$(P1_{\phi, \nu}) \quad 0 \in \partial[\phi([\nu(f_{\bar{x}})]^*)]^*(0) \iff 0 \in \partial f(\bar{x}),$$

allora il problema originale e il problema con la stabilizzazione di tipo (Un2) sono equivalenti, nel senso che hanno la stessa soluzione ottima (per ([4], Ch. XV, Th. 4.1.7)).

Il seguente lemma ci dà l'insieme di proprietà di ν e ϕ che garantiscono che vale la (P1 $_{\phi, \nu}$):

Lemma 4.3.1 *Assumiamo che valgono (P2 $_{\nu}$) e (P2 $_{\phi}$). Allora vale (P1 $_{\phi, \nu}$).*

DIMOSTRAZIONE. Ricordiamo che (P2 $_{\nu}$) \Rightarrow (P1 $_{\nu}$) e di conseguenza vale anche (P1 $_{\nu}$), e che

(1) $f_{\bar{x}}(0) = 0$ e che

(2) $\partial f_{\bar{x}}(0) \equiv \partial f(\bar{x})$.

$\Leftrightarrow 0 \in \partial f(\bar{x}) \Rightarrow_{(P1_\nu)} 0 \in \partial \nu(f_{\bar{x}}(0))$. Usando la relazione (A.1), questo si può riscrivere come $0 \in \partial \alpha \nu(f_{\bar{x}}(0))$, $\alpha > 0$. E, per le proprietà (iv) e (ix) della coniugata, questo è equivalente a $0 \in \partial \alpha [\nu(f_{\bar{x}})]^*(0)$, $\alpha > 0$. In particolare, possiamo scegliere un $\alpha \in \partial \phi([\nu(f_{\bar{x}})]^*)(0)$. Infatti, grazie al Teorema A.1.2, questo è equivalente a

$$\alpha \in \partial \phi(\min_{\beta \geq 0} \beta f_{\bar{x}}^*(0) + \nu^*(\beta)),$$

e grazie all'ottimalità di \bar{x} anche a

$$\alpha \in \partial \phi(\min_{\beta \geq 0} \nu^*(\beta)).$$

A questo punto, la relazione (A.3) ci permette di concludere che $\alpha \in \partial \phi(-\nu(0))$. Dato che $\nu(0) = 0$ e che ϕ è strettamente monotona crescente in 0, si ha $\alpha > 0$.

Quindi abbiamo $0 \in \partial \alpha [\nu(f_{\bar{x}})]^*(0)$, $\alpha \in \partial \phi([\nu(f_{\bar{x}})]^*)(0)$, $\alpha > 0$, e possiamo usare il Teorema A.1.18 per arrivare a $0 \in \partial \phi([\nu(f_{\bar{x}})]^*)(0)$. Usando la proprietà (iv) della coniugata, questo è equivalente a

$$0 \in \partial[\phi([\nu(f_{\bar{x}})]^*)]^*(0).$$

\Rightarrow) Usando la proprietà (iv) della coniugata si ha

$$0 \in \partial[\phi([\nu(f_{\bar{x}})]^*)]^*(0) \Leftrightarrow 0 \in \partial \phi([\nu(f_{\bar{x}})]^*)(0).$$

Questo significa, per la relazione (A.4) che $[\nu(f_{\bar{x}})]^*(0)$ minimizza ϕ . Per il Teorema A.1.2

$$[\nu(f_{\bar{x}})]^*(0) = \min_{\beta \geq 0} \{\beta f_{\bar{x}}^*(0) + \nu^*(\beta)\}.$$

Abbiamo già fatto vedere, nella dimostrazione del Lemma 4.3.1, che $f_{\bar{x}}^*(0) \geq 0$. Dal fatto che $\nu(0) = 0$ e dalla disuguaglianza di Fenchel si ha che anche $\nu^*(\beta) \geq 0 \forall \beta$. Quindi la somma di questi due termini è positiva, cioè $[\nu(f_{\bar{x}})]^*(0) \geq 0 \forall \bar{x}$. Come abbiamo già osservato nella dimostrazione del Lemma 4.3.1, la funzione $\phi(y)$ per $y \geq 0$ è minimizzata da $y = 0$. Quindi noi abbiamo che

$$[\nu(f_{\bar{x}})]^*(0) = \min_{\beta \geq 0} \{\beta f_{\bar{x}}^*(0) + \nu^*(\beta)\} = 0.$$

La somma di due termini positivi è 0 se e solo se sono entrambi 0. Ora, per la proprietà (vi) della coniugata, $\nu^*(\beta) = 0$ se e solo se $\beta \in \partial \nu(0)$. Questo, per la $(P1_\nu)$, implica che $\beta > 0$. Allora, $\beta f_{\bar{x}}^*(0) = 0$ se e solo se $f_{\bar{x}}^*(0) = 0$ il che significa che \bar{x} è ottima e che, quindi, $0 \in \partial f(\bar{x})$. \diamond

Capitolo 5

Dimostrazione di convergenza

5.1 Introduzione

In questo capitolo dimostriamo la convergenza dell'algoritmo di tipo Proximal Point con la stabilizzazione di tipo (Un2).

Abbiamo definito, nel capitolo precedente, il caso “good” di una stabilizzazione come un insieme di proprietà che deve avere la stabilizzazione per garantire che il problema stabilizzato è equivalente al problema originale. Qui, l'equivalenza di due problemi si intende nel senso di [4], Ch. XV, Theorem 4.1.7. Questo teorema ci dice che due problemi sono equivalenti se hanno lo stesso valore ottimo. Quindi, il nostro caso “good” descrive le stabilizzazioni tali che il valore ottimo del problema stabilizzato e del problema non stabilizzato sono equivalenti. Allora, risolvere il problema stabilizzato (Un2) equivale a risolvere (II).

Il caso “good” della stabilizzazione della funzione obiettivo (Un2) occorre quando vale $(P1_{\phi,\nu})$. Come visto in Lemma 4.3.1, questo accade se valgono $(P2_\nu)$ e $(P2_\phi)$. Tra le nostre stabilizzazioni, sono il Bundle generalizzato e SPA i rappresentanti del caso “good” di (Un2). Infatti, la funzione ϕ è l'identità nel caso di Bundle generalizzato ed è una funzione monotona e crescente ovunque, non negativa, e rispetta $(P2_\phi)$ nel caso di SPA (è la funzione ψ descritta da (2.49)). La funzione ν è l'identità per entrambi. L'identità rispetta $(P2_\nu)$ e $(P2_\phi)$. Essa è monotona crescente in 0 e $Id(0) = 0$.

Per il Level Bundle, la stabilizzazione ν è un indicatrice e ϕ è identità, mentre per il Dual Bundle è al contrario - ϕ è un indicatrice e ν è identità. Abbiamo già visto che Id è un caso “good” sia per ν che per ϕ . Questo purtroppo non è vero per l'indicatrice. Così abbiamo che la stabilizzazione

ν del Level bundle $I_{\{v \leq t\}}$ con $t < 0$ è tale che $I_{\{0 \leq t\}} = +\infty$ e $\forall v > 0$, $I_{\{v \leq t\}} = +\infty$ e quindi non vale $(P2_\nu)$. La stabilizzazione ϕ del Dual Bundle $I_{\{\sigma \leq \epsilon\}}$ con $\epsilon \geq 0$ è tale che $I_{\{0 \leq \epsilon\}} = 0$, ma $\forall \epsilon \geq \sigma > 0$, $I_{\{\sigma \leq \epsilon\}} = I_{\{0 \leq \epsilon\}} = 0$ e quindi non vale $(P2_\phi)$. In conclusione, per queste due stabilizzazioni, la soluzione ottima che trova PPA per (Un2) può essere diversa dalla soluzione ottima del problema originale (II).

Per risolvere questa situazione, abbiamo bisogno di manipolare i due parametri t ed ϵ . La manipolazione consiste nel mandare a 0 i due parametri t_k ed ϵ_k quando $k \rightarrow \infty$ (dove k è l'indice dell'iterazione). In questo modo finiamo in un caso "good" di (Un2).

Proposizione 5.1.1 *Per il Level Bundle e il Dual Bundle, la stabilizzazione del tipo (Un2) rispetta la proprietà $(P1_{\phi,\nu})$ se e solo se $\epsilon_k = t_k = 0$.*

DIMOSTRAZIONE. Dire che per (Un2) vale $(P1_{\phi,\nu})$, significa che valgono $(P2_\nu)$ e $(P2_\phi)$. Nel caso di Level Bundle, $\nu_k(v) = I_{v \leq t_k}$ e $\phi_k = Id \forall \epsilon_k$. Se $\epsilon_k = t_k = 0$, allora si ha che è vera

- $(P2_\phi)$ perché Id è strettamente monotona crescente in 0 e $Id(0) = 0$ e non è influenzabile da ϵ_k ,
- $(P2_\nu)$ perché $I_{v \leq 0}(0) = 0$ ed è strettamente monotona crescente in 0 (infatti, $I_{v \leq 0}(0) = 0$ e $\forall v > 0$, $I_{v \leq 0} = +\infty > 0$).

Nel caso di Dual Bundle, $\phi_k(\sigma) = I_{\sigma \leq \epsilon_k}$ e $\nu_k = Id \forall t_k$. Se $\epsilon_k = t_k = 0$, allora si ha che è vera

- $(P2_\nu)$ per gli stessi motivi per cui è vera $(P2_\phi)$ nel Level Bundle,
- $(P2_\phi)$ perché $I_{\sigma \leq 0}(0) = 0$ ed è strettamente monotona crescente in 0 (infatti, $I_{\sigma \leq 0}(0) = 0$ e $\forall \sigma > 0$, $I_{\sigma \leq 0} = +\infty > 0$).

◇.

5.2 Formulazione dell'algoritmo

Introduciamo la seguente notazione per la stabilizzazione della funzione obiettivo in (Un2) $([\phi([\nu(f_{\bar{x}})]^*)]^*(d))$ alla i -esima iterazione:

$$\xi_{x_i}(d) = [\phi([\nu(f_{x_i}(\cdot))]^*)]^*(d). \quad (5.1)$$

Sia

$$D_{c_i}(d) = \frac{1}{2c_i} \|d\|_2^2, \quad (5.2)$$

dove

$$\forall i, c_i \geq \underline{c} > 0 \Rightarrow \sum_{i=0}^{\infty} c_i = +\infty. \quad (5.3)$$

La funzione

$$h_{x_i}(d) = \min_d \left\{ \xi_{x_i}(d) + \frac{1}{2c_i} \|d\|_2^2 \right\} \quad (5.4)$$

è la Moreau - Yosida regularizzazione di ξ_{x_i} associata con $\|\cdot\|_2^2$. Il punto $x_i + d_i$, dove

$$d_i = \arg \min_d \left\{ \xi_{x_i}(d) + \frac{1}{2c_i} \|d\|_2^2 \right\} \quad (5.5)$$

è il cosiddetto proximal point (o proxiaml center) di x_i associato con ξ_{x_i} e $\|\cdot\|_2^2$.

(5.5) significa:

$$\begin{aligned} 0 &\in \partial h_{x_i}(d_i) = \partial \left[\xi_{x_i}(d_i) + \frac{1}{2c_i} \|d_i\|_2^2 \right] \\ &\equiv -\frac{1}{c_i} d_i \in \partial \xi_{x_i}(d_i) \end{aligned}$$

Chiamiamo

$$s_i = -\frac{1}{c_i} d_i \quad (5.6)$$

questo particolare subgradiente di ξ_{x_i} su d_i che è la soluzione unica di (5.4) (perché h_{x_i} è fortemente convessa $\forall i$).

La minimizzazione di h_{x_i} è legata alla minimizzazione di ξ_{x_i} nel modo seguente:

Lemma 5.2.1 Per ogni x_i, d_i ,

$$h_{x_i}(d_i) \leq \xi_{x_i}(0).$$

DIMOSTRAZIONE. Da (5.4) abbiamo che

$$h_{x_i}(d_i) = \xi_{x_i}(d_i) + \frac{1}{2c_i} \|d_i\|_2^2.$$

Usando la definizione (A.1.4) del subgradiente e (5.6) si ha, per ogni x_i, d_i ,

$$\begin{aligned} \xi_{x_i}(0) &\geq \xi_{x_i}(d_i) + s_i(0 - d_i) \\ &= \xi_{x_i}(d_i) + \frac{1}{c_i} d_i d_i \\ &= \xi_{x_i}(d_i) + \frac{1}{c_i} \|d_i\|_2^2 \\ &\geq \xi_{x_i}(d_i) + \frac{1}{2c_i} \|d_i\|_2^2 \\ &= h_{x_i}(d_i). \diamond \end{aligned}$$

In questo caso PPA ha seguente forma:

0) scegliere x_1 iniziale;

ripetti:

1) $d_i = \arg \min_d \{ \xi_{x_i}(d) + \frac{1}{2c_i} \|d\|_2^2 \}$;

2) se $d_i = 0$ STOP;

3) altrimenti, aggiusta c_i , $i = i + 1$ e torna al passo 1);

L'algorithmo PPA é discendente. Infatti,

$$\min_d h_{x_i}(d_i) = \min_{\gamma \geq 0, d} \{ \gamma \nu\left(\frac{f_{x_i}(d)}{\gamma}\right) + \phi^*(\gamma) + \frac{1}{2c_i} \|d\|_2^2 \}. \quad (5.7)$$

Sia $\bar{\gamma} \in \partial\phi(0)$. Per $(P2_\phi)$, $\bar{\gamma} > 0$. Allora $(\bar{\gamma}, 0)$ è ammissibile per (5.7). Per la $(P2_\nu)$ abbiamo che $\nu(0) = 0$ e per la proprietà (vi) della coniugata abbiamo che $\phi^*(\bar{\gamma}) = \bar{\gamma}0 - \phi(0) = 0$. Tutto questo ci da

$$\bar{\gamma} \nu\left(\frac{0}{\bar{\gamma}}\right) + \phi^*(\bar{\gamma}) + \frac{1}{2c_i} \|0\|_2^2 = 0.$$

Allora la soluzione ottima di (5.7) è

$$\min_{\gamma \geq 0} \{ \gamma \nu\left(\frac{f_{x_i}(d_i)}{\gamma}\right) + \phi^*(\gamma) + \frac{1}{2c_i} \|d_i\|_2^2 \} \leq 0. \quad (5.8)$$

Visto che $\phi^*(\gamma)$ e $\frac{1}{2c_i} \|d_i\|_2^2$ sono due termini non negativi, deve essere che $\gamma \nu\left(\frac{f_{x_i}(d_i)}{\gamma}\right) \leq 0$. Notiamo che γ non può essere 0 perché $(\gamma = 0, d_i)$ ci da un valore strettamente positivo in (5.7) (questo si ottiene dal fatto che $\phi(0) = 0$, dal fatto che $0 < \gamma \in \partial\phi(0)$ e applicando le proprietà (vi) e (vii) della coniugata a $\phi^*(0)$) e sappiamo, per (5.8) che tale soluzione non è ottima. Quindi, $\nu\left(\frac{f_{x_i}(d_i)}{\gamma}\right) \leq 0$ e $\gamma > 0$. Per la monotonia di ν questo implica $f_{x_i}(d_i) \leq 0$, cioè, $f(x_i + d_i) - f(x_i) \leq 0$.

5.3 Convergenza

Per dimostrare le convergenza del nostro algorithmo, assumiamo che, a parte le proprietà che descrivono il caso “good” della stabilizzazione (Un2), ν sia strettamente monotona e crescente ovunque.

Ad ogni iterazione si ha

$$s_i \in \partial \xi_{x_i}(d_i). \quad (5.9)$$

Chiamiamo $\xi_* = \inf \xi = \inf f = f_*$, e $f_\infty = \lim_{i \rightarrow \infty} \{f(x_i)\}$.

Sia

$$\delta_i = -h_{x_i}(d_i) = -\xi_{x_i}(d_i) - \frac{1}{2c_i} \|d_i\|_2^2. \quad (5.10)$$

Per ogni i , $\delta_i \geq 0$. Infatti, il Lemma 5.2.1 ci dice che

$$h_{x_i}(d_i) \leq \xi_{x_i}(0) \Leftrightarrow -h_{x_i}(d_i) \geq -\xi_{x_i}(d_i) \Rightarrow \delta_i \geq -\xi_{x_i}(0).$$

Ora,

$$\xi_{x_i}(0) = [\phi([\nu(f_{x_i}(0))]^*)]^* = [\phi([\nu(0)]^*)]^*.$$

Usando la coniugata della composizione di funzioni si ha

$$[\phi([\nu(0)]^*)]^* = \min_{\gamma \geq 0} \{\gamma \nu(0) + \phi^*(\gamma)\}.$$

Per $(P2_\nu)$, $\nu(0) = 0$ e $[\phi([\nu(0)]^*)]^*$ diventa

$$\min_{\gamma \geq 0} \{\phi^*(\gamma)\} = -\phi(0) = 0,$$

dove l'ultima uguaglianza viene da $(P2_\phi)$. Quindi, $\delta_i \geq 0$.

Proposizione 5.3.1 *Sia ϕ strettamente monotona e cerscente ovunque e sia $\text{dom } \phi^* \subseteq (-\infty, u)$. Allora $\forall \gamma \in \text{dom } \phi^*$,*

$$0 < l \leq \gamma \leq u < +\infty.$$

DIMOSTRAZIONE. Se ϕ è strettamente monotona e cerscente ovunque, allora, la Proposizione 4.2.1 ci dice che tutte le subderivate di ϕ sono strettamente positive, cioè

$$\forall \alpha, \forall \gamma \in \partial \phi(\alpha), \exists l > 0 : \gamma \geq l.$$

Insieme al fatto che $\text{dom } \phi^* \subseteq (-\infty, u)$ ($\Rightarrow \gamma \in \text{dom } \phi^* \leq u$), si ha

$$0 < l \leq \gamma \leq u < +\infty, \forall \gamma \in \text{dom } \phi^*. \diamond$$

Proposizione 5.3.2 *Siano valide le ipotesi della Proposizione 5.3.1, le proprietà $(P2_\nu)$ e $(P2_\phi)$ e sia ν continua e strettamente monotona e crescente ovunque. Supponiamo, inoltre, che*

$$\inf_{\gamma > 0} \left\{ \gamma \nu \left(\frac{f_{x_i}(d_i)}{\gamma} \right) \right\} = \gamma_i \nu \left(\frac{f_{x_i}(d_i)}{\gamma_i} \right). \quad (5.11)$$

Allora, $\delta_i \rightarrow 0$ quando $i \rightarrow \infty$.

DIMOSTRAZIONE. Supponiamo, per assurdo, che $\delta_i \geq \sigma > 0$. Allora $\forall i$,

$$\begin{aligned}
& -\delta_i \leq -\sigma < 0 \\
& \equiv h_{x_i}(d_i) \leq -\sigma < 0 \\
& \equiv \xi_{x_i}(d_i) + \frac{1}{2c_i} \|d_i\|_2^2 \leq -\sigma < 0 \\
& \equiv \xi_{x_i}(d_i) \leq -\sigma < 0 \\
& \equiv \min_{\gamma \geq 0} \left\{ \gamma \nu \left(\frac{f_{x_i}(d_i)}{\gamma} \right) + \phi^*(\gamma) \right\} \leq -\sigma < 0
\end{aligned}$$

Sappiamo, dal fatto che $\phi(0) = 0$, che

$$\forall \gamma, \gamma \nu \left(\frac{f_{x_i}(d_i)}{\gamma} \right) + \phi^*(\gamma) \geq \gamma \nu \left(\frac{f_{x_i}(d_i)}{\gamma} \right).$$

Allora anche

$$\min_{\gamma \geq 0} \left\{ \gamma \nu \left(\frac{f_{x_i}(d_i)}{\gamma} \right) \right\} \leq \min_{\gamma \geq 0} \left\{ \gamma \nu \left(\frac{f_{x_i}(d_i)}{\gamma} \right) + \phi^*(\gamma) \right\} \leq -\sigma < 0.$$

La soluzione $\gamma = 0$ non è ammissibile perché $0 \nu \left(\frac{f_{x_i}(d_i)}{0} \right) = 0 > -\sigma$. Possiamo riscrivere il problema di sopra come

$$\min_{\gamma > 0} \left\{ \gamma \nu \left(\frac{f_{x_i}(d_i)}{\gamma} \right) \right\} \leq -\sigma < 0, \quad (5.12)$$

Usando la nostra ipotesi (5.11), il problema (5.12) è equivalente a

$$\gamma_i \nu \left(\frac{f_{x_i}(d_i)}{\gamma_i} \right) \leq -\sigma < 0, \quad \forall i.$$

Dalla Proposizione 5.3.1 si ha che $\forall i$, $0 < l \leq \gamma_i \leq u < +\infty$. Allora

$$\begin{aligned}
& \gamma_i \nu \left(\frac{f_{x_i}(d_i)}{\gamma_i} \right) \leq -\sigma < 0 \\
& \equiv \nu \left(\frac{f_{x_i}(d_i)}{\gamma_i} \right) \leq \frac{-\sigma}{\gamma_i} \leq \frac{-\sigma}{u} < 0.
\end{aligned}$$

Data la stretta monotonia di ν si deduce che esiste un $-\sigma' < 0$ tale che $f_{x_i}(d_i) \leq -\sigma'$. Dunque,

$$-\delta_i \leq -\sigma < 0 \Rightarrow \nu \left(\frac{f_{x_i}(d_i)}{\gamma_i} \right) \leq \frac{-\sigma}{u} < 0 \Rightarrow f_{x_i}(d_i) \leq -\sigma' < 0 \quad \forall i.$$

Questo è corretto se $f_* = -\infty$. Supporiamo ora, che $f_* > -\infty$. Allora esiste una sottosequenza $f(x_{j_i} + d_{j_i}) - f(x_{j_i})$ che tende a 0 per $i \rightarrow \infty$. Allora $\frac{f_{x_{j_i}}(d_{j_i})}{l} \rightarrow 0$ e dato che $l \leq \gamma_i$, si ha

$$\frac{f_{x_{j_i}}(d_{j_i})}{l} \rightarrow 0 \quad \Rightarrow \quad \frac{f_{x_{j_i}}(d_{j_i})}{\gamma_i} \rightarrow 0,$$

il che, per continuità di ν , implica che

$$\nu\left(\frac{f_{x_{j_i}}(d_{j_i})}{\gamma_i}\right) \rightarrow 0$$

Questa è contraddizione perché abbiamo ipotizzato che $\nu\left(\frac{f_{x_{j_i}}(d_{j_i})}{\gamma_i}\right) \leq \frac{-\sigma}{u} < 0$. Allora, deve essere vero che $\delta_i \rightarrow 0$ per $i \rightarrow \infty$. \diamond

Lemma 5.3.3 *Sia $\{c_i\}$ data da (5.3). Siano valide le ipotesi della Proposizione 5.3.2, $(P2_\nu)$ e $(P2_\phi)$ e sia ν strettamente monotona e crescente ovunque. Allora $\{x_i\}$ è una sequenza di minimizzazione.*

DIMOSTRAZIONE. Dalla disuguaglianza del subgradiente e dalla definizione (5.6) di s_i si ha $\forall y$

$$\begin{aligned} \xi_{x_i}(y) &\geq \xi_{x_i}(d_i) + s_i(y - d_i) \\ &= \xi_{x_i}(d_i) + \frac{1}{c_i}(d_i)(d_i - y) \\ &= \xi_{x_i}(d_i) - \frac{1}{c_i}d_i y + \frac{1}{c_i}\|d_i\|_2^2. \end{aligned}$$

Dalla definizione (5.10) di δ_i si ha che

$$\xi_{x_i}(d_i) = -\delta_i - \frac{1}{2c_i}\|d_i\|_2^2.$$

Allora

$$\begin{aligned} \xi_{x_i}(y) &\geq -\delta_i - \frac{1}{2c_i}\|d_i\|_2^2 - \frac{1}{c_i}d_i y + \frac{1}{c_i}\|d_i\|_2^2 \\ &= -\delta_i + \frac{1}{2c_i}\|d_i\|_2^2 - \frac{1}{c_i}d_i y \end{aligned}$$

Da qui

$$\xi_{x_i}(y) + \delta_i - \frac{1}{2c_i}\|d_i\|_2^2 \geq -\frac{1}{c_i}d_i y \quad (5.13)$$

(5.13) ci da il seguente risultato

$$\begin{aligned}
\|y - d_i\|_2^2 &= \|y\|_2^2 - 2d_i y + \|d_i\|_2^2 \\
&\leq \|y\|_2^2 - 2c_i(\xi_{x_i}(y) + \delta_i) - \frac{1}{2c_i}\|d_i\|_2^2 + \|d_i\|_2^2 \\
&= \|y\|_2^2 + 2c_i(\xi_{x_i}(y) + \delta_i).
\end{aligned} \tag{5.14}$$

Ora, se $f_\infty = -\infty$, allora anche $\xi_* = f_* = -\infty$ e abbiamo finito. Altrimenti, se $f_\infty > -\infty$, allora dimostriamo che $f_\infty = f_*$. Per fare questo, procediamo per assurdo e supponiamo che

$$\exists \bar{x} \text{ tale che } f(\bar{x}) \leq f_\infty - \epsilon \leq f(x_i) - \epsilon \forall i, \epsilon > 0. \tag{5.15}$$

Sia $\bar{x} - x_i = y$ in (5.14). Allora

$$\|\bar{x} - x_i - d_i\|_2^2 = \|\bar{x} - x_i - x_{i+1} + x_i\|_2^2 \leq \|\bar{x} - x_i\|_2^2 + 2c_i(\xi_{x_i}(\bar{x} - x_i) + \delta_i). \tag{5.16}$$

Usando la forma estesa di ξ_{x_i} , si ha che

$$\xi_{x_i}(\bar{x} - x_i) = \min_{\gamma \geq 0} \left\{ \gamma \nu\left(\frac{f_{x_i}(\bar{x} - x_i)}{\gamma}\right) + \phi^*(\gamma) \right\} \leq \min_{\gamma \geq 0} \left\{ \gamma \nu\left(\frac{-\epsilon}{\gamma}\right) + \phi^*(\gamma) \right\}, \tag{5.17}$$

dove l'ultima disuguaglianza viene dal fatto che $f_{x_i}(\bar{x} - x_i) = f(\bar{x} - x_i + x_i) - f(x_i)$, da (5.15) e dalla monotonia di ν . Per le proprietà (vi) e (vii) della coniugata, $\phi^*(\gamma) \geq 0$ e $\phi^*(\gamma) = 0 \Leftrightarrow \gamma \in \partial\phi(0)$. Per la $(P2_\phi)$, $\gamma \in \partial\phi(0) \Rightarrow \gamma > 0$. Sia $\bar{\gamma} \in \partial\phi(0)$. $\bar{\gamma}$ è ammissibile per (5.17). Essendo ν strettamente monotona e crescente ovunque e $\nu(0) = 0$, si ha che $\bar{\gamma} \nu\left(\frac{-\epsilon}{\bar{\gamma}}\right) < 0$. Invece $\phi^*(\bar{\gamma}) = 0$. Allora,

$$\bar{\gamma} \nu\left(\frac{-\epsilon}{\bar{\gamma}}\right) + \phi^*(\bar{\gamma}) < 0$$

è una soluzione ammissibile, per cui anche

$$-\epsilon' = \min_{\gamma \geq 0} \left\{ \gamma \nu\left(\frac{-\epsilon}{\gamma}\right) + \phi^*(\gamma) \right\} < 0$$

e quindi $\xi_{x_i}(\bar{x} - x_i) \leq -\epsilon' < 0 \forall i$. Allora, possiamo riscrivere (5.16) come

$$\|\bar{x} - x_{i+1}\|_2^2 \leq \|\bar{x} - x_i\|_2^2 + 2c_i(-\epsilon' + \delta_i),$$

e sommando per $i = 1, \dots, h$ si ha

$$\|\bar{x} - x_{h+1}\|_2^2 \leq \|\bar{x} - x_1\|_2^2 + \sum_{i=1}^h 2c_i(-\epsilon' + \delta_i). \tag{5.18}$$

Dato che $\delta_i \rightarrow 0$, esiste un indice p tale che $\delta_p \leq \frac{\epsilon'}{2}$ e per ogni $i > p$, $\delta_i \leq \frac{\epsilon'}{2}$. Allora (5.18) diventa

$$\|\bar{x} - x_{h+1}\|_2^2 \leq \|\bar{x} - x_1\|_2^2 + \sum_{i=1}^{p-1} 2c_i(-\epsilon' + \delta_i) + \sum_{i=p}^h 2c_i(-\epsilon' + \delta_i).$$

Qui, l'ultima somma è minore o uguale a $\sum_{i=p}^h 2c_i \frac{-\epsilon'}{2}$ che tende a $-\infty$ per h che tende a ∞ . Questo ci dà la contraddizione

$$\|\bar{x} - x_{h+1}\|_2^2 \leq -\infty.$$

Dunque, non esiste \bar{x} tale che $f(\bar{x}) \leq f_\infty - \epsilon$ e allora $f_\infty = f_*$. \diamond

Conclusione

I risultati di questa tesi sono i seguenti:

- una rassegna di letteratura in cui si evidenziano gli aspetti primali e duali di tutti gli algoritmi bundle per la minimizzazione delle funzioni convesse non differenziabili. Questa rassegna ha evidenziato le somiglianze tra essi, cosa mai fatta prima,
- attraverso il concetto della stabilizzazione della funzione obiettivo, unita alla stabilizzazione delle iterate, è stato possibile individuare un modo unificato di scrivere quasi tutti gli algoritmi bundle noti, ad eccezione dell'algoritmo dei piani di taglio con i centri di Chebychev,
- abbiamo trovato, un insieme di proprietà della stabilizzazione unificata che garantisce la convergenza dell'algoritmo di tipo Proximal point verso la soluzione ottima per il problema originale della minimizzazione della funzione convessa, non differenziabile. Questo è il primo passo importante verso una teoria di convergenza unificata. L'equivalenza delle soluzioni ottime di (II) e di stabilizzazione unificata è sempre vera per il Bundle generalizzato e SPA. Per il Level bundle e Dual bundle è vera per opportuna scelta dei parametri t ed ϵ .

L'algoritmo Proximal point è un algoritmo teorico e la nostra dimostrazione di convergenza lo è altrettanto. Ma rappresenta una solida base sulla quale, aggiungendo un'opportuna tecnica di gestione del bundle, è possibile costruire anche un algoritmo implementabile.

Appendice A

Definizioni e notazioni

A.1 Definizioni

Qui elenchiamo alcune definizioni utili per la lettura del testo.

Definizione A.1.1 Un insieme $X \subset R^n$ si dice convesso se $\gamma x + (1 - \gamma)x' \in X$ ogni volta che $x, x' \in X$ per $\gamma \in [0, 1]$.

Sia $X \subset R^n$ un'insieme convesso. La funzione $f : X \rightarrow R$ si dice convessa quando $\forall (x, x') \in X$ e $\forall \gamma \in [0, 1]$ vale:

$$f(\gamma x + (1 - \gamma)x') \leq \gamma f(x) + (1 - \gamma)f(x').$$

Definizione A.1.2 Un insieme $X \subset R^n$ si dice chiuso se coincide con la propria chiusura, o, equivalentemente, se contiene tutti i suoi punti estremi.

Una funzione convessa $f : R^n \rightarrow R$ si dice chiusa se

$$\liminf_{x \rightarrow x_0} f(x) \geq f(x_0) \quad \forall x_0 \in R^n.$$

Definizione A.1.3 Un insieme $X \subset R^n$ si dice compatto se è chiuso e limitato.

Definizione A.1.4 Sia $f : R^n \rightarrow R$ convessa. Un subgradiente di f su x è un vettore $g \in R^n$ tale che

$$f(y) \geq f(x) + g(y - x) \quad \forall y \in R^n.$$

Insieme di tutti subgradienti di f su x si chiama il subdifferenziale di f su x e si indica con $\partial f(x)$.

Definizione A.1.5 $\partial_\epsilon f(x)$ è l'insieme di tutti gli ϵ -subgradienti di $f : R^n \rightarrow R$ su x , cioè:

$$\partial_\epsilon f(x) = \{g \in R^n : f(y) \geq f(x) + g(y - x) - \epsilon, \forall y \in R^n\}.$$

Elenchiamo alcune proprietà elementari dell' ϵ -subdifferenziale:

- (I) $\partial_\epsilon[f(x) + c] \equiv \partial_\epsilon f(x)$, $c \in R$
- (II) $\partial_\epsilon[cf(x)] \equiv c\partial_{\epsilon/c}f(x)$, $c > 0$
- (III) $\partial_\epsilon[f(cx)] \equiv c\partial_\epsilon f(cx)$, $c \neq 0$.

Una conseguenza di (II) e (III) è la seguente relazione che è utile in alcune dimostrazioni:

$$0 \in \partial f(0) \Rightarrow 0 \in \partial f(0c) \Leftrightarrow_{(III)} 0 \in c\partial f(0c), c \neq 0 \Leftrightarrow_{(II)} 0 \in \partial cf(0), c > 0 \quad (A.1)$$

Definizione A.1.6 Sia $f : R^n \rightarrow R$ convessa. La coniugata di f è la funzione $f^* : R^n \rightarrow R$ definita come

$$f^*(g) = \sup_x \{gx - f(x) : g, x \in R^n\}.$$

Coniugata è una funzione chiusa e convessa con seguenti proprietà:

- (i) $(f^*)^* = f$ (dualità della coniugata)
- (ii) $f_1 \leq f_2 \Rightarrow f_1^* \geq f_2^*$ (monotonicità della coniugata)
- (iii) $(f(\cdot+x))^*(g) = f^*(g) - gx \quad \forall g, x$ (effetto del semplice cambio della variabile)
- (iv) $g \in \partial_\epsilon f(x) \Leftrightarrow x \in \partial_\epsilon f^*(g)$ (dualità del subdifferenziale)
- (v) $g \in \partial_\epsilon f(x) \Leftrightarrow f(x) + f^*(g) \leq gx + \epsilon$ (caratterizzazione dell' ϵ -subdifferenziale)
- (vi) $gx = f(x) + f^*(g) \Leftrightarrow g \in \partial f(x)$ (relazione di base tra i valori della funzione)
- (vii) $gx \leq f(x) + f^*(g) \quad \forall g, x$ (disuguaglianza di Fenchel).
- (viii) La coniugata della funzione $f(x) + c$ è $f^*(g) - c$, dove $c \in R$.
- (ix) La coniugata della funzione $cf(x)$ è $cf^*(g/c)$, dove $c > 0$.
- (x) La coniugata della funzione $f(cx)$ è $f^*(g/c)$, dove $c \neq 0$.

$$(xi) (f(x) + g_0x)^* = f^*(g - g_0).$$

Un caso particolare è la norma Euclidea per la quale

$$(\|\cdot\|_2^2)^* = \|\cdot\|_2^2. \quad (A.2)$$

Dalla definizione di coniugata abbiamo:

$$\min_x f(x) = f(x^*) = -\max_x \{0x - f(x)\} = -f^*(0). \quad (A.3)$$

Da (A.1.4) e (A.1.5) abbiamo che:

$$x \text{ minimizza } f \iff 0 \in \partial f(x). \quad (A.4)$$

e

$$x \text{ è la soluzione } \epsilon - \text{ottima di } (\Pi) \iff 0 \in \partial_\epsilon f(x). \quad (A.5)$$

Definizione A.1.7 Sia $f : R^n \rightarrow R$ una funzione convessa. L'epigrafo di f è

$$epi f = \{(x, r) : r \geq f(x)\}.$$

L'epigrafo di f^* è

$$epi f^* = \{(r, g) : r \geq f^*(x)\}.$$

Definizione A.1.8 Sia $f : R^n \rightarrow R$ una funzione convessa. La curva di livello r di f è l'insieme

$$S_r(f) = \{x : f(x) \leq r\}.$$

Definizione A.1.9 La funzione indicatrice di un'insieme convesso X è

$$I_X(x) = \begin{cases} 0 & \text{se } x \in X, \\ +\infty & \text{altrimenti.} \end{cases}$$

Definizione A.1.10 La funzione di supporto di un'insieme convesso X è

$$\sigma_X(g) = \sup_x \{gx : x \in X\}.$$

Si ha che

$$(I_X(x))^* = (I_X)^*(g) = \sup_x \{gx - I_X(x)\} = \sup_x \{gx\} = \sigma_X(g) \quad (A.6)$$

(la coniugata della funzione indicatrice di X è la funzione di supporto di X)

Definizione A.1.11 Il semplice unitario nello spazio R^n è l'insieme

$$\Delta_n = \{\lambda \in R^n : \sum_{i=1}^n \lambda_i = 1, \lambda_i \geq 0 \text{ per } i = 1 \dots n\}.$$

Definizione A.1.12 Sia M un insieme non vuoto. Convex hull di M è l'intersezione di tutti insiemi convessi che contengono M . Come tale, è il piú piccolo tra questi insiemi. Si indica con $\text{co}\{M\}$.

Definizione A.1.13 Sia M un insieme non vuoto. Affine hull di M è l'intersezione di tutti insiemi affini che contengono M . Come tale, è il piú piccolo tra questi insiemi. Si indica con $\text{aff}\{M\}$.

Definizione A.1.14 L'interno relativo $\text{ri } C$ di un insieme convesso $C \subset R^n$ è l'interno di C per la topologia relativa all' $\text{aff}\{C\}$. In altre parole, $x \in \text{ri } C$ se e solo se

$$x \in \text{aff}\{C\} \quad e \quad \exists \delta > 0 : (\text{aff}\{C\}) \cap B(x, \delta) \subset C.$$

Notare che $\text{ri } C \subset C$.

Definizione A.1.15 La derivata direzionale di f su x in direzione $d (= \pm 1)$ è

$$f'(x, d) = \inf_{t > 0} \frac{f(x + td) - f(x)}{t}.$$

Definizione A.1.16 La ϵ -derivata direzionale di f su x è

$$f'_\epsilon(x, d) = \inf_{t > 0} \frac{f(x + td) - f(x) + \epsilon}{t}.$$

Essa è la funzione di supporto dell' ϵ -subdifferenziale di f su x :

$$f'_\epsilon(x, d) = \sigma_{\partial_\epsilon f(x)}(d) = \sup_{g \in \partial_\epsilon f(x)} \{gd\}.$$

Definizione A.1.17 Una funzione $f : R^n \rightarrow R^n$ chiusa, convessa e 0-coerciva è tale che

$$f(x) \rightarrow +\infty \quad \text{ogni volta che} \quad \|x\| \rightarrow +\infty.$$

Le funzioni chiuse, convesse e 0-coercive hanno un minimo in R^n .

Un caso particolare si ha quando f è tale che

$$\frac{f(x)}{\|x\|} \rightarrow +\infty \quad \text{ogni volta che} \quad \|x\| \rightarrow +\infty.$$

In questo caso f è detta 1-coerciva, o semplicemente coerciva.

Composizione delle funzioni convesse

Scriviamo la coniugata della composizione delle due funzioni in (1.9). Usiamo il Teorema 2.5.1 di [4] (cap.X), modificando l'ipotesi sulla ν crescente, nell'ipotesi sulla ν non decrescente (notare che questo si può fare, perché in questo teorema, l'uso del fatto che ν è crescente, sta nel dire che $-(\nu \circ f_{\bar{x}})^*(g) = \inf_d \{\nu(f_{\bar{x}}(d)) - gd\} = \inf_{d,r} \{\nu(r) - gd : f_{\bar{x}}(d) \leq r\}$ e il minimo corrisponde a $f_{\bar{x}}(d) = r$, ma questo è vero anche per una funzione ν non decrescente per la quale $\inf_d \{\nu(f_{\bar{x}}(d)) - gd\} = \inf_{d,r} \{\nu(r) - gd : f_{\bar{x}}(d) = r \forall d\}$).

Otteniamo il seguente teorema:

Teorema A.1.1 *Assumiamo che $\nu : R \rightarrow R$ sia una funzione non decrescente e convessa. Assumiamo inoltre che $f(R^n) \cap \text{intdom } \nu \neq \emptyset$. Allora la funzione $\nu \circ f_{\bar{x}}$ è chiusa e convessa su R^n e*

$$(\nu \circ f_{\bar{x}})^*(g) = \min_{\gamma \in R} \begin{cases} \gamma f_{\bar{x}}^*(\frac{1}{\gamma}g) + \nu^*(\gamma) & \text{se } \gamma > 0 \\ \sigma_{\text{dom } f_{\bar{x}}}(g) + \nu^*(0) & \text{se } \gamma = 0 \\ +\infty & \text{se } \gamma < 0 \end{cases}$$

$\gamma f_{\bar{x}}^*(\frac{1}{\gamma}g) = \sigma_{\text{epi } f_{\bar{x}}}(g, -\gamma)$ ([4], Ch.X, Proposition 1.2.1), dove $\sigma_{\text{epi } f_{\bar{x}}}(g, -\gamma)$ è il supporto dell'epigrafo di $f_{\bar{x}}$. In particolare, $\sigma_{\text{dom } f_{\bar{x}}}(g) = \sigma_{\text{epi } f_{\bar{x}}}(g, 0)$. Allora, $(\nu \circ f_{\bar{x}})^*(g) = \sigma_{\text{epi } f_{\bar{x}}}(g, -\gamma) + \nu^*(\gamma)$. Possiamo eliminare il caso $\gamma < 0$, perché stiamo calcolando \min_{γ} . Tutto questo ci permette di riscrivere il Teorema A.1.1 come:

Teorema A.1.2 *Assumiamo che $\nu : R \rightarrow R$ sia una funzione non decrescente e convessa e $f : R^n \rightarrow R$ una funzione chiusa e convessa. Assumiamo inoltre che $f(R^n) \cap \text{intdom } \nu \neq \emptyset$. Allora la funzione $\nu \circ f_{\bar{x}}$ è chiusa e convessa su R^n e*

$$(\nu \circ f_{\bar{x}})^*(g) = \min_{\gamma \geq 0} \left\{ \gamma f_{\bar{x}}^*\left(\frac{1}{\gamma}g\right) + \nu^*(\gamma) \right\}. \quad (\text{A.7})$$

Teorema A.1.18 *Riportiamo qui la definizione ([4], Ch.XI, Th.3.6.1) dell' ϵ -subdifferenziale di una composizione di funzioni:*

Assumiamo che $f : R^n \rightarrow R$ sia convessa e che $\nu : R \rightarrow R$ sia crescente, chiusa e convessa. Assumiamo inoltre che $f(R^n) \cap \text{intdom } \nu \neq \emptyset$. Allora per ogni x tale che $f(x) \in \text{dom } g$,

$$\begin{aligned} g \in \partial_{\epsilon} \nu(f(x)) & \iff \\ \exists \epsilon_1, \epsilon_2, \alpha \geq 0 & \text{ tali che} \\ \epsilon_1 + \epsilon_2 = \epsilon, \quad g \in \partial_{\epsilon_1}(\alpha f)(x), & \quad \alpha \in \partial_{\epsilon_2} \nu(f(x)). \end{aligned}$$

Duali dell'Analisi Convessa

Definizione A.1.19 Consideriamo un generico problema d'ottimizzazione vincolata

$$(\Pi L) \quad \min\{f(x) : g_j(x) \leq 0, j = 1, \dots, m \text{ e } x \in X\} \quad (\text{A.8})$$

con f convessa. Allora il problema

$$(\Delta L) \quad \max_{\lambda \geq 0} \{\underline{L}(\lambda) \equiv \inf_{x \in X} \mathcal{L}(x, \lambda)\} \quad (\text{A.9})$$

è il duale Lagrangiano di (ΠL) . Qui,

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{j=1}^m \lambda_j g_j(x) \quad (\text{A.10})$$

è la funzione Lagrangiana di (ΠL) .

Definizione A.1.20 Il duale Lagrangiano di un problema di programmazione quadratica (QP) con i vincoli lineari si chiama duale quadratico. In questo caso, (ΠL) della definizione (A.1.19) è della forma

$$(\Pi Q) \quad \min_x \{f(x) = \frac{1}{2}x^T Qx + qx + c^T y : Ax + By \geq b\}. \quad (\text{A.11})$$

La funzione Lagrangiana è:

$$\begin{aligned} \mathcal{L}(x, y, \lambda) &= \frac{1}{2}x^T Qx + qx + c^T y + \lambda(b - Ax - By) \\ &= \frac{1}{2}x^T Qx + (q - \lambda A)x + (c - \lambda B)y + \lambda b. \end{aligned}$$

Poniamo $\gamma = -x$. Dunque $\inf_{x,y} \mathcal{L}(x, y, \lambda)$ si ottiene quando $\lambda A + \gamma Q = q$ e $\lambda B = c$. Allora, rimettendo queste condizioni nella Lagrangiana e massimizzando rispetto a λ e γ , si ottiene il duale quadratico:

$$(\Delta Q) \quad \max_{\lambda, \gamma} \left\{ -\frac{1}{2} \gamma^T Q \gamma + \lambda b : \lambda A + \gamma Q = q, \lambda B = c, \lambda \geq 0 \right\}. \quad (\text{A.12})$$

Definizione A.1.21 Siano f_1 ed f_2 due funzioni chiuse e convesse e tali che $\text{ri} \text{dom} f_1 \cap \text{ri} \text{dom} f_2 \neq \emptyset$ (dove $\text{ri} C$, con C insieme convesso, è tale che $x \in \text{ri} C \iff x \in \text{aff} C$ e $\exists \delta > 0 : \text{aff} C \cap B(x, \delta) \subset C$ e $\text{aff} C = \bigcap \{S : S \text{ è affine e contiene } C\}$). Se la soluzione ottima del problema primale

$$m = \inf_x \{ f_1(x) + f_2(x) \} \quad (\text{A.13})$$

è un numero finito, allora la soluzione ottima del problema duale è l'opposto di m :

$$-m = \inf_g \{ f_1^*(g) + f_2^*(-g) \}. \quad (\text{A.14})$$

Questo è il duale di Fenchel. Più in generale,

$$\inf_x \left\{ \sum_{i=1}^n f_i(x) \right\} = - \inf_g \left\{ \sum_{i=1}^n f_i^*(g_i) : \sum_{i=1}^n g_i = 0 \right\} \quad (\text{A.15})$$

Il duale Lagrangiano di (A.1.19) si può scrivere sempre al posto di quello di Fenchel di (A.1.21):

consideriamo il primale di Fenchel (A.13) e lo riscriviamo come

$$\inf_{x_1, x_2} \{ f_1(x_1) + f_2(x_2) : x_1 - x_2 = 0 \},$$

e ne scriviamo la Lagrangiana

$$\mathcal{L}(x_1, x_2, \lambda) = f_1(x_1) + f_2(x_2) + \lambda(x_1 - x_2).$$

La funzione duale da minimizzare è

$$\begin{aligned} & - \inf_{x_1, x_2} \mathcal{L}(x_1, x_2, \lambda) \\ &= - \inf_{x_1} [f_1(x_1) + \lambda x_1] - \inf_{x_2} [f_2(x_2) - \lambda x_2] \\ &= f_1^*(-\lambda) + f_2^*(\lambda), \end{aligned}$$

che è esattamente (A.14).

Il duale di Fenchel di (A.1.21) si può scrivere qualche volta al posto di quello di Lagrange di (A.1.19), ad esempio nel caso di $f_2 = g_2(Ax - b)$ con g_2 convessa (tipicamente g_2 è I oppure $\| \cdot \|_2^2$) e A lineare. Il problema primale è della forma

$$\inf_x \{ f_1(x) + g_2(Ax - b) \}. \quad (\text{A.16})$$

Assumendo che esiste $x \in \text{ridom } f_1 : Ax - b \in \text{ridom } g_2$, le soluzioni del problema duale di Fenchel associate con (A.16) sono $s = -A^T \lambda$, dove λ è la soluzione di

$$\min \{ g_2^*(\lambda) - \inf \{ \mathcal{L}(x, \lambda) \} \},$$

con $\mathcal{L}(x, \lambda) = f_1(x) + \lambda(Ax - b)$ la Lagrangiana.

A.2 Notazioni

In questo testo si usano le seguenti notazioni:

- $\beta = \{f(x_i), g_i \in \partial f(x_i)\}$, è il bundle e $\{x_i\}$ è la sequenza di iterate.
- f_β è l'approssimazione di f , costruita a partire dal bundle.
- \hat{f}_β è l'approssimazione dei piani di taglio di f .
- $D_t(d)$ è il termine di stabilizzazione; il parametro t indica la “forza” di D .
- \bar{x} è l'iterata corrente, d è direzione di decrescita e $\bar{x}_+ = \bar{x} + d$ è l'iterata successiva.
- L'errore di linearizzazione tra x_i e \bar{x} è

$$\alpha_i = f(\bar{x}) - f(x_i) - g_i(\bar{x} - x_i) \quad i \in \beta. \quad (\text{A.17})$$

Poiché f è convessa, $\alpha_i \geq 0$ per ogni $i \in \beta$.

Dato che $g_i \in \beta$ sono tali che $g_i \in \partial f(x_i)$ e grazie alla proprietà (vi) della coniugata, l'errore di linearizzazione si può riscrivere nel modo seguente:

$$\begin{aligned} \alpha_i &= f(\bar{x}) - f(x_i) - g_i(\bar{x} - x_i), \\ &= f(\bar{x}) - f(x_i) - g_i\bar{x} + g_ix_i \\ &= f(\bar{x}) - \bar{x}g_i + f^*(g_i), \quad i \in \beta \\ &= f_{\beta, \bar{x}}^*(g) \end{aligned} \quad (\text{A.18})$$

- (Π) è il problema originale astratto di minimizzazione di una funzione convessa.
- $(\Pi_{\bar{x}})$ è la versione traslata di (Π) .
- $(\Pi_{\beta, \bar{x}})$ è la forma concreta di $(\Pi_{\bar{x}})$
- $(\Pi_{\bar{x}, t})$ è la versione stabilizzata di $(\Pi_{\bar{x}})$ ed è ancora un problema astratto.
- $(\Pi_{\beta, \bar{x}, t})$ è la forma concreta di $(\Pi_{\bar{x}, t})$.
- $(\hat{\Pi}_{\beta, \bar{x}, t})$ è la forma implementabile di $(\Pi_{\beta, \bar{x}, t})$.

- (Δ) è la versione duale astratta del problema originale (Π) .
- $(\Delta_{\bar{x}})$ è la versione traslata di (Δ) .
- $(\Delta_{\beta, \bar{x}})$ è la forma concreta di $(\Delta_{\bar{x}})$
- $(\Delta_{\bar{x}, t})$ è la versione stabilizzata di $(\Delta_{\bar{x}})$ ed è ancora un problema astratto.
- $(\Delta_{\beta, \bar{x}, t})$ è la forma concreta di $(\Delta_{\bar{x}, t})$.
- $(\hat{\Delta}_{\beta, \bar{x}, t})$ è la forma implementabile di $(\Delta_{\beta, \bar{x}, t})$.

Bibliografia

- [1] A.FRANGIONI, *Generalized bundle methods*, SIAM J. Optim., Vol. 13, No. 1, (1998), pp. 117–156
- [2] J.E. KELLEY *The Cutting-Plane Method for Solving Convex Programs*, Journal of the SIAM, 8 (1960), pp. 703–712
- [3] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms I - Fundamentals*, Grundlehren Math. Wiss. 305, Springer-Verlag, New York, 1993
- [4] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms II - Advanced Theory and Bundle Methods*, Grundlehren Math. Wiss. 306, Springer-Verlag, New York, 1993.
- [5] C. LEMARÉCHAL, A. NEMIROVSKII AND Y. NESTEROV, *New Variants of Bundle Methods*, Math. Programming, 69 (1995), pp. 111–147
- [6] E. A. NURMINSKI, *Separating Plane Algorithms for Convex Optimization*, Math. Programming, 76 (1997), pp. 373–391
- [7] A. OUOROU, *A Proximal Cutting Plane Method Using Chebychev Center for Nonsmooth Convex optimization*,
- [8] A. OUOROU, *A Proximal Cutting Plane Method Using Chebychev Center for Nonsmooth Convex optimization*, Springer-Verlag, New York, 2008.
- [9] T.ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM J. Control and Optimisation, Vol. 14, No. 5, August (1976)