La Matematica nella Società e nella Cultura

Rivista dell'Unione Matematica Italiana

Enrico Gorgone

Ottimizzazione non-differenziabile: teoria e algoritmi

La Matematica nella Società e nella Cultura. Rivista dell'Unione Matematica Italiana, Serie 1, Vol. **2** (2009), n.2 (Fascicolo Tesi di Dottorato), p. 251–254.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=RIUMI_2009_1_2_2_251_0>

La Matematica nella Società e nella Cultura. Rivista dell'Unione Matematica Italiana, Unione Matematica Italiana, 2009.

La Matematica nella Società e nella Cultura
Rivista dell'Unione Matematica Italiana
Serie I, Vol. II, Agosto 2009, 251-254

Ottimizzazione non-differenziabile: teoria e algoritmi

ENRICO GORGONE

1. – Il Problema

Un problema di ottimizzazione convessa non-differenziabile è del tipo:

$$(\mathcal{P}) = \min f(x) : \quad x \in \mathbb{R}^n,$$

dove $f : \mathbb{R}^n \to \mathbb{R}$ è convessa e non è necessariamente differenziabile.

Lo sviluppo dell'analisi convessa [5] fornì le fondamenta teoriche alla messa a punto di metodi numerici per l'ottimizzazione convessa non-differenziabile. I risultati storicamente più significativi sono: il *metodo del piano di taglio* di E.W. Cheney e A. Goldestein (1959) e, contemporaneamente, di J.E. Kelley (1960) e i *metodi del subgradiente* di N.Z Shor (1961) e di B. Polyak (1967). Questi ultimi estendono l'algoritmo del gradiente nel caso non-differenziabile. Successivamente da alcuni lavori di C. Lemaréchal e, separatamente, di P. Wolfe, furono sviluppati i *metodi bundle* (1975).

Quando la funzione obiettivo f non è differenziabile né convessa, $(\mathcal{P})$ prende il nome di problema di ottimizzazione non-differenziabile non-convessa. L'ottimizzazione non-differenziabile non-convessa nacque adeguando gli algoritmi esistenti, progettati per il caso convesso, al caso non-convesso (vedi, per esempio, Schramm e Zowe, 1992). La letteratura odierna offre invece algoritmi originali, concepiti appositamente per la risoluzione del problema non-convesso (vedi, per esempio, [1, 2]).

La tesi descrive i principali metodi numerici per l'ottimizzazione non-differenziabile, considerando sia il caso convesso e sia il caso non-convesso, e presenta un nuovo metodo di tipo bundle per il caso non-convesso.

2. – Considerazioni Introduttive

La tesi è organizzata come segue. Vengono richiamati nella prima parte gli elementi essenziali dell'analisi convessa, quindi vengono descritti i principali metodi sopra menzionati per il caso convesso e alcune recenti tecniche di "smoothing" (Yu. Nesterov, 2005). Le tecniche di smoothing trasformano classi di problemi ben strutturati non-differenziabili in problemi differenziabili, cosicché per la loro risoluzione è possibile applicare metodi a punto interno specifici dell'ottimizzazione convessa differenziabile. Un'analisi approfondita viene rivolta in special modo ai metodi bundle e alle rispettive proprietà di convergenza (vedi, per esempio, [3]).

Vengono richiamati nella seconda parte alcuni elementi di analisi non-differenziabile e viene presentata, in particolare, una semplice dimostrazione del teorema di H. Rademacher (1919) [4], che spiega come per un'ampia classe di funzioni reali definite su $\mathbb{R}^n$, l'insieme dei punti nei quali la funzione non è differenziabile ha misura nulla nel senso di Lebesgue. In seguito vengono esaminati alcuni algoritmi per il caso non-convesso, in particolare l'algoritmo del *gradiente campionato* (J. Burke, A. Lewis e L. Overton, 2005) e una classe di algoritmi di tipo bundle (A. Fuduli, M. Gaudioso e G. Giallombardo, 2004).

Infine viene presentato un nuovo algoritmo per la risoluzione del problema non-differenziabile non-convesso, che adotta la tecnica di "splitting" usata in [1] e che dà origine a due approssimazioni poliedrali, una convessa e l'altra concava, della funzione obiettivo f. La direzione di ricerca è calcolata, ad ogni iterazione, con l'obiettivo di massimizzare la differenza tra i suddetti modelli; ciò equivale a risolvere un sottoproblema di programmazione quadratica di struttura speciale. Per garantire la convergenza dell'algoritmo ad un punto stazionario si è adottato un approccio di tipo prossimale, che ha consentito di dimostrare che il metodo proposto converge, per ogni scelta del punto iniziale x_0, sotto le ipotesi che l'insieme di livello

$$L_0 \triangleq \{x \in \mathbb{R}^n : \quad f(x) \leq f(x_0)\}$$

sia compatto e f sia debolmente semidifferenziabile (weakly semismooth). L'algoritmo è stato implementato in Fortran ed applicato a 25 problemi test raccolti, disponibili in letteratura.

3. – Ottimizzazione Non-differenziabile: un nuovo metodo di tipo bundle

L'algoritmo, qui presentato, è di tipo bundle [3, 1], in questa sezione viene descritto il cuore della procedura, cioè il sottoproblema quadratico.

Per ogni coppia di punti (x_i, x) l'errore di linearizzazione a_i è definito come la differenza tra il valore effettivo di f in x e l'espansione lineare di f generata in x_i valutata in x:

$$a_i \triangleq f(x) - f(x_i) - g_i^T(x - x_i),$$

con $g_i \in \partial f(x_i)$, dove con $\partial f(\,\cdot\,)$ viene indicato il gradiente generalizzato di Clarke.

Per una generica funzione, l'errore di linearizzazione non è, come nel caso convesso, sempre non negativo. Si è pensato, quindi, di suddividere l'insieme degli indici del bundle I, cioè l'insieme degli indici dei punti generati precedentemente dall'algoritmo, nel seguente modo:

$$I^+ \triangleq \{i \in I : \quad a_i \geq -\sigma\} \quad \text{and} \quad I^- \triangleq \{i \in I : \quad a_i < -\sigma\}, \tag{1}$$

per un assegnato valore $\sigma > 0$. Più in particolare, I^+ contiene i punti che mostrano un tipo di comportamento "convesso" e I^- i punti che mostrano un tipo di comportamento "concavo" rispetto al punto x.

Si definisce una funzione $h(d) \triangleq f(x+d) - f(x)$, chiamata *funzione differenza*, e vengono costruiti due modelli poliedrali di h, usando separatamente i due bundle. Nello specifico vengono definite due funzioni affini a tratti:

$$\Delta^+(d) \triangleq \max_{i \in I^+} \{g_i^T d - a_i\}$$

$$\Delta^-(d) \triangleq \min\left\{0, \ \min_{i \in I^-} \{g_i^T d - a_i\}\right\}.$$

Risulta evidente che Δ^+ è convessa mentre Δ^- è concava.

TABELLA 1. – Risultati Computazionali

Problem				NCNS		NCVX	
#	Problem	n	f^*	N_f	f	N_f	f
1	Rosenbrock	2	0	54	5.137e-06	70	5.009e-07
2	Crescent	2	0	53	5.112e-06	22	8.022e-06
3	CB2	2	1.9522245	16	1.9522255	18	1.9522245
4	CB3	2	2	14	2.0000000	15	2.0000000
5	DEM	2	-3	13	-3.0000000	21	-2.9999999
6	QL	2	7.2	15	7.2000001	28	7.2000005
7	LQ	2	-1.4142136	15	-1.4142136	9	-1.4142135
8	Mifflin1	2	-1	165	-0.9999389	127	-0.9999977
9	Mifflin2	2	-1	14	-0.9999991	13	-1.0000000
10	Rosen-Suzuki	4	-44	33	-43.999997	29	-44.000000
11	Shor	5	22.600162	27	22.600163	44	22.600162
12	Maxquad	10	-0.8414083	90	-0.8413860	56	-0.8414078
13	Maxq	20	0	187	1.561e-06	293	1.660e-07
14	Maxl	20	0	23	4.493e-15	44	1.110e-15
15	Goffin	50	0	56	1.984e-13	148	1.142e-13
16	El-Attar	6	0.5598131	172	0.5598143	152	0.5598163
17	Wolfe	2	-8	43	-7.9999998	21	-7.9999998
18	MXHILB	50	0	24	1.764e-05	33	1.768e-05
19	L1HILB	50	0	30	1.709e-05	104	6.978e-07
20	Colville1	5	-32.348679	36	-32.348677	47	-32.348679
21	Gill	10	9.7857721	308	9.7858381	164	9.7857746
22	HS78	5	-2.9197004	237	-2.9191783	159	-2.9196589
23	TR48	48	-638565	1662	-638514.80	353	-638565.00
24	Shell Dual	15	32.348679	642	32.348687	1497	32.349404
25	Steiner2	12	16.703838	96	16.703844	196	16.703838

La direzione di ricerca d^* è determinata risolvendo il seguente problema convesso:

$$(\mathcal{PQ}) = \min_{d \in \mathbb{R}^n} \Delta(d) + \frac{1}{2}\rho\|d\|^2$$

dove $\Delta \triangleq \Delta^+ - \Delta^-$ e $\rho > 0$ è il parametro di prossimità introdotto per garantire la soluzione di $(\mathcal{PQ})$ e per evitare problemi di instabilità. La logica del modello è di collocare il "nuovo punto" $x + d^*$ in maniera che entrambi i modelli Δ^+ e Δ^- predicano una riduzione del valore della funzione obiettivo, ma, allo stesso tempo, le loro predizioni siano il più possibile discordi.

Nella tesi è provata la terminazione dell'algoritmo, ad un punto di ottimo approssimato, sotto le ipotesi che f sia localmente di Lipschitz e debolmente semidifferenziabile.

Nella tabella 1 sono riportati i risultati computazionali in termini del numero N_f di valutazioni di funzione, indicando con f^* e f, rispettivamente, il minimo della funzione obiettivo e il valore di funzione determinato dall'algoritmo quando il test di stop risulta soddisfatto. Le prestazioni del codice *NCNS* del nuovo algoritmo vengono confrontate con quelle dell'algoritmo *NCVX* di [1].

BIBLIOGRAFIA

[1] FUDULI A., GAUDIOSO M. e GIALLOMBARDO G., *Minimizing nonconvex nonsmooth functions via cutting planes and proximity control*, SIAM Journal on Optimization, **14** (2004), 743-756.

[2] GAUDIOSO M., GORGONE E. e MONACO M. F., *Piecewise linear approximation in nonconvex nonsmooth optimization*, Numerische Mathematik, accettato per la pubblicazione

[3] HIRIART-URRUTY J. B. e LEMARÉCHAL C., *Convex analysis and minimization algorithms Vol. II* (Springer-Verlag,1993).

[4] NEKVINDA A. e ZAJÍČEK L., *A simple proof of the Rademacker Theorem*, Časopis Pro Pěstování Matematiky, 4 (1988), 337-341.

[5] ROCKAFELLAR R. T., *Convex analysis*, Princeton University Press (1970).

Dipartimento di Elettronica Informatica e Sistemistica, Università della Calabria
e-mail: egorgone@deis.unical.it
Dottorato in Ricerca Operativa
con sede presso l'Università della Calabria (Rende) - Ciclo XX
Direttore di Ricerca: Proff. Manlio Gaudioso (Università della Calabria),
Maria Flavia Monaco (Università della Calabria)