La Matematica nella Società e nella Cultura

Rivista dell'Unione Matematica Italiana

Margherita Porcelli

Metodi trust-region a convergenza quadratica per problemi ai minimi quadrati non lineari con vincoli semplici e problemi di ammissibilitaÁ non lineari

La Matematica nella Società e nella Cultura. Rivista dell'Unione Matematica Italiana, Serie 1, Vol. **4** (2011), n.1 (Fascicolo Tesi di Dottorato), p. 67–70.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=RIUMI_2011_1_4_1_67_0>

La Matematica nella Società e nella Cultura. Rivista dell'Unione Matematica Italiana, Unione Matematica Italiana, 2011.

La Matematica nella Società e nella Cultura
Rivista dell'Unione Matematica Italiana
Serie I, Vol. IV, Aprile 2011, 67-70

Metodi trust-region a convergenza quadratica per problemi ai minimi quadrati non lineari con vincoli semplici e problemi di ammissibilità non lineari

MARGHERITA PORCELLI

1. – I problemi trattati

Consideriamo la risoluzione numerica di due classi di problemi di ottimizzazione strettamente correlati: problemi ai minimi quadrati nonlineari con vincoli semplici e problemi di ammissibilità non lineari.

La formulazione generale della prima classe di problemi è la seguente

$$\min_{x \in \Omega} \theta(x) = \frac{1}{2} \|\Theta(x)\|_2^2, \tag{1}$$

dove la funzione obiettivo $\theta : \mathbb{R}^n \to \mathbb{R}$ e la funzione residuo $\Theta : \mathbb{R}^n \to \mathbb{R}^m$ è una funzione differenziabile con continuità. L'insieme dei vincoli Ω è definito come segue

$$\Omega = \{x \in \mathbb{R}^n \,|\, l \le x \le u\}, \tag{2}$$

con $l \in (\mathbb{R} \cup -\infty)^n, u \in (\mathbb{R} \cup \infty)^n, l < u$. Le disuguaglianze in (2) sono dette vincoli di forma "semplice" o "box" e sono intese componente a componente. Alcune componenti di x possono essere illimitate inferiormente e/o superiormente.

È interessante notare che se il problema ai minimi quadrati (1) ammette una soluzione a residuo zero, ovvero una soluzione che annulla la funzione θ, allora tale soluzione risolve il sistema di equazioni $\Theta(x) = 0$, $x \in \Omega$. Viceversa, il sistema rettangolare di equazioni con vincoli semplici

$$F(x) = 0, \quad x \in \Omega, \tag{3}$$

con $F : \mathbb{R}^n \to \mathbb{R}^m$ differenziabile con continuità, può essere riformulato come problema (1) ponendo $\Theta = F$.

La seconda classe di problemi trattati consiste nei problemi di ammisibilità non lineari formulati come sistemi di equazioni e disequazioni non lineari della forma

$$\begin{aligned} C_E(v) &= 0, \\ C_I(v) &\le 0, \\ v_l \le v &\le v_u, \end{aligned} \tag{4}$$

dove le funzioni $C_E : \mathbb{R}^p \to \mathbb{R}^{m_E}$ e $C_I : \mathbb{R}^p \to \mathbb{R}^{m_I}$ sono differenziabili con continuità, $v_l \in (\mathbb{R} \cup -\infty)^p$, $v_u \in (\mathbb{R} \cup \infty)^p$.

La presenza dei vincoli semplici sulle variabili nei problemi (1), (3) e (4) ha grande rilevanza applicativa poiché permette di limitare la ricerca delle soluzioni ad una specifica regione di $\mathbb{R}^n$ tenendo conto del loro significato fisico ed eventualmente del dominio delle funzioni. In pratica sono molto frequenti i vincoli di non negatività.

Data la generalità dei problemi trattati, le principali difficoltà teoriche e numeriche nella loro risoluzione sono dovute al fatto che, in molti casi, questi problemi non ammettono soluzione isolata. Infatti, è importante notare che non poniamo alcuna restrizione sul rapporto tra le dimensioni m ed n in (1) e (3) e le dimensioni m_E, m_I e p in (4).

La risoluzione dei problemi indicati è richiesta in vari contesti applicativi. Infatti, il problema ai minimi quadrati (1) descrive problemi di identificazione dei parametri di un sistema il cui stato è noto mediante misure sperimentali, mentre i sistemi di equazioni e disequazioni (4) descrivono il calcolo di punti ammissibili per problemi di ottimizzazione vincolata con vincoli generali, il calcolo degli zeri di una funzione non lineare e non quadrata all'interno di una specificata regione dello spazio ($m_I = 0$ in (4)), il problema di "feasibility restoration" nella recente tecnica di globalizzazione detta metodo dei filtri e proposta da R. Fletcher e S. Leyffer.

Risolvere il problema (4) consiste nel determinare un vettore $v \in \mathbb{R}^p$ che soddisfi le equazioni e disequazioni. Se tale vettore non può essere trovato, l'obiettivo consiste nel minimizzare la somma delle violazioni dei vincoli in (4). La soluzione del problema (4) è quindi tipicamente affrontata riformulandolo come un problema ai minimi quadarti non lineari.

Le riformulazioni proposte variano a seconda delle proprietà della funzione obiettivo e dalla possibile presenza di vincoli semplici. In letteratura, usando l'equivalenza tra $t \leq 0$ e $max\{t, 0\} = 0$, vengono presentate riformulazioni del problema (4) come problema ai minimi quadrati non vincolato in cui la funzione residuo non è differenziabile con continuità. Al contrario, noi proponiamo riformulazioni che preservano la regolarità del problema e mantengono i vincoli semplici. In particolare, le disuguaglianze di tipo box sono distinte da quelle di forma generale $C_I(v) \leq 0$ e queste ultime sono ricondotte ad uguaglianze mediante opportune trasformazioni in modo da ottenere un problema della forma (1). Nelle trasformazioni da noi usate, disuguaglianze del tipo $t \leq 0$, vengono sostitutite da $s - t = 0, s \leq 0$ oppure da $max\{t, 0\}^2/2 = 0$. Notiamo che la riformulazione (1) consente di sfruttare algoritmi per problemi di ottimizzazione con vincoli semplici che assicurano il calcolo della funzione obiettivo solo in punti ammissibili, cioè in punti di Ω.

Nella Tesi consideriamo il problema (1) come problema unificante per i problemi trattati e studiamo nuovi metodi numerici per la sua risoluzione. In particolare, il nostro contributo consiste nella definizione di nuovi procedimenti, nell'analisi delle loro proprietà teoriche e nella loro validazione sperimentale. Inoltre, uno degli algoritmi proposti viene implementato in un codice Matlab di dominio pubblico.

2. – Metodi numerici proposti

Nel lavoro di Tesi presentiamo algoritmi per la risoluzione di problemi non lineari della forma (1) che appartengono alla classe dei metodi di tipo trust-region con scalatura affine.

Pertanto, i metodi proposti sono iterativi e generano una successione di approssimazioni $\{x_k\}$ ammissibili.

L'idea di base di un procedimento di tipo trust-region, consiste nel definire all'iterazione k-sima un modello quadratico per θ in un intorno dell'iterata corrente x_k e una regione di affidabilità ("trust-region") in cui si confida che il modello rappresenti adeguatamente la funzione obiettivo θ. Il passo di prova p_{tr} per correggere l'iterata corrente, viene calcolato minimizzando, eventualmente in maniera approssimata, il modello quadrato all'interno della trust-region. Il problema di trust-region ha in generale, la seguente forma

$$\min \{m_k(x_k + p) \,:\, \|p\|_2 \leq \Delta_k\}, \tag{5}$$

dove m_k è un modello quadratico per θ e Δ_k è uno scalare positivo detto "raggio di trust-region" che definisce la regione di affidabilità. Una volta risolto il problema (5), il passo viene accettato se viene misurato un sufficiente accordo tra la funzione θ e il modello m_k. In tal caso, l'iterata corrente viene aggiornata e il raggio di trust-region aumentato; altrimenti il passo viene scartato e il raggio di trust-region viene ridotto. Un ingrediente fondamentale nella strategia di trust-region è dato dal cosiddetto passo di Cauchy definito come il minimizzatore del modello m_k lungo la direzione del gradiente soggetto al vincolo di trust-region. La sua importanza è data dal fatto che per ottenere la convergenza globale della successione $\{x_k\}$, cioè convergenza che non dipende dalla scelta dell'approssimazione iniziale x_0, è sufficiente garantire che il passo di trust-region produca una decrescita sufficiente nel modello m_k rispetto alla decrescita ottenuta con il passo di Cauchy.

Nei procedimenti presentati nelle Tesi, modifichiamo la strategia classica di trust-region, utilizzando un'opportuna matrice di scalatura per tener conto della presenza dei vincoli.

Tale matrice è definita osservando che le condizioni di ottimalità del prim'ordine per il problema (1) possono essere formulate come il seguente sistema non lineare

$$D(x)\nabla\theta(x) = D(x)\,J(x)^T\,\Theta(x) = 0, \tag{6}$$

dove J rappresenta la matrice Jacobiana di Θ e la matrice di scalatura $D(x) = diag(|v_1(x)|, \ldots, |v_n(x)|)$ è tale che

$$v_i(x) = \begin{cases} x_i - u_i & \text{se } \ (\nabla\theta(x))_i < 0, \ u_i < \infty, \\ x_i - l_i & \text{se } \ (\nabla\theta(x))_i \geq 0, \ l_i > -\infty, \\ 1 & \text{se } \ (\nabla\theta(x))_i \geq 0, \ l_i = -\infty \ \text{ or } \ (\nabla\theta(x))_i < 0, \ u_i = \infty, \end{cases}$$

per $i = 1, \ldots, n$. Una delle più importanti proprietà della matrice D consiste nel rendere la direzione di più ripida discesa $(-\nabla\theta(x_k))$ ben angolata rispetto ai bordi di Ω. Infatti, la direzione di più ripida discesa scalata di θ in x_k

$$d_k = -D(x_k)\nabla\theta(x_k),$$

consente un più ampio passo all'interno di Ω rispetto alla direzione non scalata $(-\nabla\theta(x_k))$.

L'ammisibilità delle iterate è imposta calcolando la proiezione del passo di trust-region p_{tr} sull'insieme Ω, i.e. calcolando $\bar{p}_{tr} = \max\{l, \min\{x_k + p_{tr}, u\}\} - x_k$. Per ottenere la convergenza globale dei procedimenti, si impone che il passo per aggiornare x_k produca una riduzione in m_k almeno pari ad una frazione fissata della riduzione raggiunta dal passo di Cauchy Generalizzato definito come

$$p_k^C \stackrel{\text{def}}{=} \underset{p \in span\{d_k\}}{\text{argmin}} \; m_k(x_k + p) \quad \text{tale che} \quad \|p\|_2 \leq \Delta_k, \quad x_k + p \in \Omega.$$

Gli algoritmi proposti si diversificano nel tipo di modello quadratico m_k utilizzato e nella procedura impiegata per calcolare una soluzione approssimata di (5). Si considera infatti un modello della forma

$$m_k(x_k + p) = \frac{1}{2}\|J(x_k)\,p + \Theta(x_k)\|_2^2 + \frac{1}{2}\mu_k\|p\|_2^2,$$

dove μ_k è uno scalare non negativo, che rappresenta un modello detto di tipo Gauss-Newton se $\mu_k = 0, k \geq 0$, oppure un modello di tipo Gauss-Newton regolarizzato se $\mu_k > 0, k \geq 0$.

L'implementazione della strategia di trust-region, coinvolge una fase di algebra lineare. In tale fase, gioca un ruolo cruciale per l'analisi di convergenza locale, il cosiddetto passo di minima norma, cioè il minimizzatore globale del modello m_k di norma euclidea minima. Tenendo in considerazione le proprietà di questo passo, discutiamo sia l'uso di procedimenti diretti sia l'uso di procedimenti iterativi tipo Gradienti Coniugati adatti alla risoluzione di problemi di grandi dimensioni.

Per tutte le implementazione proposte, proviamo proprietà di convergenza globale ad un punto stazionario per il problema (1), i.e. ad un punto che soddisfa le condizioni (6). L'analisi delle proprietà di convergenza locale viene svolta individuando condizioni che generalizzano le classiche ipotesi per problemi con soluzione isolata e garantiscono la convergenza asintoticamente veloce della successione ad una soluzione x^* tale che $\Theta(x^*) = 0$.

Infine, il metodo di tipo Gauss-Newton che sfrutta l'algebra lineare diretta nella soluzione di (5), viene implementato in un codice Matlab che gestisce in modo automatico i sistemi di equazioni e disequazioni non lineari (4) e per questo prende il nome di `TRESNEI` (Trust-REgion Solver for Nonlinear Equalities and Inequalities). `TRESNEI` viene ampiamente testato sui problemi applicativi ben noti in letteratura e confrontato con funzioni del Toolbox di Ottimizzazione di Matlab. In particolare, si rivela robusto ed efficiente ed ha una applicabilià superiore rispetto ai codici attualmente disponibili in ambiente Matlab. `TRESNEI` è disponibile all'indirizzo `http://tresnei.de.unifi.it`.

naXys - Department of Mathematics,
FUNDP - University of Namur, Namur, Belgio
e-mail: margherita.porcelli@math.unifi.it
Dottorato in Matematica
con sede presso l'Università di Firenze – Ciclo XXII
Direttore di ricerca: Prof.ssa Benedetta Morini, Università di Firenze