3 – Metodi iterativi

M. Bertero – DISI – Università di Genova

- Metodo di Landweber
- Metodo di Landweber proiettato
- Metodo della massima pendenza
- Metodo del gradiente coniugato

Nel caso del problema di deconvoluzione, sia il funzionale ai minimi quadrati che le sue forme regolarizzate possono essere facilmente diagonalizzati mediante DFT. I problemi possono diventare troppo onerosi dal punto di vista computazionale nel caso di una matrice arbitraria e di grandi dimensioni. Tuttavia alcuni metodi iterativi hanno interesse anche nel caso del problema di deconvoluzione per due motivi: 1) questi metodi iterativi hanno proprietà regolarizzanti, nel senso che si ottengono soluzioni sensate se si arrestano opportunamente le iterazioni; 2) possono essere utilizzati per minimizzare il funzionale ai minimi quadrati con vincoli aggiuntivi quali la n0n-negatività della soluzione.

3.1 – Metodo di Landweber

Il metodo di Landweber è il piu' semplice tra i metodi iterativi per la minimizzazione del funzionale ai minimi quadrati. Si possono infatti studiare facilmente le sue proprietà di convergenza nonchè comprendere le sue proprietà regolarizzanti. Inoltre può essere facilmente generalizzato alla risoluzione di problemi ai minimi quadrati con vincoli aggiuntivi, quali la non-negatività o altri.

Si ricorda che un metodo iterativo per la minimizzazione di un funzionale ha la seguente struttura: è definito da un operatore, lineare o non, T nel modo seguente:

$$i)\ \text{si assegna un'approssimazione iniziale } f^{(0)};$$

$$ii)\ \text{per } j = 0,1,2,\ldots, \text{ data l'approssimazione } f^{(j)}, \text{ si calcola}$$

$$\text{l'approssimazione successiva } f^{(j+1)} \text{ mediante la regola:}$$

$$f^{(j+1)} = T\left(f^{(j)}\right).$$

L'approssimazione $f^{(j)}$ viene anche detta l' iterata di ordine j.

Il metodo di Landweber è un metodo di discesa e, piu' precisamente, un metodo di tipo gradiente: data l'iterata $f^{(j)}$, l'iterata successiva si ottiene muovendosi di un certo passo in direzione opposta a quella del gradiente, cioè nella direzione di massima pendenza.

Si ricordi che la variazione prima del funzionale ai minimi quadrati, che non è altro che il quadrato del funzionale discrepanza, è data da:

$$\varepsilon^2(f+h;g) = \varepsilon^2(f;g) + 2\left(A^T Af - A^T g, h\right) + \|Ah\|^2 \quad .$$

Pertanto il gradiente del funzionale ha la forma seguente:

$$\nabla_f \varepsilon^2(f;g) = 2\left(A^T Af - A^T g\right) . \qquad (3.1)$$

Se a partire dall'iterata $f^{(j)}$ ci si muove di un passo $\tau/2$ in direzione opposta al gradiente si ottiene il seguente schema iterativo:

$$f^{(j+1)} = f^{(j)} + \tau\left(A^T g - A^T Af^{(j)}\right) \quad . \qquad (3.2)$$

Osservazione - Tale relazione definisce l'operatore T che è quindi la somma di un operatore lineare e di una traslazione.

Tenendo conto dell'osservazione precedente, scriviamo la formula iterativa nel modo seguente:

$$f^{(j+1)} = \tau\ A^T g + \left(I - \tau\ A^T A\right) f^{(j)} \quad . \qquad (3.3)$$

A partire da questa equazione scriviamo le prime iterate:

$$f^{(0)} \quad , \quad f^{(1)} = \tau\, A^T g + \left(I - \tau\, A^T A\right) f^{(0)} \,,$$

$$f^{(2)} = \tau\, A^T g + \left(I - \tau\, A^T A\right) A^T g + \left(I - \tau\, A^T A\right)^2 f^{(0)} \quad ,$$

$$f^{(3)} = \tau\, A^T g + \left(I - \tau\, A^T A\right) A^T g + \left(I - \tau\, A^T A\right)^2 A^T g + \left(I - \tau\, A^T A\right)^3 f^{(0)} \quad ,$$

.....

Da queste si può arguire che la generica iterata di ordine j deve avere la forma seguente:

$$f^{(j)} = \tau \sum_{l=0}^{j-1} \left(I - \tau\ A^T A\right)^l A^T g + \left(I - \tau\ A^T A\right)^j f^{(0)} \quad . \qquad (3.4)$$

La dimostrazione può essere fatta per induzione: si ammette che la formula sia vera per l'iterata di ordine k e si dimostra che allora è vera anche per l'iterata di ordine k+1. Infatti, sostituendo la (3.4) nella (3.3), si ottiene:

$$f^{(j+1)} = \tau A^T g + \left(I - \tau A^T A\right)\left\{\sum_{l=0}^{j-1}\left(I - \tau A^T A\right)^l A^T g + \left(I - \tau A^T A\right)^j f^{(0)}\right\} =$$

$$= \tau \sum_{l=0}^{j}\left(I - \tau A^T A\right)^l A^T g + \left(I - \tau A^T A\right)^{j+1} f^{(0)} \quad ,$$

che è quanto dovevasi dimostrare.

L'equazione (3.4) permette di ottenere una forma chiusa generale per l'iterata di ordine j; tuttavia, per semplicità, ci limitiamo al caso del problema di deconvoluzione. Quindi, scrivendo la (3.4) in termini di DFT, si ottiene:

$$\hat{f}^{(j)}(\underline{k}) = \tau \sum_{l=0}^{j-1}\left(1 - \tau\,|\hat{K}(\underline{k})|^2\right)^l \hat{K}^*(\underline{k})\hat{g}(\underline{k}) + \left(1 - \tau\,|\hat{K}(\underline{k})|^2\right)^j \hat{f}^{(0)}(\underline{k}) \ . \quad (3.5)$$

La somma può essere calcolata mediante la solita espressione della ridotta della serie geometrica.

Dopo alcune semplici manipolazioni algebriche, il risultato è il seguente:

$$\hat{f}^{(j)} = \left\{ 1 - \left(1 - \tau \, | \hat{K}(\underline{k}) |^2 \right)^j \right\} \frac{\hat{g}(\underline{k})}{\hat{K}(\underline{k})} + \left(1 - \tau \, | \hat{K}(\underline{k}) |^2 \right)^j \hat{f}^{(0)} . \quad (3.6)$$

Tale equazione è stata ricavata per i pixels in cui la funzione di trasferimento è diversa da zero, cioè per i pixels interni alla banda. Tuttavia, mediante passaggio al limite, si riconosce che essa è vera anche al di fuori della banda. Infatti in tali pixels, come si ricava direttamente dalla (3.5), si ha:

$$\hat{f}(\underline{k}) = \hat{f}^{(0)}(\underline{k}) \qquad , \qquad \underline{k} \notin B \quad . \qquad (3.7)$$

Un metodo iterativo ha senso se esso è convergente. Dalla (3.6) è facile dedurre che il metodo di Landweber produce una successione convergente se e solo se:

$$-1 < 1 - \tau \, | \hat{K}(\underline{k}) |^2 < 1 \qquad .$$

Questa è una condizione sul passo τ, che viene anche detto parametro di rilassamento.

Dalle condizioni precedente segue che:

$$0 < \tau < \frac{2}{|\hat{K}(\underline{k})|^2} \quad ,$$

per ogni pixel appartenente alla banda. Ciò è vero per tutti i pixels se e solo se τ soddisfa alle condizioni seguenti:

$$0 < \tau < \frac{2}{|\hat{K}_{\max}|^2} \quad , \quad \hat{K}_{\max} = \max_{\underline{k} \in B} |\hat{K}(\underline{k})| \quad , \qquad (3.8)$$

che pertanto sono condizioni necessarie e sufficienti per la convergenza del metodo di Landweber.

Dalle equazioni (3.6) e (3.7), passando al limite per j che tende all'infinito, si ottiene:

$$\lim_{j \to \infty} \hat{f}^{(j)}(\underline{k}) = \begin{cases} \dfrac{\hat{g}(\underline{k})}{\hat{K}(\underline{k})} & , \quad \underline{k} \in B \quad ; \\ \hat{f}^{(0)}(\underline{k}) & , \quad \underline{k} \notin B \quad . \end{cases} \qquad (3.9)$$

Dalla (3.7) segue che le iterazioni di Landweber non modificano i valori fuori banda del'oggetto usato per inizializzare il processo; questi valori vengono ovviamente ritrovati nel limite delle iterazioni. Tranne che in casi del tutto eccezzionali, non si sa nulla dei valori fuori banda dell'oggetto da ricostruire. Pertanto è buona norma non fare alcuna ipotesi su di essi; il risultato è che il metodo di Landweber viene solitamente inizializzato con un oggetto identicamente nullo. In tal caso, dalla (3.9) segue che:

$$\lim_{j\to\infty} f^{(j)}(\underline{n}) = f^{+}(\underline{n}) \qquad . \qquad (3.10)$$

Inoltre dalle (3.6) e (3.7) si vede che l'iterata di ordine j è una versione filtrata della soluzione generalizzata, con un filtro in frequenze dato da:

$$\hat{K}^{(j)}(\underline{k}) = 1 - \left(1 - \tau\,|\hat{K}(\underline{k})|^2\right)^j \qquad , \qquad (3.11)$$

da confrontarsi con il filtro di Tikhonov, precedentemente introdotto:

$$\hat{K}_{\mu}(\underline{k}) = \frac{|\hat{K}(\underline{k})|^2}{|\hat{K}(\underline{k})|^2 + \mu} \qquad . \qquad (3.12)$$

Si possono confrontare le proprietà dei due filtri, rispettivamente con j fissato e μ fissato. Si ha:

$i)$ per $\underline{k} = \underline{0}$, tenendo conto della normalizzazione della PSF :

$$\hat{K}^{(j)}(\underline{0}) = 1 - (1-\tau)^j$$

$$\hat{K}_{\mu}(\underline{0}) = \frac{1}{1+\mu}$$

e quindi entrambi tendono a 1, rispettivamente per $j \to \infty$ e $\mu \to 0$

(grazie alla normalizzazione della PSF si ha $0 < \tau < 2$).

ii) per valori di $\underline{k}$ per i quali $\hat{K}(\underline{k})$ è piccolo, si ha :

$$\hat{K}^{(j)}(\underline{k}) \cong j\tau \,|\, \hat{K}(\underline{k})|^2 \qquad ,$$

$$\hat{K}_{\mu}(\underline{k}) \cong \frac{1}{\mu} \,|\, \hat{K}(\underline{k})|^2 \qquad .$$

Pertanto i due filtri hanno un comportamento simile sia a basse che a alte frequenze.

Le precedenti osservazioni suggeriscono che il numero di iterazioni si comporti come un parametro di regolarizzazione o, piu' esattamente, che il prodotto j τ si comporti come l'inverso del parametro di regolarizzazione. Questa osservazione può essere resa piu' quantitativa se, in analogia con quanto fatto nel caso delle soluzioni regolarizzate alla Tikhonov, si pone:

$$f^{(j)} = R^{(j)} g \qquad , \qquad g = A\overline{f} + w \qquad\qquad (3.13)$$

e si definisce un errore di ricostruzione come segue:

$$\rho^{(j)} = \left\| f^{(j)} - \overline{f} \right\| \quad .$$

Dalla diseguaglianza triangolare si ottiene nuovamente:

$$\rho^{(j)} \leq \left\| \left(R^{(j)} A - I \right) g \right\| + \left\| R^{(j)} w \right\| = \rho_{appro}^{(j)} + \rho_{noise}^{(j)} \quad ,$$

e quindi l'errore di ricostruzione può essere maggiorato dalla somma di un errore di approssimazione e di un errore di propagazione del noise. Si tratta di studiare il loro comportamento in funzione del numero di iterazioni.

Se si tiene conto che $R^{(j)}A$ è la matrice circolante associata al filtro definito nella (3.11), che si ha quindi $R^{(j)}A=K^{(j)}$, utilizzando l'eguaglianza di Parseval si ottiene:

$$\left(\rho_{appro}^{(j)}\right)^2 = \left\|\left(R^{(j)}A - I\right)g\right\|^2 = \frac{1}{N^2}\sum_{\underline{k}}\left(1-\tau\,|\hat{K}(\underline{k})|^2\right)^{2j}\,|\hat{\overline{f}}(\underline{k})|^2 =$$

$$= \frac{1}{N^2}\sum_{\underline{k}\notin B}|\hat{\overline{f}}(\underline{k})|^2 + \frac{1}{N^2}\sum_{\underline{k}\in B}\left(1-\tau\,|\hat{K}(\underline{k})|^2\right)^{2j}\,|\hat{\overline{f}}(\underline{k})|^2 \quad .$$

Pertanto l'errore di aprossimazione è una funzione decrescente del numero di iterazioni, con valori compresi nell'intervallo:

$$\left\|\overline{f}\right\| \geq \rho_{appro}^{(j)} \geq \left\|\overline{f}_{out}\right\| \quad .$$

Si noti che l'intervallo dei valori coincide con quello che si era trovato per l'errore di approssimazione nel caso della soluzione di Tikhonov. Tuttavia quello era una funzione crescente del parametro di regolarizzazione, in accordo con l'osservazione precedentemente fatta.

Analogamente, per l'errore di propagazione del noise si ha:

$$\left(\rho_{noise}^{(j)}\right)^2 = \frac{1}{N^2}\sum_{\underline{k}\in B}\left|1-\left(1-\tau\,|\,\hat{K}(\underline{k})|^2\right)^j\right|^2\left|\frac{\hat{w}(\underline{k})}{\hat{K}(\underline{k})}\right|^2 \quad .$$

Pertanto l'errore di propagazione del noise è una funzione crescente del numero di iterazioni, con valori compresi nell'intervallo:

$$0 \le \rho_{noise}^{(j)} \le \left\|w^+\right\| \quad .$$

Ne risulta che l'errore di ricostruzione, in funzione del numero di iterazioni, è maggiorato dalla somma di una funzione crescente e di una decrescente. Pertanto esso avrà un minimo. Più esattamente, alle prime iterazioni l'errore decresce, passa per un valore minimo e poi ricresce, tendendo a valori molto grandi dovuti alla propagazione del noise. Esiste quindi un valore ottimale, j_{opt} del numero di iterazioni. Tale comportamento è detto semiconvergenza.

3.2 – Metodo di Landweber proiettato

Il metodo di Landweber, quando applicato al problema di deconvoluzione, non è un vero e proprio metodo iterativo poichè, come si è visto, è possibile trovare una forma chiusa dell'iterata di ordine j, la quale, a sua volta, può essere facilmente calcolata mediante la DFT. Il risultato è che l'iterata di ordine j fornisce un filtraggio della soluzione generalizzata.

Tuttavia, una caratteristica importante del metodo di Landweber è che esso può essere facilmente modificato per risolvere problemi ai minimi quadrati con vincoli opportuni. Uno di tali problemi è già stato studiato: determinare la soluzione ai minimi quadrati che soddisfa al vincolo:

$$\| f \| \leq E \quad ,$$

che si trova cioè all'interno di un insieme sferico di raggio E assegnato. Si ricordi che la soluzione si trova sulla frontiera. Un altro vincolo, che risulta importante in ricostruzione di immagini, quello di non-negatività:

$$f \geq 0 \quad ,$$

ossia si cerca una soluzione ai minimi quadrati appartenente al cono degli oggetti non-negativi.

I due insiemi precedentemente ricordati sono casi particolari di insiemi chiusi e convessi.

Definizione 1 – Un insieme C si dice convesso, se dati due qualsiasi elementi $f_1$ed f_2 appartenenti all'insieme, allora appartiene all'insieme anche il ''segmento di retta'' che li congiunge, ossia:

$$f = \lambda\, f_1 + (1-\lambda) f_2 \in C \ , \ \forall \ \lambda \in [0,1] \quad .$$

E' facile verificare che i due insiemi precedentemente introdotti sono convessi. Infatti, nel primo caso, grazie alla diseguaglianza triangolare, si ha:

$$\|f\| \le \lambda \|f_1\| + (1-\lambda)\|f_2\| \le \lambda\, E + (1-\lambda)\, E = E \quad .$$

Quanto al secondo caso, è del tutto ovvio che:

$$f_1 \ge 0 \ , \ f_2 \ge 0 \ \Rightarrow \ f \ge 0 \quad .$$

E' anche facile verificare che i due insiemi in questione sono chiusi.

Definizione 2 – Dato un insieme chiuso e convesso C ed un oggetto f, dicesi proiezione convessa di f su C ogni elemento di C, f_C , tale che:

$$f_C = \arg\min_{h \in C} \|h - f\| \quad .$$

Se f appartiene a C, è ovvio che esso è la proiezione di se stesso. Se non appartiene, vale il seguente risultato:

Per ogni f esiste un'unica proiezione convessa di f su C. Se f non appartiene a C allora la sua proiezione sta sulla frontiera di C.

Esempio 1 - Nel caso dell'insieme sferico di raggio E è facile verificare che, se f non appartiene ad esso, allora la proiezione convessa di f è data da:

$$f_C = \frac{E}{\|f\|} f \quad . \qquad (3.14)$$

Infatti, si osservi preliminarmente che:

$$\|f_C - f\| = \left\| \left(\frac{E}{\|f\|} - 1 \right) f \right\| = \|f\| - E \quad .$$

Inoltre, dalla diseguaglianza di Schwarz, segue che, per ogni h con ||h||=E, si ha:

$$\|h-f\|^2 = \|h\|^2 + \|f\|^2 - 2(h,f) = E^2 + \|f\|^2 - 2(h,f) \geq$$
$$\geq E^2 + \|f\|^2 - 2\|h\|\|f\| = (\|f\| - E)^2 \quad ,$$

con il segno eguale che vale solo quando h è parallelo a f, cioè h = α f.

Esempio 2 – Nel caso del cono degli oggetti non-negativi si ha:

$$f_C(\underline{n}) = \begin{cases} f(\underline{n}) & , \quad \text{se} \quad f(\underline{n}) \geq 0 \quad ; \\ 0 & , \quad \text{se} \quad f(\underline{n}) \leq 0 \quad . \end{cases} \qquad (3.15)$$

Basta infatti osservare che:

$$\|h-f\|^2 = \sum_{f(\underline{n})\geq 0} |h(\underline{n}) - f(\underline{n})|^2 + \sum_{f(\underline{n})\leq 0} \left|h(\underline{n}) + |f(\underline{n})|\right|^2 \quad ,$$

per verificare che il minimo è dato dall’espressione precedente.

L'esistenza ed unicità della proiezione convessa permette di definire un operatore di proiezione convessa associato all'insieme chiuso e convesso C e dato da:

$$P_C \; f = f_C \quad .$$

Tale operatore è, in genere, non lineare, a meno che l'insieme in questione non sia un sottospazio ortogonale, nel qual caso l'operatore di proiezione convessa coincide con il solito operatore di proiezione ortogonale.

Dato un insieme chiuso e convesso C, si possono ora considerare soluzioni ai minimi quadrati vincolate a C, cioè gli oggetti che minimizzano su C il funzionale ai minimi quadrati; esse sono quindi date da:

$$f_{LS}^{(C)} = \arg \min_{f \in C} \left\| Af - g \right\| \quad . \qquad (3.16)$$

Nel caso dell'insieme dell'Esempio 1, questo problema coincide con quello già studiato con il metodo dei moltiplicatori di Lagrange e di cui si è visto che ammette una ed una sola soluzione.

Più in generale è stato dimostrato che, qualunque sia l'insieme chiuso e convesso C, il seguente algoritmo iterativo:

$$\begin{cases} f_C^{(0)} \in C \quad , \quad \text{assegnato} \quad ; \\ f_C^{(j+1)} = P_C \left\{ f_C^{(j)} + \tau \left(A^T g - A^T A f_C^{(j)} \right) \right\}, \end{cases} \tag{3.17}$$

converge ad una soluzione del problema (3.16) se τ soddisfa alle condizioni (3.8).

Ovviamente, se la soluzione è unica, il limite non dipende dall'inizializzazione dell'algoritmo. Altrimenti si possono ottenere diverse soluzioni utilizzando diversi oggetti iniziali. Anche in vista di ciò, si preferisce inizializzare l'algoritmo con l'oggetto identicamente nullo. Non è tuttavia dimostrato (anche se è plausibile) che, in tal modo, si ottenga la soluzione di norma minima tra tutte quelle del problema (3.16).

Esperimenti numerici evidenziano che anche questo metodo iterativo ha la proprietà di semiconvergenza. Occorre pertanto arrestare le iterazioni prima della convergenza, anche se il minimo dell'errore di ricostruzione è, in genere, molto piatto, per cui la scelta del numero di iterazioni può non essere critica.

Si descrive l'algoritmo nel caso in cui C è il cono non-negativo. Si indicherà con P_+ l'operatore di proiezione sul cono e con $f_+^{(j)}$ l'iterata di ordine j.

- si calcola $\hat{K}^* \hat{g}$ e si pone $f_+^{(0)} = 0$;
- dato $f_+^{(j)}$ si calcola $\hat{f}_+^{(j)}$ mediante FFT;
- si calcola $\hat{h}^{(j)} = \hat{f}_+^{(j)} + \tau \left(\hat{K}^* \hat{g} - \widehat{\left| \hat{K} \right|^2} \hat{f}_+^{(j)} \right)$;
- si calcola $h^{(j)}$ mediante FFT inversa ;
- si azzerano i valori negativi di $h^{(j)}$, cioè si calcola $P_+ h^{(j)}$;
- si pone $f_+^{(j)} = P_+ h^{(j)}$.

3.3 – Metodo della massima pendenza

Sia il metodo di Landweber che il metodo di Landweber proiettato sono metodi molto lenti. Poichè occorre arrestare le iterazioni per evitare un'eccessiva propagazione del noise, questa lentezza può non essere un difetto, dato che la scelta del numero di iterazioni non è critica. Tuttavia è ovvio che i metodi non sono computazionalmente efficienti. Una loro ''accelerazione'' può essere ottenuta mediante un'opportuna scelta del parametro di rilassamento τ , cioè del passo nella direzione della massima pendenza. Come è abbastanza ovvio, si ottiene la migliore convergenza se si sceglie τ vicino, ma non troppo, al suo estremo superiore, dato dalla (3.8). Si osservi che, nel caso di PSF normalizzate a volume 1, si ha $0 < \tau < 2$; dall'esperienza numerica segue che un valore a 1.8 può essere ottimale dal punto di vista della convergenza.

Tuttavia il metodo di Landweber appartiene alla categoria dei *metodi stazionari* , cioè dei metodi per i quali i parametri, del tipo del parametro di rilassamento, non dipendono dall'iterazione. La convergenza può essere migliorata considerando *metodi non-stazionari* , cioè metodi nei quali la scelta del parametro di rilassamento dipende dall'iterazione.

L'idea di base consiste nel muoversi nella direzione di massima pendenza a partire dall'iterata di ordine j e nel scegliere il passo in modo da minimizzare il funzionale lungo questa retta.

Al fine di semplificare le notazioni ed i calcoli, sostituiamo il funzionale ai minimi quadrati con il seguente:

$$\eta(f;g) = \frac{1}{2}\left(A^T A f, f\right) - \left(A^T g, f\right) \quad , \qquad (3.18)$$

la cui minimizzazione è equivalente a quella del suddetto funzionale, dato che si ha:

$$\varepsilon^2(f;g) = 2\,\eta(f;g) + \|g\|^2 \quad .$$

Il gradiente del nuovo funzionale è dato da:

$$\nabla_f \eta(f;g) = A^T A f - A^T g \quad .$$

Inoltre, data l'iterata di ordine j, si definisce il suo residuo di ordine j come segue:

$$r^{(j)} = A^T g - A^T A f^{(j)} \quad . \qquad (3.19)$$

E' ovvio che il residuo non è altro che il gradiente, calcolato nell'iterata e cambiato di segno; pertanto l'iterata di ordine successivo sarà del tipo:

$$f^{(j+1)} = f^{(j)} + \tau\, r^{(j)} \quad . \qquad (3.20)$$

Al variare di t si ottengono punti che stanno sulla semiretta che passa per l'iterata di ordine j e punta nella direzione di massima pendenza. Si tratta di minimizzare il funzionale η(f;g) su questa semiretta.

Sostituendo la (3.20) nella (3.18), si ottiene:

$$\eta\left(f^{(j+1)};g\right) = \frac{1}{2}\left(A^T A\left(f^{(j)} + \tau\, r^{(j)}\right), f^{(j)} + \tau\, r^{(j)}\right) - \left(A^T g, f^{(j)} + \tau\, r^{(j)}\right) =$$

$$= \eta\left(f^{(j)};g\right) + \tau\left(A^T A f^{(j)}, r^{(j)}\right) + \frac{1}{2}\tau^2 \left\| A r^{(j)} \right\|^2 - \tau\left(A^T g, r^{(j)}\right) =$$

$$= \eta\left(f^{(j)};g\right) - \tau \left\| r^{(j)} \right\|^2 + \frac{1}{2}\tau^2 \left\| A r^{(j)} \right\|^2 \quad .$$

Pertanto, lungo la semiretta, il funzionale è una funzione quadratica di τ.

Derivando rispetto a τ e ponendo eguale a zero, si ottiene che il funzionale è minimo per il seguente valore di τ :

$$\tau^{(j)} = \frac{\left\| r^{(j)} \right\|^2}{\left\| A r^{(j)} \right\|^2} \qquad . \qquad (3.21)$$

Pertanto, il seguente algoritmo ottimizza, ad ogni passo, la scelta del parametro di rilassamento:

- porre $f^{(0)} = 0$;
- dato $f^{(j)}$, calcolare $r^{(j)} = A^T g - A^T A f^{(j)}$;
- calcolare $\tau^{(j)}$ mediante la (3.21) ;
- calcolare $f^{(j+1)}$ mediante :

  $$f^{(j+1)} = f^{(j)} + \tau^{(j)} r^{(j)} \qquad .$$

Questo metodo viene detto metodo della massima pendenza; è dunque un metodo di tipo gradiente con il passo ottimizzato. Riduce il numero di iterazioni rispetto a Landweber, con un modesto aumento del costo computazionale dovuto al calcolo del passo.

E' utile, in vista della prossima generalizzazione, considerare il metodo della massima pendenza come un metodo iterativo per il calcolo dei residui, che permette di aggiornare la soluzione ad ogni passo:

- porre $r^{(0)} = A^T g\,,\ f^{(0)} = 0$;
- dati $r^{(j)}$ e $f^{(j)}$, calcolare $\tau^{(\mathrm{j})}$ mediante la (3.21) e inoltre :

$$f^{(j+1)} = f^{(j)} + \tau^{(\mathrm{j})} r^{(j)} \ ;$$

$$r^{(j+1)} = r^{(j)} - \tau^{(\mathrm{j})} A^T A\, r^{(j)} \ .$$

Dalla relazione ricorrente tra i residui si ricava che:

$$\left(r^{(j+1)}, r^{(j)}\right) = \left\| r^{(j)} \right\|^2 - \tau^{(j)} \left\| A r^{(j)} \right\|^2 = 0.$$

Poichè il residuo è ortogonale alla superficie di livello che passa per una iterata, ciò significa che ci si muove lungo la direzione del residuo fino al punto in cui tale direzione è tangente ad un'altra superficie di livello del funzionale.

Anche per questo metodo vale la semiconvergenza.

3.4 – Metodo del gradiente conigato

I metodi precedentemente considerati, in generale, non sono convenienti dal punto di vistacomputazionale perchè la convergenza, o semiconvergenza, è molto lenta; in altri termini, occorre un gran numero di iterazioni per ottenere la soluzione ottimale. Tale comportamento è dovuto al fatto che il ''bacino'' che contiene il punto di minimo è molto ''piatto'' e dunque i metodi iterativi suddetti, che procedono nella direzione del gradiente, avanzano con ''passi'' molto piccoli. Il metodo del gradiente coniugato propone una diversa direzione di avanzamento.

Al fine di formulare il metodo, occorre la seguente definizione:

Definizione – Due tabelle f, h si dicono coniugate rispetto ad A , o anche A-ortogonali , se vale la seguente proprietà:

$$\left(Af\ ,\ Ah\ \right) = 0\ \ .$$

Infatti, nel metodo del gradiente coniugato ci si muove non in direzione ortogonale alle curve di livello ma in direzione A-ortogonale ad esse, cioè non in direzione del gradiente ma, appunto, in direzione del gradiente coniugato.

Un'altra definizione importante è quella di sottospazio di Krylov. Se si analizzano i metodi iterativi di Landweber e della massima pendenza, si riconosce che all'iterazione j, essi producono un'approssimazione della soluzione che, posto:

$$\tilde{g} = A^T g \quad , \quad \tilde{A} = A^T A \quad ,$$

è una combinazione lineare delle tabelle:

$$\tilde{g} \ , \ \tilde{A}\tilde{g} \ , \tilde{A}^2\tilde{g}_{\bullet}, \ldots\ldots , \tilde{A}^{j-1}\tilde{g} \quad .$$

Il sottospazio generato da queste tabelle viene detto appunto sottospazio di Krylov di ordine j e viene indicato con la notazione:

$$K^{(j)}\left(\tilde{g}\,;\tilde{A}\right) = span\left\{\tilde{g}\,,\ \tilde{A}\,\tilde{g}\ ,\tilde{A}^2\tilde{g}\,,\ldots\ldots,\tilde{A}^{j-1}\tilde{g}\right\}. \qquad (3.22)$$

Inoltre si indica con P_j l'operatore di proiezione ortogonale su tale sottospazio; più esattamente $P_j f$ è l'elemento del sottospazio di Krylov di ordine j che ha distanza minima da f.

Poichè i metodi iterati suddetti producono approssimazioni dell'oggetto incognito f che appartengono via via a sottospazi di Krylov di dimensioni crescenti, è ovvio che le migliori approssimazioni sarebbero le proiezioni di tale oggetto su tali sottospazi. Tuttavia tali proiezioni non possono essere determinate poichè richiedono la conoscenza dell'oggetto stesso. Come si vedrà il metodo del gradiente coniugato, pur non potendo fare questo, farà qualcosa di meglio dei metodi precedenti poichè, all'iterazione j, minimizza il funzionale ai minimi quadrati, o funzionale discrepanza, sul sottospazio di Krylov di ordine j. Detto altrimenti, esso risolve ricursivamente le equazioni di Eulero proiettate:

$$P_j A^T A P_j f = P_j A^T g \quad \rightarrow \quad P_j \tilde{A} P_j f = P_j \tilde{g} \,. \qquad (3.23)$$

Per tale motivo, anche se il gradiente coniugato viene implementato come un metodo iterativo, esso appartiene ugualmente alla classe dei metodi di proiezione, cioè dei metodi che forniscono la soluzione dell'equazione d'Eulero proiettata su sottospazi opportuni (in questo caso, i sottospazi di Krylov).

Il metodo del gradiente coniugato si basa sulla costruzione recursiva di due basi, una ortogonale e l'altra A-ortogonale, in sottospazi di Krylov di dimensioni via via crescenti. Si noti infatti che:

$$K^{(j)}\left(\tilde{g}\,;\,\tilde{A}\right) \supset K^{(l)}\left(\tilde{g}\,;\,\tilde{A}\right) \quad \text{per} \quad l = 1,2,\ldots,\ j-1\ .$$

Lo schema iterativo è il seguente:

- porre $r^{(0)} = p^{(0)} = \tilde{g}$;
- dati $r^{(j)}, p^{(j)}$ calcolare :

$$\alpha^{(\mathrm{j})} = \frac{\left\| r^{(j)} \right\|^2}{(r^{(j)}, \tilde{A} p^{(j)})} \quad , \quad r^{(j+1)} = r^{(j)} - \alpha^{(\mathrm{j})} \tilde{A} p^{(j)} \ ;$$

$$\beta^{(j)} = -\frac{(r^{(j+1)}, \tilde{A} p^{(j)})}{(p^{(j)}, \tilde{A} p^{(j)})} \quad , \quad p^{(j+1)} = r^{(j+1)} + \beta^{(j)} p^{(j)}$$

Lo schema iterativo per le soluzioni approssimate è poi il seguente:

– porre $f^{(0)} = 0$;

– dati $f^{(j)}$, $p^{(j)}$ calcolare :

$$f^{(j+1)} = f^{(j)} + \alpha^{(j)} p^{(j)} \quad . \qquad (3.24)$$

Le espressioni precedentemente date dei coefficienti α e β non sono quelle solitamente date nella formulazione dell'algoritmo (che saranno date in seguito); tuttavia esse sono utili per capire come l'algoritmo procede. E' infatti facile verificare che esse assicurano che ogni tabella tipo r è ortogonale alla precedente, cosi' come ogni tabella tipo p è A-ortogonale alla precedente. Vale a dire tali scelte dei coefficienti assicurano che:

$$(r^{(j+1)}, r^{(j)}) = 0 \quad , \quad (p^{(j)}, \tilde{A} p^{(j)}) = (Ap^{(j)}, Ap^{(j)}) = 0 \ .$$

Si noti che la prima condizione di ortogonalità coincide con quella soddisfatta dai residui nel caso del metodo della massima pendenza. si vedrà che la scelta della notazione non è casuale.

Valgono dei risultati più forti dei precedenti. Si può infatti dimostrare per induzione che:

- le tabelle $r^{(0)}, r^{(1)}, \ldots\ldots, r^{(j-1)}$ formano una base ortogonale in $K^{(j)}\left(\tilde{g}\,;\,\tilde{A}\right)$;
- le tabelle $p^{(0)}, p^{(1)}, \ldots\ldots, p^{(j-1)}$ formano una base $A-$ortogonale in $K^{(j)}\left(\tilde{g}\,;\,\tilde{A}\right)$;
- le tabelle $r^{(j)}$ sono precisamente i residui associati a $f^{(j)}$:

$$r^{(j)} = \tilde{g} - \tilde{A}\, f^{(j)} \quad ; \qquad (3.25)$$

- valgono inoltre le seguenti espressioni :

$$\alpha^{(j)} = \frac{\left\| r^{(j)} \right\|^2}{(p^{(j)}, \tilde{A} p^{(j)})} \quad , \quad \beta^{(j)} = \frac{\left\| r^{(j+1)} \right\|^2}{\left\| r^{(j)} \right\|^2} \quad .$$

E’ particolarmente facile dimostrare la (3.25), ne diamo quindi la dimostrazione. Innanzi tutto tale equazione è vera per j=0; si ammette quindi che sia vera per j e si dimostra che è ancora vera per j+1. Si ha infatti:

$$r^{(j+1)} = r^{(j)} - \alpha^{(j)} \tilde{A} p^{(j)} = \tilde{g} - \tilde{A} f^{(j)} - \alpha^{(j)} \tilde{A} p^{(j)} =$$
$$= \tilde{g} - \tilde{A}\left(f^{(j)} + \alpha^{(j)} p^{(j)}\right) = \tilde{g} - \tilde{A} f^{(j+1)} \quad .$$

Nell’ultimo passaggio si è usata la (3.24). Pertanto il gradiente coniugato permette di costruire una successione di residui che sono tutti tra loro ortogonali.

Ciò implica che il residuo di ordine j è ortogonale al sottospazio di Krylov di ordine j, cioè, in termini dell’operatore di proiezione ortogonale su tale sottospazio:

$$P_j r^{(j)} = 0 \quad \Rightarrow \quad P_j\left(\tilde{g} - \tilde{A} f^{(j)}\right) = 0 \quad , \; P_j f^{(j)} = f^{(j)} \, ,$$

il che implica che vale la (3.23).

In definitiva l'algoritmo, che può essere implementato direttamente in termini di DFT è il seguente:

- porre $r^{(0)} = p^{(0)} = \tilde{g}$, $f^{(0)} = 0;$
- dati $r^{(j)}, p^{(j)}, f^{(j)}$ calcolare :

$$\alpha^{(j)} = \frac{\left\| r^{(j)} \right\|^2}{(p^{(j)}, \tilde{A} p^{(j)})} \quad , \quad r^{(j+1)} = r^{(j)} - \alpha^{(j)} \tilde{A} p^{(j)} \quad ;$$

$$\beta^{(j)} = \frac{\left\| r^{(j+1)} \right\|^2}{\left\| r^{(j)} \right\|^2} \quad , \quad p^{(j+1)} = r^{(j+1)} + \beta^{(j)} p^{(j)} \quad ;$$

- calcolare :

$$f^{(j+1)} = f^{(j)} + \alpha^{(j)} p^{(j)} \quad .$$