

*Consiglio Nazionale delle Ricerche*

**ISTITUTO DI ELABORAZIONE  
DELLA INFORMAZIONE**

**PISA**

**UNA SOLUZIONE DUALE AL RESTAURO DI SEGNALI CON  
IL METODO DI MASSIMA ENTROPIA**

**ANNA TONAZZINI**

**Nota Interna B4 - 30  
Giugno 1989**

**UNA SOLUZIONE DUALE AL RESTAURO DI  
SEGNALI CON IL METODO DI MASSIMA ENTROPIA**

*Anna Tonazzini*

Istituto di Elaborazione della Informazione  
Consiglio Nazionale delle Ricerche

Via S.Maria, 46  
56100 PISA

## Sommario

Molti problemi di elaborazione di segnali e immagini possono essere formulati in termini di ricostruzione di una funzione continua sulla base di un insieme finito di sue misure distorte ed affette da rumore. In questo lavoro tale problema viene affrontato con riferimento al restauro di segnali monodimensionali.

Riconoscendo nel problema un tipico esempio di problema inverso mal-posto, si propone una tecnica di regolarizzazione basata sull'ottimizzazione di un funzionale costo soggetto a vincoli derivanti da conoscenze a priori. Nell'ipotesi che opportune condizioni di convessità siano soddisfatte, il problema di ottimizzazione vincolata originale, di dimensione infinita, può essere ricondotto ad un problema equivalente di ottimizzazione non vincolata, di dimensione finita, risolubile mediante un metodo di gradiente coniugato.

In questo lavoro vengono analizzati i casi in cui si utilizzano, come funzionali costo, l'entropia e la cross-entropia. Le prestazioni ottenibili con tali funzionali vengono confrontate con quelle fornite dal più tradizionale metodo di minima norma. I risultati dimostrano che il metodo di massima entropia consente il recupero di segnali continui di tipo impulsivo da un insieme molto limitato di misure.

## 1. Introduzione

Recentemente, il Metodo di Massima Entropia (MEM) è stato applicato alla risoluzione del problema del restauro e della ricostruzione di segnali e immagini [4,19,20,21].

In generale il problema del restauro si configura come il problema di stimare una funzione continua sulla base di un insieme discreto e finito di sue misure distorte ed affette da rumore. In questo senso esso costituisce un tipico esempio di problema inverso mal-posto.

L'utilizzo del MEM per la risoluzione del problema del restauro ha dato luogo a molte controversie circa le prestazioni fornite dal metodo stesso in relazione a quelle di altri metodi più tradizionali [1]. Molta parte di tali controversie derivano dal fatto che i fondamenti teorici su cui il MEM si basa sono di tipo probabilistico, e quindi idonei a giustificare principalmente l'utilizzo nella stima di distribuzioni di probabilità [2] o nella stima dello spettro di potenza di processi stocastici [3]. In tali ambiti, il MEM è stato di volta in volta considerato come l'approccio massimamente non impegnativo (non-committal) rispetto ai dati mancanti, il solo metodo consistente per combinare dati differenti, quello che fornisce la soluzione più uniforme fra quelle consistenti coi dati disponibili, o la soluzione più probabile.

Tuttavia non è ancora del tutto chiaro come le proprietà statistiche del MEM si adattino alla risoluzione del problema più generale, e usualmente affrontato da un punto di vista deterministico, della ricostruzione di segnali [11]. In questo senso l'approccio forse più corretto è quello di considerare il metodo di massima entropia come un particolare metodo variazionale per ottenere la regolarizzazione di problemi inversi mal-posti, cioè per rimuovere l'incertezza inerente ai problemi sottodeterminati dell'elaborazione di segnali.

Caratteristica comune di tutti i metodi basati sull'ottimizzazione vincolata è che la regolarizzazione è ottenuta riducendo la dimensione del problema di stima a quella dei dati, mediante l'uso di un modello a dimensione finita per la soluzione [4]. Ogni funzionale costo produce un modello differente, e la scelta del "miglior" funzionale costo per un particolare problema dipende dalla corrispondenza fra la classe di soluzioni ammesse dal modello e la classe delle funzioni a cui il segnale originale appartiene. In questo senso il MEM costituisce una valida tecnica di regolarizzazione se la funzione originale da stimare può essere espressa in termini del modello di massima entropia che da esso deriva.

Una conferma della validità di questa interpretazione modellistica dei principi variazionali nella regolarizzazione può essere trovata in alcuni lavori sul problema della stima spettrale monodimensionale. In essi è stato infatti messo in evidenza che la stima spettrale ottenuta massimizzando l'entropia, con vincoli sulla funzione di autocorrelazione, equivale all'uso del ben noto modello spettrale autoregressivo a tutti poli [6], mentre quella ottenuta minimizzando la norma  $L_2$  (o energia) soggetta agli stessi vincoli equivale all'uso dello stimatore di Blackmann-Tukey a tutti zeri [8]. Chiaramente, in presenza di pochi valori di correlazione, il MEM funzionerà meglio per spettri di tipo impulsivo, poichè i picchi possono essere modellati facilmente mediante poli, mentre il metodo di minima norma funzionerà meglio per spettri a larga banda, lentamente variabili, poichè spettri di questo tipo sono generalmente modellati meglio da zeri che da poli.

Argomenti simili si applicano anche al problema del restauro o ricostruzione di segnali, cosicchè l'interpretazione modellistica dei metodi basati sull'ottimizzazione vincolata non propone un funzionale costo come superiore a tutti gli altri, ma piuttosto stabilisce che il funzionale costo debba essere scelto in base alle conoscenze a priori sulla forma della soluzione [5,7].

In questo lavoro viene descritta un'applicazione dei metodi di ottimizzazione vincolata alla ricostruzione di segnali da dati sparsi e affetti da rumore, e in particolare vengono analizzate e confrontate le prestazioni del MEM e del metodo di minima norma, di più comune utilizzo. La relazione modellistica fra la soluzione e il funzionale costo è esplicitata mediante l'uso del teorema di dualità di Fenchel [9], che consente di trasformare il problema di ottimizzazione vincolata originale (problema primale) in un altro, ad esso equivalente, in generale di più facile risoluzione (problema duale). Un tale approccio è stato adottato per la prima volta in [6] per risolvere il problema della stima spettrale multidimensionale di tipo MEM e in [4] con applicazione al recupero di segnali monodimensionali. Mentre in [10] sono state fornite simulazioni del metodo per un problema di deconvoluzione in assenza di rumore, in questo lavoro i risultati sono relativi sia a dati esatti che a dati affetti da rumore.

Il lavoro è organizzato come segue. Nel paragrafo 2 è posizionato il problema della ricostruzione di segnali e la sua regolarizzazione mediante tecniche di ottimizzazione vincolata. Nel paragrafo 3, con riferimento all'Appendice, è applicato il Teorema di Fenchel per derivare i problemi duali relativi ai metodi di massima entropia e di minima norma. Nel paragrafo 4 è descritto un algoritmo di gradiente coniugato per la stima della soluzione del problema duale. Le simulazioni a calcolatore, effettuate per la valutazione delle prestazioni del metodo in presenza di dati sparsi ed affetti da rumore, sono illustrate nel paragrafo 5. In Appendice sono derivati in dettaglio i risultati presentati nel paragrafo 3.

## 2. Formulazione e risoluzione ottima del problema

La formulazione del problema della ricostruzione di segnali è fatta con riferimento al modello di sistema di misura lineare

continuo-discreto fornito dall'equazione integrale del primo tipo [16]:

$$g(s_k) = \int_D f(t) h(t, s_k) dt + n_k \quad k=1,2,\dots,N \quad (2.1)$$

dove  $f$  è una funzione continua e positiva su una regione compatta  $D$ ,  $h(t,s)$  rappresenta un kernel di convoluzione, i  $g(s_k)$ ,  $k=1,2,\dots,N$ , rappresentano i dati discreti visti come campioni della convoluzione continua fra  $f$  ed  $h$ , e gli  $n_k$ ,  $k=1,2,\dots,N$ , rappresentano campioni di un processo di rumore additivo.

Il problema della ricostruzione di un segnale si configura quindi come il problema inverso di ricavare  $f$  noti i dati  $g$ .

Benchè il modello di eq. (2.1) rispecchi al meglio la realtà fisica di ogni sistema di misura, l'approccio più comune alla risoluzione del problema della ricostruzione consiste nel campionare la funzione  $f(t)$  e nell'approssimare l'integrale mediante una somma. Alternativamente,  $f$  viene talvolta modellata come una combinazione lineare di un insieme di vettori di base continui. Usando il formalismo degli spazi vettoriali, tuttavia, il problema può essere direttamente espresso come la stima della funzione continua  $f$ , a partire da un insieme finito e discreto di misure o dati  $g$  [4,5,7].

In questo modo la soluzione è assunta appartenere allo spazio vettoriale  $X=C[D]$ , cioè lo spazio delle funzioni continue definite sull'insieme compatto  $D$ , con norma uniforme definita da:

$$\|f\| = \sup_{t \in D} |f(t)| \quad (2.2)$$

La relazione integrale (2.1) che lega i dati alla soluzione può essere riscritta come una trasformazione lineare  $G(f)$ , dallo spazio delle soluzioni  $X$  allo spazio  $R^N$  dei dati:

$$g=G(f)+n \quad (2.3)$$

dove  $n \in \mathbb{R}^N$  è il vettore rumore. Per denotare l'insieme dei kernel integrali si adotta la notazione vettoriale  $h(t) = \{h(t, s_1), h(t, s_2), \dots, h(t, s_N)\}$ . L'insieme dei dati disponibili verrà indicato con  $g^0$ .

La natura mal-posta del problema [17], così come formulato, è evidenziata dalle due seguenti osservazioni: (i) lo spazio delle soluzioni è a dimensione infinita, mentre i dati a disposizione sono in numero finito, cosicché si hanno certamente infinite funzioni che forniscono gli stessi dati; (ii) a causa della presenza del rumore, anche quando la soluzione ha la stessa dimensione dei dati, essa può essere poco attendibile, a meno che il rumore non sia esplicitamente incluso nella formulazione del problema.

Scopo della teoria della regolarizzazione di problemi mal-posti è quello di individuare metodi per superare entrambe queste difficoltà, cioè per selezionare una soluzione unica e robusta al rumore [18].

Una tecnica di regolarizzazione molto popolare è quella basata su principi di ottimizzazione vincolata, la cui idea fondamentale consiste nel restringere la classe delle soluzioni ammissibili, imponendo vincoli derivanti dalle conoscenze a priori che si hanno sul problema e sulla soluzione desiderata, e quindi di scegliere un funzionale costo da ottimizzare su tale classe.

In generale i vincoli sono rappresentati da proprietà note della soluzione (positività, estensione finita, limitatezza, ecc.) e da informazioni di natura statistica sul rumore, mentre il funzionale costo rappresenta una certa misura globale (entropia, energia, norma, ecc.) che deve essere ottimizzata dalla soluzione desiderata.

In questo lavoro sono stati considerati tre diversi funzionali costo, di particolare interesse nel settore dell'elaborazione dei segnali, e più precisamente:

$$\begin{aligned}
\text{(i) energia:} \quad & h_1(f) = \int_D \frac{f(t)^2}{2} dt \\
\text{(ii) entropia:} \quad & h_2(f) = - \int_D \log [f(t)] dt \\
\text{(iii) cross-entropia:} \quad & h_3(f) = \int_D f(t) \log [f(t)] dt
\end{aligned} \tag{2.4}$$

Una derivazione formale di  $h_2(f)$  è data in [2] per il caso in cui  $f(t)$  denota una distribuzione di probabilità, mentre una derivazione dell'entropia  $h_3(f)$  per il caso in cui  $f(t)$  denota uno spettro di potenza è data in [3].

Per molti problemi l'informazione disponibile a priori può essere separata in due gruppi di vincoli: quelli che governano particolari proprietà della soluzione e quelli basati sulle proprietà statistiche del rumore. Ogni vincolo può essere usato per definire un sottoinsieme dello spazio delle soluzioni contenente tutte quelle funzioni che obbediscono a quel vincolo. L'intersezione degli insiemi vincolo relativi alle informazioni sulla soluzione sia denotato con  $R$ , mentre quello relativo ai vincoli sul rumore sia chiamato  $S$ . Allora l'insieme  $R \cap S$  denoterà l'insieme delle soluzioni possibili.

Un tipico vincolo sulla soluzione, necessario per poter utilizzare i funzionali entropia  $h_2$  e  $h_3$ , è quello di positività:

$$R = \{f \in X : f(t) \geq 0 \text{ per ogni } t \in D\} \tag{2.5}$$

Per quanto riguarda il rumore, è sempre possibile assumere che qualche sua proprietà statistica sia nota o possa essere misurata. Il rumore di fondo, per esempio, può essere misurato in assenza di segnale e le sue statistiche possono essere ottenute empiricamente, oppure il processo di rumore può essere caratterizzato da un modello fisico noto. Una scelta tipica è quella che assume che il rumore sia additivo e gaussiano e che una stima a priori della sua matrice di

covarianza  $C$  sia disponibile. E' ben noto che per un processo gaussiano multivariato la forma quadratica  $n^T C^{-1} n$  ha una distribuzione di tipo chi-quadro con  $N$  gradi di libertà. Così si può definire l'insieme vincolo:

$$S = \{f \in X : (g^0 - g)^T C^{-1} (g^0 - g) \leq \delta \text{ con } g = G(f)\} \quad (2.6)$$

dove  $\delta$  denota la regione di confidenza.

In queste assunzioni, il problema del recupero di  $f(t)$  può essere così ricondotto al seguente problema di ottimo vincolato (problema primale):

$$\min h(f) \quad \text{tale che } f \in R \cap S \quad (2.7)$$

dove  $h(f)$  denota un opportuno funzionale costo, ad esempio uno dei tre definiti in eq. (2.4). Nel seguito verrà chiamato MEM I il metodo di massima entropia che utilizza il funzionale  $h_2(f)$  e MEM II quello riferito al funzionale  $h_3(f)$ .

Se la struttura di  $h(f)$  e dell'insieme vincolo  $R \cap S$  è tale da garantire l'esistenza e l'unicità del minimo, il problema della ricostruzione risulterà adesso ben-posto, cioè sarà risolto da una singola funzione consistente con le informazioni a priori sulla soluzione stessa e sui dati.

### 3. La formulazione duale di un problema di ottimo

La risoluzione del problema (2.7), che regolarizza il problema originariamente mal-posto della ricostruzione di  $f(t)$ , è in generale non banale, in quanto implica la risoluzione di un problema di ottimizzazione vincolata di dimensione infinita, per il quale è spesso impossibile individuare una soluzione esplicita. A tale scopo la tecnica forse più utilizzata fa uso della teoria dei moltiplicatori di Lagrange, la quale tuttavia fornisce condizioni solo necessarie all'esistenza di un punto stazionario per  $h(f)$ .

Se però il problema è convesso o concavo, cioè se valgono le ipotesi che l'insieme delle soluzioni possibili individuato dai

vincoli sia convesso, e che il funzionale costo sia convesso o concavo, è possibile far ricorso al potente teorema di dualità di Fenchel [9], che consente di ridurre il problema di ottimo vincolato originale, di dimensione infinita, a un problema equivalente di dimensione finita [4,5,7].

In generale, i teoremi di dualità rivestono un ruolo molto importante nella teoria dell'ottimizzazione. Il duale di uno spazio vettoriale, cioè l'insieme di tutti i funzionali lineari definiti su quello spazio, rappresenta infatti un ambiente alternativo in cui formulare un problema di ottimo. Associato ad ogni problema primale di ottimizzazione sullo spazio vettoriale originale vi è un problema duale, in cui un funzionale alternativo è ottimizzato sullo spazio duale.

In particolare il teorema di dualità di Fenchel stabilisce un'equivalenza tra il problema primale di minimizzare la differenza di un funzionale convesso  $h(f)$  e di uno concavo  $k(f)$ , definiti entrambi sull'intersezione di due insiemi convessi  $R$  ed  $S$ , e quello duale di massimizzare un funzionale concavo nello spazio duale. Il problema duale è equivalente al problema primale nel senso che il funzionale obiettivo del problema primale è minimizzato allo stesso valore per cui il funzionale obiettivo del problema duale è massimizzato. In formule si ha:

$$\inf_{f \in R \cap S} [h(f) - k(f)] = \max_{f^* \in R^* \cap S^*} [k^*(f^*) - h^*(f^*)] \quad (3.1)$$

dove l'asterisco indica coniugazione [9], cioè appartenenza ai relativi spazi duali.

Molto spesso, come nel caso trattato in questo lavoro, il problema duale è più semplice di quello primale. In particolare esso può essere di dimensione finita e non vincolato. Inoltre il calcolo del funzionale duale  $\phi = [k^*(f^*) - h^*(f^*)]$  nel termine destro della (3.1), quando può essere eseguito per via analitica, dà luogo ad una relazione che lega fra loro la soluzione del problema primale e quella del problema duale, in termini di un

modello dipendente dal funzionale costo. E' proprio quest'ultimo il risultato che consente di giustificare in modo rigoroso la visione modellistica dei metodi di ottimizzazione vincolata nella regolarizzazione proposta in questo lavoro.

Questa tecnica fu utilizzata per la prima volta da Lang e McClellan [6] per risolvere il problema della stima spettrale di massima entropia. Un metodo di ottimizzazione generalizzato basato su questo teorema è dato in [4] per il recupero di segnali con insiemi vincolo convessi arbitrari.

In questo lavoro il teorema di Fenchel è stato applicato al problema primale di eq. (2.7), in cui per  $k(f)$  è da intendersi il funzionale identicamente nullo (concavo), e  $h(f)$  rappresenta uno dei tre funzionali di eq.(2.4), evidentemente convessi in  $X=C[D]$ . Per quanto riguarda gli insiemi vincolo  $R$  ed  $S$ , sono stati considerati rispettivamente gli insiemi convessi (2.5) e (2.6) per i metodi MEM, mentre per il metodo di minima energia è stato considerato il solo insieme (2.6), essendo, in assenza di vincoli sulla soluzione,  $R=X$ . In pratica, in entrambi i casi, si è assunto di possedere informazioni a priori solo sul rumore. In queste specifiche ipotesi il teorema risulta in:

$$\min h(f) \quad \text{tale che} \quad f \in R \cap S \equiv \max \phi(\lambda) \quad \lambda \in R^N \quad (3.2)$$

Come già anticipato, in questo caso il problema duale risulta notevolmente semplificato rispetto a quello primale, in quanto a dimensione finita e non vincolato. Il vettore dei parametri del problema duale,  $\lambda \in R^N$ , è legato alla soluzione del problema primale da un modello a dimensione finita che deriva in modo naturale dall'applicazione del teorema. Il funzionale duale  $\phi$  è in questo caso la somma delle soluzioni di due problemi di ottimizzazione vincolata, uno risolvibile analiticamente, e l'altro risolvibile mediante il Teorema di dualità di Lagrange (si veda l'Appendice).

Per i tre funzionali costo definiti in eq. (2.4), i funzionali primale e duale sono legati dai seguenti modelli:

$$\begin{aligned}
\hat{f}_1(t) &= \lambda^T \mathbf{h}(t) \\
\hat{f}_2(t) &= \frac{1}{\lambda^T \mathbf{h}(t)} \\
\hat{f}_3(t) &= \frac{1}{e} \exp[\lambda^T \mathbf{h}(t)]
\end{aligned}
\tag{3.3}$$

I parametri di questi modelli vengono calcolati ottimizzando gli appropriati funzionali duali:

$$\begin{aligned}
\phi_1(\lambda) &= \lambda^T \mathbf{g}^0 - \int_D \frac{\hat{f}(t)^2}{2} dt - [\delta \lambda^T C \lambda]^{1/2} \\
\phi_2(\lambda) &= \lambda^T \mathbf{g}^0 + \int_D \log [\hat{f}(t)] dt - [\delta \lambda^T C \lambda]^{1/2} \\
\phi_3(\lambda) &= \lambda^T \mathbf{g}^0 + \int_D \hat{f}(t) dt - [\delta \lambda^T C \lambda]^{1/2}
\end{aligned}
\tag{3.4}$$

Per i dettagli sulla derivazione delle (3.3) e (3.4) si rimanda all'Appendice.

Dalle eqq. (3.3) è evidente che l'adozione di un particolare funzionale costo, da ottimizzare soggetto ai vincoli derivanti dai dati, induce un modello per la soluzione in termini di un vettore incognito  $\lambda$  e dei kernel di eq.(2.1) denotati dal vettore  $\mathbf{h}(t)$ . Sebbene questo modello si espliciti solo quando si risolve il problema duale, il fatto che la soluzione primale e duale siano equivalenti [9] significa che questi modelli sono implicitamente adottati anche quando si esegue l'ottimizzazione nello spazio primale. Così le prestazioni di ogni metodo di ottimizzazione saranno dipendenti dalla relazione fra le soluzioni permesse dal modello per  $f(t)$  e la funzione originale incognita. E' per questo motivo che sembra poco plausibile che un singolo metodo di ottimizzazione, ad esempio il MEM, possa essere superiore a tutti gli altri [1,11].

L'adozione del vincolo di consistenza esatta coi dati, in luogo del vincolo di eq.(2.6), produrrebbe modelli identici a quelli dati in eq. (3.3); il funzionale duale  $\phi(\lambda)$  differirebbe solo per l'assenza del termine finale in eq.(3.4). In assenza di rumore un tale vincolo è formalmente corretto e la massimizzazione di  $\phi(\lambda)$  produce una soluzione che è esattamente consistente con i dati disponibili. Tuttavia, in presenza di rumore, l'adozione di tale vincolo renderebbe il metodo molto instabile. Per evitare questo, si sceglie di utilizzare il vincolo rilassato definito dall'eq. (2.6). Così il valore  $\delta$ , che definisce una regione di confidenza, ad esempio al 95%, per la forma quadratica  $n^T C^{-1} n$ , può essere visto come un fattore di regolarizzazione predeterminato, scelto sulla base della stima a priori disponibile per la covarianza del rumore.

#### 4. L'algoritmo di ottimizzazione

Si è detto che per calcolare la soluzione al problema duale occorre massimizzare il funzionale  $\phi(\lambda)$  per  $\lambda \in R^N$ . Poichè, per quanto affermato nel teorema di Fenchel,  $\phi(\lambda)$  è concavo ed ammette un massimo finito, tale massimo sarà raggiunto nel vettore  $\lambda$  per cui il gradiente del funzionale stesso si annulla. Il parametro  $\lambda$  così calcolato, sostituito nel modello per  $f(t)$ , fornisce poi la soluzione del problema primale originale.

Ponendo a zero il gradiente di  $\phi(\lambda)$  si ottiene un insieme di equazioni in generale non lineari. Per esempio, considerando il funzionale entropia  $h_3(f)$  di eq.(2.4), il gradiente del relativo funzionale duale  $\phi_3(\lambda)$  è dato da:

$$\nabla \phi_3(\lambda) = g^0 - \hat{g} - \left[ \frac{\delta}{\lambda^T C \lambda} \right]^{\frac{1}{2}} C \lambda \quad (4.1)$$

dove

$$\hat{g}_k = \int_D \frac{1}{e} \exp [\lambda^T h(r)] h_k(r) dr \quad (4.2)$$

Ripetendo l'operazione anche per gli altri funzionali duali, si osserva che il gradiente in tutti i casi è dato dall'eq.(4.1), in cui  $g$  rappresenta i dati ottenuti sostituendo  $\lambda$  nell'appropriato modello per  $f(t)$ .

Si è già notato che utilizzando il vincolo di esatta consistenza con i dati, in luogo del vincolo espresso da eq. (2.6), l'ultimo termine in  $\phi(\lambda)$  non è presente e quindi, ponendo a zero il gradiente, in questo caso si ottiene l'equazione:

$$g^0 = \hat{g} \quad (4.3)$$

che evidenzia come il vettore duale possa essere trovato rimpiazzando  $f(t)$  con il modello  $f(t)$  nell'equazione vincolo e risolvendo in  $\lambda$ . Questo risultato può ovviamente essere derivato direttamente applicando la condizione di Kuhn-Tucker [9] al problema primale con il vincolo di esatta consistenza dei dati.

Quindi, dal confronto fra la (4.3) e la corrispondente condizione per il vincolo rilassato, che può essere riscritta come

$$g^0 = \hat{g} + \left[ \frac{\delta}{\lambda^T C \lambda} \right]^{\frac{1}{2}} C \lambda \quad (4.4)$$

si vede in che modo il problema duale esplicita l'effetto di rilassamento prodotto dal parametro  $\delta$ .

I risultati presentati nel paragrafo successivo sono stati ottenuti usando il metodo del gradiente coniugato con restart [12,13,14], piuttosto che risolvere direttamente l'eq.(4.4). Lo schema iterativo per tale algoritmo è dato dalle seguenti formule:

$$\lambda^{n+1} = \lambda^n + \gamma^n v(\lambda^n) \quad (4.5)$$

dove

$$v(\lambda^n) = \nabla \phi(\lambda^n) + \beta^n (\lambda^{n+1})$$

$$\beta^n = \begin{cases} 0 & n=0 \\ \frac{\|\nabla \phi(\lambda^n)\|}{\|\nabla \phi(\lambda^{n-1})\|} & n>0 \end{cases} \quad (4.6)$$

e  $\gamma^n$  è scelto in modo tale da massimizzare  $\phi(\lambda^{n+1})$ . Una espressione algebrica per  $\gamma^n$  può essere trovata sostituendo eq.(4.5) in  $\phi$  e uguagliando a zero la derivata rispetto a  $\gamma^n$ . Alternativamente se ne può trovare un valore approssimato usando l'algoritmo di ricerca di Fibonacci [14]. Nelle simulazioni eseguite è stato riscontrato che questo metodo è più efficiente del calcolo esatto per  $\gamma^n$ .

La procedura di restart è implementata occasionalmente reinizializzando la direzione di descent  $v(\lambda^n)$  alla direzione dello steepest descent, cioè

$$v(\lambda^n) = \nabla \phi(\lambda^n)$$

Questo processo di restart è usualmente ripetuto ogni N iterazioni ed è impiegato per ottenere convergenza lineare per l'algoritmo del gradiente coniugato. Direzioni alternative di restart sono date in [13].

Nell'applicazione in esame, l'algoritmo del gradiente coniugato richiede, per ogni iterazione k, il calcolo di un integrale che, per il caso dell'entropia  $h_3(f)$  è dato da eq. (4.2), mentre per il modello di minima norma si riduce a:

$$\hat{g} = \int_D [\lambda^T h(r)] h(r) dr = H\lambda \quad (4.7)$$

valido per ogni valore di  $k$  e dove

$$H = \int_D h(r) h(r)^T dr \quad (4.8)$$

Così si osserva che con il modello lineare il sistema  $H$  può essere precalcolato usando integrazione numerica e quindi memorizzato.

Per i modelli non lineari l'integrale è a sua volta non lineare in  $\lambda$  e deve essere ricalcolato ad ogni iterazione. Quindi il numero di termini usati per stimare l'integrale sarà limitato da considerazioni di costo. In questo lavoro è stata usata una semplice regola di integrazione numerica rettangolare, e si è riscontrato che il passo di integrazione influenza la qualità della soluzione. In particolare il passo di integrazione che dà i risultati migliori è quello che corrisponde alla risoluzione desiderata per la soluzione. Viceversa, integrare con un passo più grande dà risultati peggiori. Un approccio alternativo, particolarmente utile quando  $h(t,s)$  è una funzione a rapida variazione, potrebbe essere quello di usare tecniche più sofisticate, come ad esempio l'integrazione di Romberg.

## 5. Risultati delle simulazioni

In questo paragrafo vengono analizzate le prestazioni degli stimatori di massima entropia, minima cross-entropia e minima norma per la deconvoluzione di dati sparsi e rumorosi.

Per scopi di confronto sono stati usati dati simili a quelli usati in [10], cioè si è assunto che la funzione incognita fosse una coppia di impulsi e la funzione di convoluzione fosse la funzione gaussiana. In particolare i dati sono stati generati in accordo all'eq. (2.1) con

$$f(t) = \delta(t-t_1) + \delta(t-t_2) + 0.000001$$

dove  $\delta(t)$  denota la funzione Delta di Dirac e il termine di fondo è usato per garantire la positività. Come kernel di convoluzione è stata usata la funzione

$$h(t,s) = 0.1 \exp \left\{ -\frac{(s-t)^2}{2\sigma_h^2} \right\}$$

e come rumore additivo è stato generato un segnale casuale incorrelato e gaussiano con varianza scelta per fornire 20 dB e 10 dB di rapporto segnale-disturbo (SNR). I campioni prelevati come dati, in assenza di rumore e con rumore di 10 dB e 20 dB rispettivamente, sono mostrati in Fig. 1.

La ricerca dell'ottimo dei tre funzionali di eq. (3.6) col metodo del gradiente coniugato è stata implementata così come descritta nel precedente paragrafo e i risultati relativi sono mostrati nelle Figg. 2, 3 e 4 per i tre insiemi di dati di Fig.1.

Sulla base di questi risultati possono essere fatte alcune osservazioni. Innanzi tutto il MEM (I e II) dà risultati migliori del metodo di minima norma. Una prima ragione per questo può risiedere nel fatto che le soluzioni MEM sono automaticamente vincolate ad essere positive mentre questo vincolo dovrebbe essere specificatamente incluso nel modello di minima energia [4,8]. La seconda ragione che spiega le prestazioni superiori del MEM sta nella forma del modello duale. Dalle eqq.(3.5) si può osservare che i vari modelli sono definiti in termini di una somma pesata di gaussiane ritardate. Ora, poichè il segnale scelto per le simulazioni è discontinuo e quindi non limitato in banda, è chiaro che una buona soluzione non potrà essere ottenuta con il modello lineare quando  $N$  è piccolo, dato che un impulso non può essere espresso come una somma finita di gaussiane. Per l'algoritmo MEM, invece, la nonlinearità del modello consente una rappresentazione molto migliore degli impulsi, anche con pochi termini. Questo può

essere spiegato, per il modello di entropia derivato da  $h_2(f)$ , dal fatto che si ha

$$\lambda^T h(r) \rightarrow 0 \quad \text{quando} \quad \hat{f}(r) \rightarrow \infty$$

Un effetto simile è ottenuto anche con il modello di entropia derivato da  $h_3(f)$ .

Nell'esempio mostrato il MEM I presenta migliori prestazioni del MEM II, il quale a sua volta dà migliori prestazioni del metodo di minima norma, sia per il caso senza rumore, che per il caso con rumore a 20 dB. Per il caso di rumore a 10 dB, le prestazioni dei due MEM sono rovesciate. L'algoritmo di MEM II dà due picchi di uguale peso, mentre il MEM I presenta picchi di altezza molto diversa; il metodo di minima norma rileva appena la presenza di due picchi.

Da questi risultati scaturiscono due osservazioni fondamentali. Come prima cosa, nella scelta del funzionale costo è importante tener conto delle proprietà che ci si aspetta la soluzione possieda. Da questo punto di vista, nell'esempio qui considerato il modello MEM I è chiaramente il migliore dei tre. Secondariamente, poichè la presenza del rumore altererà sempre le prestazioni degli stimatori, sebbene un funzionale costo possa modellare la soluzione meglio di un altro, la robustezza dei diversi modelli al rumore nei dati può differire, e quindi un funzionale alternativo può talvolta rappresentare una scelta migliore. Per l'esempio mostrato in questo lavoro appare che, per rapporti segnale-disturbo alti, il MEM I funziona meglio, ma per bassi rapporti segnale-disturbo il metodo MEM II è superiore.

## 6. Conclusioni

In questo lavoro è stata data una formulazione continuo-discreta alla soluzione del problema del restauro di segnali, basata su un metodo di ottimizzazione vincolata che usa il

teorema di dualità di Fenchel. Tale formulazione consente di evidenziare che l'effetto di regolarizzazione prodotto dalle tecniche di ottimizzazione vincolata può essere spiegato in termini di un modello a dimensione finita implicitamente adottato per la soluzione. Attraverso la simulazione è stato dato un esempio in cui entrambe le forme del metodo di massima entropia danno prestazioni che sono significativamente migliori di quelle fornite dal metodo di minima norma. La dipendenza dell'algoritmo dai dati è stata messa in evidenza: affinché il metodo funzioni bene la vera soluzione deve essere approssimata dalle funzioni ammesse dal modello associato con il funzionale costo scelto. Infine, la formulazione continuo-discreta consente la stima del segnale ricostruito in un numero di punti potenzialmente infinito. Quindi essa è più generale della formulazione discreto-discreta in cui eq.(2.1) è approssimata da un prodotto matrice-vettore. Tuttavia, una soluzione esatta al problema duale richiede il calcolo di un integrale, e poichè questo deve essere calcolato numericamente, eccetto che per il modello lineare, ne risulta che un qualche grado di approssimazione è ancora necessario anche nella formulazione continuo-discreta.

## Appendice

Con riferimento al problema di ricostruzione formulato secondo le eq. (2.1) e (2.3), si consideri il caso in cui si voglia minimizzare il funzionale convesso  $h(f)$  sull'insieme  $R \cap S$ , dove  $R$  è l'insieme convesso contenente tutte le funzioni consistenti con le conoscenze a priori sulla forma della soluzione, e  $S$  è l'insieme convesso contenente tutte le funzioni tali che il residuo  $g^0 - G(f)$  soddisfa i vincoli basati sulle statistiche del rumore. Se il funzionale costo è uno dei tre dati in eq. (3.4), esso può essere riscritto come:

$$h(f) = \int_D H[f(r)] dr \quad (A.1)$$

Applicando il teorema di dualità di Fenchel e considerando il funzionale concavo  $k(f)=0$  si ha:

$$\inf_{f \in R \cap S} [h(f) - k(f)] = \max_{f^* \in R^* \cap S^*} [k^*(f^*) - h^*(f^*)] \quad (A.2)$$

dove i funzionali coniugati  $k^*$  e  $h^*$  e gli insiemi coniugati  $S^*$  e  $R^*$  sono dati rispettivamente da [9]:

$$S^* = \{ f^* \in X^* : \inf_{f \in S} [\langle f, f^* \rangle - k(f)] > -\infty \}$$

$$k^*(f^*) = \inf_{f \in S} [\langle f, f^* \rangle - k(f)] \quad (A.3)$$

e da:

$$R^* = \{ f^* \in X^* : \sup_{f \in R} [\langle f, f^* \rangle - h(f)] < \infty \}$$

$$h^*(f^*) = \sup_{f \in R} [\langle f, f^* \rangle - h(f)] \quad (A.4)$$

dove  $X^*$  indica lo spazio duale di  $X=C[D]$  e  $\langle f, f^* \rangle$  indica il risultato dell'applicazione del funzionale lineare  $f^* \in X^*$  all'elemento  $f \in X$ . In particolare  $X^*$  è lo spazio delle funzioni a variazione limitata normalizzate definite su  $D$ , cioè  $X^* = NBV[D]$  [9].

Dalle (A.3) e (A.4) risulta che il funzionale costo duale del teorema di Fenchel è dato dalla differenza di due funzionali espressi essi stessi come soluzione di due sottoproblemi di ottimo vincolato. Tuttavia, si può osservare che il primo sottoproblema dipende solo dai vincoli sul rumore (insieme  $S$ ), mentre il secondo dipende solo dai vincoli sulla soluzione (insieme  $R$ ) e dalla forma del funzionale primale  $h$ . Molto spesso questa partizione dei vincoli, unita al fatto che i due sottoproblemi possono talvolta essere risolti per via analitica, semplifica grandemente il problema originale. Si vedrà in dettaglio come questo si verifica per il problema analizzato in questo lavoro.

Nell'ipotesi di  $k(f)=0$  si ha che

$$S^* = \{ f^* \in X^* : \inf_{f \in S} [\langle f, f^* \rangle] > -\infty \} \quad (A.5)$$

cioè  $S^*$  è costituito da quei funzionali di  $X^*$  che hanno un minimo finito su  $S$ .

Allora si consideri un  $f \in S$  e un  $d \in X$ , tale che  $G(d)=0$ . Poichè  $S$  è definito interamente in termini dell'operatore lineare  $G(f)$ , vale che  $f^0 = f + \alpha d \in S$  per ogni  $\alpha$ . La quantità

$$\begin{aligned} \langle f^0, f^* \rangle &= \langle f + \alpha d, f^* \rangle = \\ &= \langle f, f^* \rangle + \alpha \langle d, f^* \rangle \end{aligned}$$

può essere fatta piccola a piacere scegliendo  $\alpha$  arbitrariamente piccolo (o grande), a meno che non sia  $\langle d, f^* \rangle = 0$ . Questo implica che deve necessariamente essere:

$$\langle f, f^* \rangle = \int_D f(r) \lambda^T h(r) dr \quad (\text{A.6})$$

per ogni  $f^* \in S^*$ , dove  $\lambda$  è un vettore di  $R^N$ . Questa asserzione è derivata rigorosamente in [15]. Si ha quindi che:

$$\begin{aligned} k^*(f^*) &= \inf_{f \in S} [\langle f, f^* \rangle] = \\ &= \inf_{f \in S} \int_D f(r) \lambda^T h(r) dr = \inf_{g \in \Gamma} \lambda^T g \end{aligned} \quad (\text{A.7})$$

con

$$\Gamma = G(S) = \{ g \in R^N : G(f) = g \text{ per qualche } f \in S \} \quad (\text{A.8})$$

Il sottoproblema (A.7) è usualmente più semplice rispetto al problema primale originale; in questo caso in particolare è a dimensione finita e dipende solo dai vincoli sul rumore espressi mediante l'insieme  $S$ . La sua risoluzione può essere eseguita mediante il metodo primale-duale di Lagrange [9], nel seguente modo.

Si consideri l'insieme  $S$  dato dalla (2.6). Secondo il metodo primale-duale di Lagrange si ha:

$$\inf_{g \in \Gamma} \lambda^T g = \max_{\eta > 0} \inf_{g \in R^N} \{ \lambda^T g + \eta [ (g - g^0)^T C^{-1} (g - g^0) - \delta ] \} \quad (\text{A.9})$$

Per il calcolo dell'inf alla destra della (A.9) si calcola la derivata del termine fra parentesi graffa rispetto a  $\mathbf{g}$  e si pone a zero. Si ottiene:

$$\lambda^T + 2\eta (\mathbf{g} - \mathbf{g}^0) C^{-1} = \mathbf{0}^T \quad (\text{A.10})$$

da cui, per la simmetria di  $C$

$$\mathbf{g} = \mathbf{g}^0 - \frac{C\lambda}{2\eta} \quad (\text{A.11})$$

Il valore di  $\mathbf{g}$  dato dalla (A.11) risostituito nella (A.9) fornisce il termine in  $\eta$  che deve essere reso massimo cioè

$$\max_{\eta > 0} \lambda^T \mathbf{g}^0 - \frac{\lambda^T C \lambda}{4 \eta} - \eta \delta \quad (\text{A.12})$$

Ancora derivando, questa volta in  $\eta$ , e ponendo a zero si ottiene

$$\frac{\lambda^T C \lambda}{4 \eta^2} - \delta = 0 \quad (\text{A.13})$$

da cui

$$\eta = \frac{1}{2} \sqrt{\frac{\lambda^T C \lambda}{\delta}} \quad (\text{A.14})$$

Il valore calcolato per  $k^*(f^*)$  sarà allora dato da

$$k^*(f^*) = \lambda^T \mathbf{g}^0 - \sqrt{\delta \lambda^T C \lambda} \quad (\text{A.15})$$

in cui è evidenziato il legame con i soli vincoli sul rumore e la dipendenza dai parametri  $\lambda$ .

Si consideri adesso l'insieme coniugato  $R^*$  e il funzionale coniugato  $h^*(f^*)$ . Dalla definizione (A.4) e da quanto detto sulla forma di  $\langle f, f^* \rangle$  (A.6) si ha:

$$\begin{aligned} h^*(f^*) &= \sup_{f \in R} [ \langle f, f^* \rangle - h(f) ] = \\ &= \sup_{f \in R} \left\{ \int_D f(r) \lambda^T h(r) dr - h(f) \right\} \end{aligned} \quad (A.16)$$

Anche il calcolo di  $h^*(f^*)$  richiede la risoluzione di un sottoproblema di ottimizzazione vincolata, in questo caso eseguibile per via analitica. Se  $R$  è un aperto di  $X$ , la soluzione di questo sottoproblema è infatti data da quella funzione  $f$  che annulla il differenziale di Frechet del funzionale integrale in (A.16). Tale funzione risulterà espressa in termini di  $\lambda$ , dei kernel  $h(r)$  e dipenderà dalla forma del funzionale costo  $h$ . Si avrà:

$$\hat{f}(r) = \psi(\lambda, h(r)) \quad (A.17)$$

soggetto ai vincoli espressi da  $R$ .

Equazione (A.17) esprime la relazione fra la soluzione e il funzionale costo in termini di un modello, dipendente dai parametri  $\lambda$ , e quindi di dimensione finita pari a quella dei dati.

Con riferimento al problema analizzato in questo lavoro, tenendo presente che i funzionali costo  $h(f)$  considerati sono del tipo di eq. (A.1), il sottoproblema (A.16) può essere riscritto come:

$$h^*(f^*) = \sup_{f \in R} \int_D \{ f(r) \lambda^T h(r) - H[f(r)] \} dr \quad (A.18)$$

Calcolando il differenziale di Frechet del funzionale integrale in (A.18) si ottiene la condizione necessaria per l'esistenza di un punto di massimo data da [9]:

$$\int_D \left\{ \lambda^T h(r) - \frac{dH[f(r)]}{df(r)} \right\} \tau(r) dr = 0 \quad \forall \tau(r) \quad (\text{A.19})$$

e quindi

$$\frac{dH[f(r)]}{df(r)} = \lambda^T h(r) \quad (\text{A.20})$$

Sostituendo al posto dell'operatore H le forme da esso assunte nei tre funzionali costo di eq.(2.4) si ottengono i tre modelli per la soluzione forniti in eq. (3.3).

Sostituendo nella (A.18) l'espressione ricavata per la soluzione nella (A.20) si ottiene un'espressione analitica per  $h^*(f^*)$ , che risulta in una funzione dei parametri  $\lambda$ .

Combinando assieme le espressioni calcolate per  $k^*(f^*)$  e  $h^*(f^*)$  e sostituendole nel termine destro della (A.2) questo fornisce la forma del funzionale duale, che risulta una funzione concava [9] dei parametri  $\lambda$  e che dovrà essere resa massima.

Il problema duale è così completamente definito come un problema di ottimo non vincolato e di dimensione finita N. Nei tre casi considerati in questo lavoro, il funzionale duale risulta in particolare dato dalle (3.4). La sua soluzione, calcolata con un qualche algoritmo di descent, sostituita nel modello per f (eqq. (A.17) e (A.20)), darà la soluzione desiderata per il problema primale.

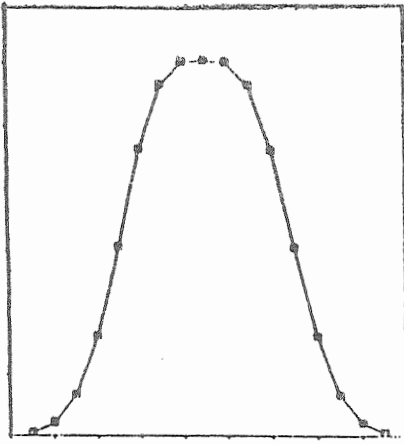


Fig.1a

Dati simulati, SNR =  $\infty$

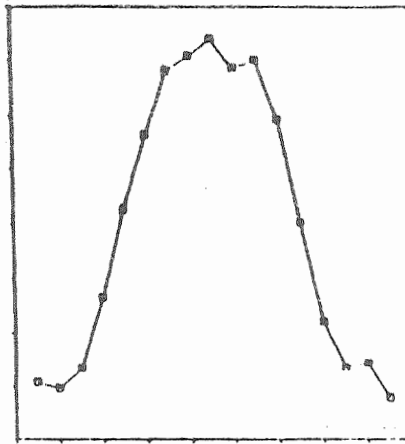


Fig.1b

Dati simulati, SNR=20dB

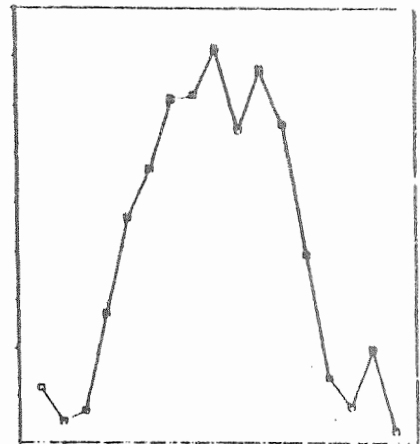


Fig.1c

Dati simulati, SNR=10dB

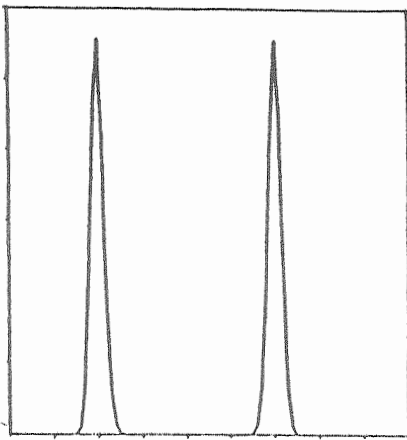


Fig.2a

Restauro MEM I, SNR= $\infty$

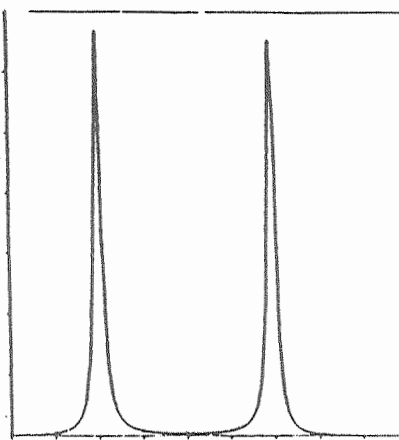


Fig.2b

Restauro MEM I, SNR=20dB

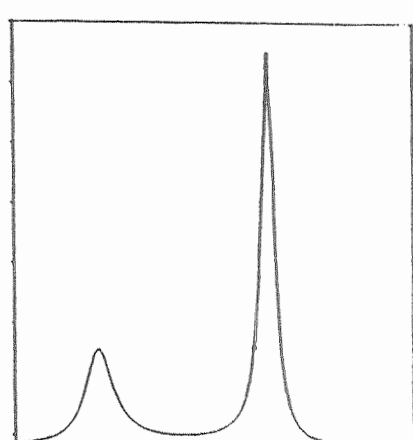


Fig.2c

Restauro MEM I, SNR=10dB

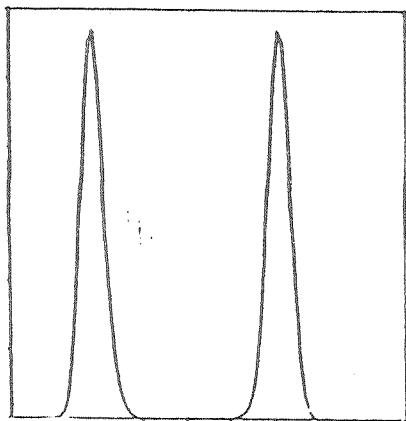


Fig.3a

Restaurom MEM II, SNR= $\infty$

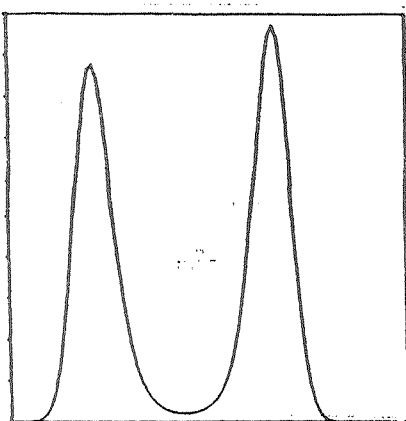


Fig.3b

Restaurom MEM II, SNR=20dB

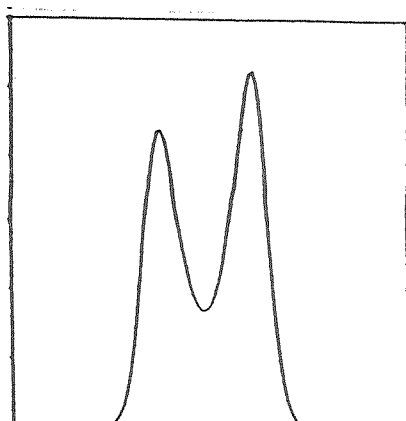


Fig.3c

Restaurom MEM II, SNR=10dB

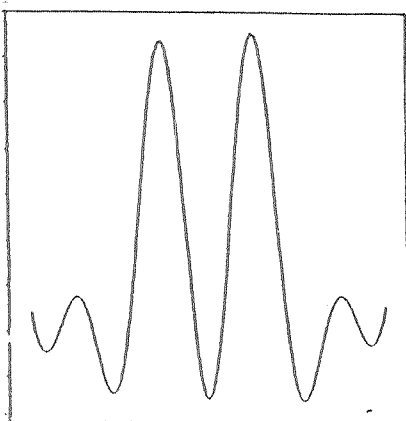


Fig.4a

Restaurom ML2, SNR= $\infty$

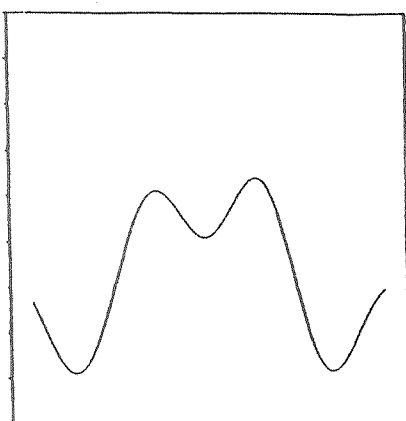


Fig.4b

Restaurom ML2, SNR=20dB

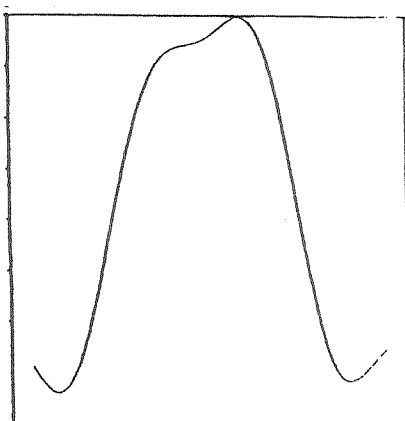


Fig.4c

Restaurom ML2, SNR=10dB

## Bibliografia

1. D.Titterington "The maximum entropy method for data analysis", più risposta di Skilling J., *Nature*, Vol.312, 1984, pp.381-382.
2. Shore J., Johnson R. "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy", *IEEE Trans. on Inform. Theory*, Vol.26, 1980, pp.16-27.
3. Papoulis A. Probabilities, *Random variables and Stochastic Processes*, McGraw-Hill, 1984.
4. Leahy R., Goutis C.E. "An optimal technique for constraint based image restoration and reconstruction", *IEEE Trans on Acoustic, Speech and Signal Processing*, Vol.34, 1986, pp.1629-1642.
5. Leahy R.M., Tonazzini A. "Maximum entropy signal restoration from short data records", *Proc. 8-th IASTED Int. Symp. MECO '86*, Taormina, Italy, 1986.
6. Lang S., McClellan J. "Multidimensional MEM spectral estimation", *IEEE Trans on Acoustic, Speech and Signal Processing*, Vol.30, 1982, pp.880-887.
7. Leahy R.M., Tonazzini A., Wang H. "A dual space optimization technique for maximum entropy signal reconstruction and restoration", *Proc. 20-th Asilomar Conf. on Signals, Systems and Computers*, Monterey, CA, 1986.
8. Steinhardt A., Goodrich R., Roberts R. "Spectral estimation via minimum energy correlation extension", *IEEE Trans. on Acoustic, Speech and Signal Processing*, Vol.33, 1985, pp.1509-1516.

9. Luenberger D.G. *Optimization by Vector Space Methods*, Wiley, 1969, New York.
10. Au-Yeung C., Merserau R. "Maximum entropy deconvolution", *Proc. of Int. Conf. on Acoustic, Speech and Signal Processing*, ICASSP '86, Japan, 1986.
11. Gull S., Skilling J. "Maximum entropy method in image processing", *IEE Proc. Pt.F*, Vol.131, N.6, 1984, pp.646-659.
12. Fletcher R., Reeves C.M. "Function minimization by conjugate gradients", *Computer Journal*, Vol.7, N.2, 1964, pp.149-154.
13. Powell M. "Restart procedures for the conjugate gradient method", *Math. Prog.*, Vol.12, 1977, pp.241-254.
14. Scales L.E. *Introduction to Non-Linear Optimization*, Springer-Verlag, 1985.
15. Mc Clellan J.H., Lang S.W. "Duality for multidimensional MEM spectral analysis", *IEE Proc. Pt.F*, Vol.130, N.3, pp.230-235, 1983.
16. Andrews H.C., Hunt B.R. *Digital Image Restoration*, Prentice-Hall, 1977, Englewood Cliffs.
17. Bertero M. "Linear inverse and ill-posed problems", *Internal Report INFN/TC-88/2*, 1988.
18. Tikhonov A.N., Arsenin V.Y. *Solutions of Ill-posed Problems*, Winston-Wiley, 1977, Washington.
19. Burch S.F., Gull S.F., Skilling J. "Image restoration by a powerful maximum entropy method", *Computer Vision, Graphics, and Image Processing*, Vol.23, 1983, pp.113-128.

20. Wernecke S.J., D'Addario L.R., "Maximum entropy image reconstruction", *IEEE Trans. on Computers*, Vol.26, 1977, pp.351-364.
21. Bedini L., Tonazzini A. "The use of neural networks in maximum entropy image restoration", 1989, in stampa.