

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea Magistrale in Matematica

**Tecniche di regolarizzazione
per analisi perfusionali
da immagini tomografiche**

Tesi di Laurea in Analisi Numerica e laboratorio

Relatore:
Chiar.mo Prof.
ELENA LOLI PICCOLOMINI

Presentata da:
ELENA MOROTTI

‡ Sessione Terza
Anno Accademico 2011/2012

*A mamma e papà,
e a mio marito*

Introduzione

Questa tesi nasce da un progetto di ricerca avviato nel 2009 dall'ARCES (Advanced Research Center on Electronic Systems, for information and communication technologies E. De Castro) di Bologna, in collaborazione con l'IRST, Istituto scientifico Romagnolo per lo Studio e la cura dei Tumori di Meldola (FC), e coordinato dal professor Alessandro Bevilacqua. Negli ultimi anni, la ricerca in campo oncologico ha portato allo sviluppo di nuove terapie antitumorali, alcune delle quali utilizzano tecniche di imaging che rendono possibile valutarne l'efficacia, prima ancora che siano osservabili le variazioni morfologiche delle masse tumorali in cura. A tal proposito, la tomografia computerizzata perfusionale (TCp) rappresenta una tecnica importante di imaging radiologico, perché é in grado di produrre indicatori funzionali, relativi alla vascolarizzazione dei tessuti analizzati, che caratterizzano accuratamente la zona di interesse clinico.

In letteratura, sono stati proposti diversi modelli matematici per il calcolo dei parametri perfusionali, relativi a una massa tumorale studiata mediante immagini di TCp. In questa tesi viene approfondita la risoluzione del *modello cinetico deconvoluzionale*, che viene formulato come un problema inverso, discreto e mal-posto. Per la risoluzione di tale problema, é necessario utilizzare tecniche di regolarizzazione capaci di limitare la propagazione degli errori, di cui sono affetti i dati sperimentalmente acquisiti.

L'obiettivo di questa tesi consiste pertanto nell'applicare e testare i metodi di regolarizzazione, nel valutarne i risultati e, non meno importante, nel-

l'automatizzare l'intero processo di risoluzione.

Precisamente, in precedenza il progetto includeva già una parte di regolarizzazione ed il contributo della tesi consiste nell'applicare queste tecniche in maniera più rigorosa e nel determinare autonomamente i parametri ottimali di cui necessitano, senza richiedere l'intervento di un utente esterno. Inoltre, sono state modificate o implementate ex novo delle function Matlab, sia per l'analisi e la risoluzione dei problemi inversi, che per la valutazione degli output prodotti.

Nello specifico, nel primo capitolo vengono illustrate le caratteristiche dei problemi mal-posti, sia nella formulazione continua che in quella discreta, poi sono presentate le tecniche di regolarizzazione di decomposizione troncata ai valori singolari e del metodo di Tikhonov, assieme alle strategie capaci di determinare i migliori parametri di regolarizzazione.

Nel secondo capitolo, invece, troviamo una panoramica sulle neoplasie e le analisi di tomografia perfusionale, seguita dalla definizione del modello cinetico deconvoluzionale.

Sulla base di quest'ultimo, nel terzo capitolo viene simulato un problema test e risolto con i metodi di indagine matematica, che ora possono essere valutati sulla base dei risultati forniti.

Infine, nel quarto capitolo, i metodi di regolarizzazione e valutazione matematica sono applicati su un problema reale di neoplasia epatica, individuata e studiata mediante una sequenza di 36 immagini TCp.

Indice

Introduzione	i
1 La regolarizzazione per la risoluzione di problemi lineari inversi	1
1.1 Problemi continui mal posti	2
1.1.1 L'espansione a valori singolari SVE	3
1.1.2 La condizione continua di Picard	5
1.2 Problemi discreti, inversi e mal posti	7
1.2.1 La SVD	7
1.2.2 La condizione discreta di Picard	9
1.2.3 Tipologie e analisi del rumore	10
1.3 Metodi diretti di regolarizzazione	11
1.3.1 Metodo di Tikhonov	14
1.3.2 Metodo della TSVD	15
1.4 Scelta del parametro di regolarizzazione	17
1.4.1 Principio di discrepanza	20
1.4.2 L-curva	21
1.4.3 Generalised Cross Validation	25
2 Le TC perfusionali	27
2.1 Le neoplasie e la tomografia computerizzata	27
2.1.1 Diagnosi e terapie	29
2.1.2 La tomografia computerizzata	30
2.1.3 La TC perfusionale	32

2.1.4	Considerazioni generali	35
2.2	Il modello della TCp	35
2.2.1	La curva di concentrazione nel tempo, TCC	35
2.2.2	Principii della tecnica	37
2.2.3	Il post-processing	40
2.2.4	Problemi legati alla mancanza di un protocollo d'ac- quisizione standard	42
2.2.5	Lo studio delle neoplasie tramite TCp	43
2.3	Il modello cinetico deconvoluzionale	44
2.3.1	Formulazione matematica	44
2.3.2	Formulazione matematica discreta	47
3	Applicazione al modello simulato	49
3.1	Creazione del problema test	50
3.2	Analisi degli output BF	55
4	Applicazione al caso reale	67
4.1	Definizione del problema	68
4.1.1	Preprocessing	72
4.2	Risultati	76
4.2.1	Risultati con la Tsvd	76
4.2.2	Risultati con Tikhonov	82
4.2.3	Osservazioni	84
	Conclusioni	91
	Bibliografia	93

Elenco delle figure

1.1	Esempi per l'analisi di Picard nel caso discreto	9
1.2	L-curva	22
2.1	Esempio di TCC	36
2.2	Slice di tessuto nelle diverse fasi dell'esame	37
2.3	Time Attenuation Curve	37
2.4	Andamento del tracciante	38
2.5	ROI necessarie per definire il modello	41
3.1	Funzione di input arterioso	51
3.2	Funzione residua teorica	51
3.3	Funzioni di segnale tissutale, pulito e con rumore aggiunto . .	52
3.4	Grafici per analizzare preventivamente il sistema da risolvere .	54
3.5	Grafico dei valori singolari della matrice, con evidenziati quelli relativi al 20% e 4% del valore singolare piú grande	56
3.6	Errori relativi per SNR=10 e seed variabile	57
3.7	L-curva ottenuta per il metodo di Tikhonov con operatore di derivata prima	58
3.8	Ricostruzioni ottenute con la TSVD, della funzione residuo . .	58
3.9	Esempi di L-curves con diversi vertici trovati e relative soluzioni	59
3.10	Varianti della L-curva, per determinare un buon parametro di regolarizzazione	61
3.11	Errori relativi commessi dalla TSVD con parametro scelto con diverse curve e funzioni	62

3.12	L-curve della TSVD al crescere del rumore	62
4.1	Alcune scansioni di TCp epatiche, relative allo stesso esame	68
4.2	ROI utilizzate per ricavare i dati	69
4.3	Dati del problema	70
4.4	Grafici per lo studio preliminare del problema	72
4.5	Grafici per lo studio preliminare del problema	73
4.6	Esempi di alcune curve, di output tissutale, mediate con un filtro di media mobile 3×3 e relative mappe colorimetriche dei residui, descritti tramite gli operatori di media e deviazione standard	75
4.7	Output con la TSVD classica	77
4.8	Esempi di L-curve della TSVD	78
4.9	Curve alternative alla L-curve	79
4.10	Grafici di Picard	80
4.11	Output della TSVD con parametro k di regolarizzazione fissato	81
4.12	Output con il metodo di Tikhonov	82
4.13	L-curve relative al metodo di Tikhonov	83
4.14	Istogrammi delle mappe riportate nelle figure 4.7 e 4.12	88
4.15	Mappe e relativi istogrammi, ottenuti cancellando l'ultimo 5 percentile	89

Elenco delle tabelle

3.1	Errori relativi compiuti sul calcolo di BF al variare del SNR, con la TSVD	65
3.2	Errori relativi compiuti sul calcolo di BF al variare del SNR, con Tikhonov	65
3.3	Tempi di calcolo per la simulazione	65
4.1	Coefficienti di correlazione delle mappe migliori	85

Capitolo 1

La regolarizzazione per la risoluzione di problemi lineari inversi

La maggior parte dei sistemi lineari, del tipo $Ax = b$, sono caratterizzati da un valore molto alto del numero di condizione della matrice A dei coefficienti. Ciò comporta che i metodi standard dell'algebra lineare non sono più applicabili per la risoluzione del problema inverso $x = A^{-1}b$; da qui la necessità di introdurre tecniche numeriche di risoluzione dei sistemi inversi e tecniche di regolarizzazione di questi ultimi.

Trattando sistemi con matrici mal condizionate, è importante capire la tipologia del mal condizionamento stesso, per poter scegliere il metodo migliore per il caso in analisi. Sono molte infatti le variabili che entrano in gioco: dalla scelta del metodo di risoluzione (metodi diretti o metodi iterativi), alla scelta della tecnica di regolarizzazione (si possono aggiungere informazioni per stabilizzare la soluzione e, in tal caso, è opportuno stabilire quanto tali nozioni debbano incidere sul processo).

Infine non dimentichiamo che, nella maggior parte dei casi, i sistemi lineari a dimensione finita sono la discretizzazione di un problema reale continuo: alcune tecniche consentono di limitare la propagazione degli errori di

discretizzazione e sono pertanto molto utilizzate in letteratura.

1.1 Problemi continui mal posti

Il concetto di problema ben-posto o mal-posto nasce con Hadamard nel 1923. In [6], Hadamard definí un problema *ben-posto* se é tale per cui valgono le seguenti condizioni:

1. esistenza della soluzione;
2. unicitá della soluzione;
3. stabilitá della soluzione.

Le prime due condizioni costituiscono il problema analitico, il terzo il problema numerico della risoluzione di problemi inversi. La mancata veridicitá di anche solo una delle condizioni rende il problema mal-posto.

In veritá, l'esistenza della soluzione si puó forzare allargando lo spazio delle soluzioni o riformulando il problema (ad esempio ai minimi quadrati), mentre si puó ottenere l'unicitá imponendo opportuni vincoli al modello. Quindi supporremo sempre che il problema analitico sia ben-posto.

La ricerca della stabilitá é piú complicata, anche se si basa ancora su riformulazioni del problema e aggiunta di informazioni vincolanti le soluzioni: questa é la natura del mal-posizionamento che dobbiamo considerare e studiare.

I problemi continui mal-posti possono essere formulati l'espressione

$$\int_{\Omega} input * sistema d\Omega = output$$

In tale formulazione, il problema diretto consiste nel calcolare l'output, noti l'input e la descrizione matematica del sistema, sullo spazio funzionale Ω . Nei problemi inversi, invece, si vuole ricavare l'input che produce la data misurazione dell'output, spesso affetta da varie forme di rumore o errori.

Un classico esempio di problema lineare mal posto é l'equazione integrale

di Fredholm di prima specie, definita nello spazio funzionale di Hilbert L^2 , ossia:

$$\int_0^1 K(s, t) \cdot f(t) dt = g(s) \quad \forall s \in [0, 1] \quad (1.1)$$

dove $f, g \in L^2(\Omega)$ e $K \in L^2(\Omega \times \Omega)$ é il *kernel* dell'equazione.

Si suppone di conoscere K esattamente, grazie al modello matematico sottostante, mentre il dato g corrisponde solitamente a quantità misurate, quindi é noto con una certa accuratezza solo in un insieme finito di punti s_1, s_2, \dots, s_m . Per questo motivo l'equazione integrale di Fredholm é spesso presentata in letteratura nella formulazione continua sulla sola variabile t :

$$\int_0^1 k_i(t) \cdot f(t) dt = g_i \quad \forall i = 1, 2, \dots, m$$

dove $k_i(t) := K(s_i, t)$ e $g_i := g(s_i)$.

Le difficoltà legate alle equazioni integrali sono strettamente connesse alla compattezza dell'operatore K associato al kernel K , che ha un effetto di smoothing sull'input f .

1.1.1 L'espansione a valori singolari SVE

Uno dei migliori metodi per analizzare l'equazione integrale di Fredholm di prima specie, nel caso in cui il kernel sia a quadrato sommabile, é studiare l'espansione in valori singolari (*SVE*) del kernel stesso.

Infatti, per $K \in L^2$ esiste l'espansione in valori singolari, cioè é possibile scrivere $K(s, t)$ come serie

$$K(s, t) = \sum_{i=1}^{\infty} \sigma_i u_i(s) \cdot v_i(t) \quad (1.2)$$

dove u_i e v_i , dette *funzioni singolari*, sono ortonormali rispetto al prodotto interno e i *valori singolari* σ_i sono reali non negativi, ordinati in successione non crescente.

4 1. La regolarizzazione per la risoluzione di problemi lineari inversi

Si osserva che, poiché K é un operatore a quadrato sommabile, deve risultare $\|K\|^2 := \int_0^1 \int_0^1 K(s,t)^2 ds dt < \infty$. Ne segue che

$$\|K\|^2 = \sum_{i=1}^{\infty} \sigma_i^2 < \infty,$$

quindi i valori singolari devono decrescere piú veloci di $i^{-\frac{1}{2}}$, altrimenti non si ha convergenza.

Inoltre

$$f(t) = \sum_{i=1}^{\infty} \frac{\langle v_i, f \rangle}{\sigma_i} v_i(t) \quad (1.3)$$

mentre, per definizione di SVE, risulta anche

$$\begin{cases} K v_i = \sigma_i u_i \\ K u_i = \sigma_i v_i \end{cases} \quad (1.4)$$

da cui la formulazione

$$\int_0^1 K(s,t) v_i(t) dt = \sigma_i u_i.$$

che é forse una delle relazioni piú importanti della SVE, dato che mostra come ogni v_i viene mappata nella corrispondente u_i con amplificazione dettata dal valore singolare σ_i . Inserendo tale risultato nell'equazione integrale 1.1, considerando ancora la definizione di SVE (1.2), si ricava

$$\begin{aligned} g(s) &= \int_0^1 K(s,t) \left(\sum_{i=1}^{\infty} \langle v_i, f \rangle v_i(t) \right) dt = \sum_{i=1}^{\infty} \int_0^1 \langle v_i, f \rangle K(s,t) \cdot v_i(t) dt = \\ &= \sum_{i=1}^{\infty} \sigma_i \langle v_i, f \rangle u_i(s) \end{aligned}$$

Ora, poiché

$$g(s) = \sum_{i=1}^{\infty} \langle u_i, g \rangle u_i(s),$$

risulta

$$f(t) = \sum_{i=1}^{\infty} \frac{\langle u_i, g \rangle}{\sigma_i} v_i(s) \quad (1.5)$$

Da quest'ultimo risultato si vede che chiedere l'esistenza della soluzione f , ossia che $g \in \text{Range}(K)$, equivale ad accertarsi che i suoi coefficienti $\frac{\langle u_i, g \rangle}{\sigma_i}$ siano ben definiti, per ogni $i = 1, 2, \dots$. É quindi necessaria l'ipotesi che $\sigma_i \neq 0 \quad \forall i$. In tal caso, f sar  completamente determinata dai suoi coefficienti e dalle funzioni singolari v_i , come mostrato nell'equazione 1.5.

L'andamento dei valori singolari   certamente fondamentale ed   strettamente collegato sia al comportamento delle funzioni singolari che alle propriet  del kernel K . Pi  precisamente, quanto pi  velocemente i σ_i tendono a zero, tanto pi  il kernel ha un effetto di smoothing sui dati e, si   osservato sperimentalmente, le funzioni u_i e v_i mostrano forti oscillazioni e numerosi zero-crossing. Queste propriet  rendono molto instabile la ricostruzione dell'input f con la formula 1.5 e possono fornire un indice del mal condizionamento del problema stesso.

1.1.2 La condizione continua di Picard

Considerando ancora l'equazione 1.5,   importante analizzare anche il contributo dato dal termine g , che compare nella definizione dei coefficienti di f .   chiaro infatti che non tutte le funzioni g rendono sia convergente la serie sopracitata che una soluzione $f \in L^2$. Nasce cos  la condizione di Picard, la quale afferma che affin  esista la soluzione a quadrato sommabile f , il termine g deve soddisfare

$$\sum_{i=1}^{\infty} \left(\frac{\langle u_i, g \rangle}{\sigma_i} \right)^{\frac{1}{2}} < \infty \quad (1.6)$$

In altri termini, si richiede che la successione $\langle u_i, g \rangle$ decada pi  velocemente della successione dei valori singolari e questo   un vincolo per g molto pi  forte del chiedere che $g \in L^2$. In tal caso, $g \in \text{Range}(K)$.

Se $g \notin \text{Range}(K)$, sicuramente non esiste una soluzione f al sistema, a quadrato sommabile:   quindi opportuno verificare che sia soddisfatta la condizione di Picard, prima di applicare il metodo scelto di risoluzione per il sistema inverso. Se ci si accorge che $g \notin \text{Range}(K)$, sia g_k l'approssimazione di g

6 1. La regolarizzazione per la risoluzione di problemi lineari inversi

ottenuta troncando l'espansione SVE dopo k termini, ossia

$$g_k = \sum_{i=1}^k \langle u_i, g \rangle u_i$$

Sicuramente tale funzione soddisfa la condizione di Picard, per ogni scelta di k , quindi é possibile calcolare un'approssimazione di f , detta f_k , come

$$f_k = \sum_{i=1}^k \frac{\langle u_i, g \rangle}{\sigma_i} v_i$$

Si osserva che mentre

$$g_k \xrightarrow{k \rightarrow \infty} g \in L^2,$$

lo stesso limite non vale per le f_k poiché

$$\|f_k\|_2 = \langle f_k, f_k \rangle^{\frac{1}{2}} \xrightarrow{k \rightarrow \infty} \infty.$$

Questo corrisponde alla mancanza di stabilit  che rende l'equazione integrale 1.1 mal posta.

Sfortunatamente nelle situazioni reali non si conosce mai il valore esatto di g , ma solo una sua approssimazione dovuta alla comparsa di inevitabili perturbazioni sui dati acquisiti e ora riassunti nella componente η , come segue:

$$g = g^{exact} + \eta \quad \text{dove } \|\eta\|_2 < \|g^{exact}\|_2$$

Ne segue che dobbiamo necessariamente calcolare

$$f = \sum_{i=1}^{\infty} \frac{\langle u_i, (g^{exact} + \eta) \rangle}{\sigma_i} v_i$$

e non possiamo aspettarci che il termine $g^{exact} + \eta$ appartenga al range dell'operatore K . Sorge da questo aspetto la necessit  di introdurre metodi di regolarizzazione, per riformulare il sistema assegnato come un problema avente soluzione stabile.

1.2 Problemi discreti, inversi e mal posti

Tipicamente, i problemi discreti mal posti nascono dai problemi continui descritti dall'equazione integrale di Fredholm 1.1, dove la discretizzazione del kernel K produce una matrice mal condizionata. Per precisione, i problemi con matrici A mal condizionate si distinguono in due categorie: la prima é formata dai sistemi dove la matrice A ha rango non massimo, la seconda dai problemi dove A é mal posta. Noi analizzeremo solo quest'ultima classe di problemi.

É doverosa una nota sui problemi *rank-deficient*, ossia quelli in cui la matrice A ha rango numerico non massimo e un gap elevato tra i valori singolari alti e quelli prossimi allo zero (il che non toglie che il rango analitico di A possa essere massimo e, quindi, che il problema analitico ammetta un'unica soluzione). Il mal condizionamento é dovuto strettamente ad aspetti numerici del problema: la presenza di valori singolari di A molto piccoli fa sí che una o piú righe e colonne della matrice siano quasi combinazioni lineari di altre righe o colonne. Ne segue che A contiene informazioni "ridondanti": un buon metodo di risoluzione deve per tanto estrapolare le informazioni linearmente indipendenti dalla matrice e costruire un nuovo problema con matrice ben-condizionata.

1.2.1 La SVD

Considerando il caso discreto, l'espansione a valori singolari SVE si traduce nella decomposizione a valori singolari SVD.

Se $A \in \mathbb{R}^{m \times n}$ (con $m > n$), la sua SVD é definita come

$$A = USV^t = \sum_{i=1}^n u_i \sigma_i v_i^t \quad (1.7)$$

dove $U = (u_1, u_2, \dots, u_n) \in \mathbb{R}^{m \times n}$ e $V = (v_1, v_2, \dots, v_n) \in \mathbb{R}^{n \times n}$ sono matrici con colonne ortogonali (per cui vale $U^t U = V^t V = I_n$), mentre S é una matrice diagonale di elementi $(\sigma_1, \sigma_2, \dots, \sigma_n)$ non negativi e in ordine non crescente, chiamati "*valori singolari*". I vettori u_i e v_i sono invece detti

8 1. La regolarizzazione per la risoluzione di problemi lineari inversi

rispettivamente "vettori singolari sinistri e destri" della matrice A .

Considerando la SVD, la matrice A mappa vettori x in

$$b := Ax = USV^t x = \sum_{i=1}^n \sigma_i (v_i^t \cdot x) u_i$$

per cui é evidente che eventuali alte frequenze del vettore x vengono attenuate dal prodotto con i valori singolari. Al contrario, risolvere il problema inverso e determinare x amplifica le oscillazioni ad alta frequenza presenti nel termine noto b .

Nel caso di problemi discreti mal posti, si verificano inoltre le seguenti proprietà:

1. i valori singolari σ_i decadono a zero gradualmente, senza particolari salti nello spettro, e gli ultimi tendono a livellarsi verso un valore τ ;
2. all'aumento di dimensione della matrice, corrisponde l'aumento di valori singolari piccoli;
3. i vettori singolari tendono ad aumentare i cambi di segno tra i loro elementi, con l'avanzare dell'indice i , cioè col ridursi dei valori singolari;
4. vettori e valori singolari si legano ad A anche con le relazioni

$$\begin{cases} A v_i = \sigma_i u_i \\ A u_i = \sigma_i v_i \end{cases} \quad \text{da cui} \quad \begin{cases} \|A v_i\|_2 = \sigma_i \\ \|A u_i\|_2 = \sigma_i \end{cases} \quad \forall i = 1, 2, \dots, n.$$

Come quest'ultimo punto riprende il sistema 1.4, così l'equazione 1.3 trova la sua formulazione corrispondente in

$$x = \sum_{i=1}^n (v_i^t \cdot x) v_i$$

da cui si ricava

$$x = \sum_{i=1}^{\text{rank}(A)} \frac{(u_i^t \cdot b)}{\sigma_i} v_i$$

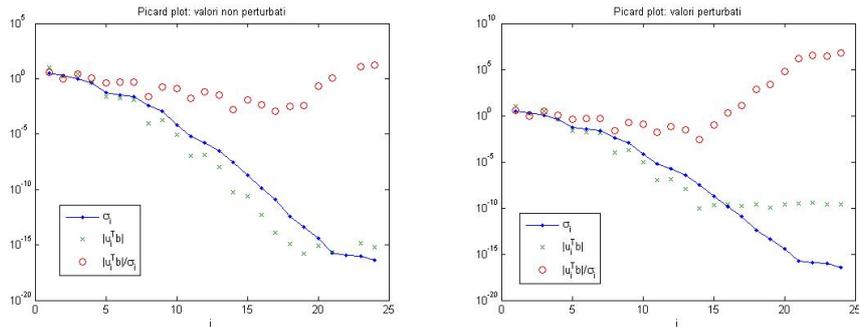
che corrisponde alla soluzione dei minimi quadrati, con la matrice pseudo-inversa $A^\dagger = \sum_{i=1}^{\text{rank}(A)} v_i \sigma_i^{-1} u_i$.

1.2.2 La condizione discreta di Picard

Vediamo ora come viene riformulata, nel caso discreto, la condizione di Picard precedentemente mostrata.

Sia b^{exact} il termine noto esatto. Si dice soddisfatta la *condizione discreta di Picard* se, per tutti i valori singolari $\sigma_i > \tau$, i corrispondenti coefficienti $(u_i^t \cdot b^{\text{exact}})$ decadono a zero piú velocemente dei σ_i stessi.

La figura 1.1 mostra graficamente l'analisi che deriva dall'enunciato della



(a) Caso in cui la condizione di Picard é verificata (b) Caso in cui la condizione di Picard non é verificata

Figura 1.1: Esempi per l'analisi di Picard, nel caso discreto

condizione di Picard: in blu sono riportati i valori singolari, nel loro andamento regolare e decrescente, in verde sono rappresentate le norme dei coefficienti $|u_i^t \cdot b|$, mentre in rosso ci sono i valori dei coefficienti della soluzione $\frac{|u_i^t \cdot b|}{\sigma_i}$. Nel caso (a), dove la successione verde decresce piú velocemente della blu (se non per pochissimi punti finali), i coefficienti rossi assumono sempre valori limitati che rendono la soluzione finale convergente. Al contrario, la figura (b) riporta un caso dove la successione $|u_i^t \cdot b|$ é quasi costante per l'ultima metà dei suoi valori e rimane quindi maggiore dei valori singolari per molti punti: in corrispondenza di questi indici i , i coefficienti $\frac{|u_i^t \cdot b|}{\sigma_i}$ crescono, assumendo valori

anche superiori di 10^6 che portano ad una soluzione altamente irregolare e inaffidabile.

1.2.3 Tipologie e analisi del rumore

La figura 1.1 paragona due grafici relativi a due sistemi molto simili tra loro: il primo é stato creato con un problema test esatto, del tipo $Ax = b$, mentre il secondo corrisponde al sistema $Ax = b + \eta$ dove il termine noto é stato perturbato con una piccola componente η , ma la matrice rimane identica. Nonostante η consista in un vettore di valori molto bassi, il grafico (b) mostra che la sua influenza non é affatto irrilevante.

Come già accennato in precedenza, il termine η rappresenta quella componente di rumore che é presente in ogni problema reale, dove i dati vengono acquisiti da macchinari con un margine di incertezza che si può solo studiare o stimare, ma mai annullare. Molto frequente é il *rumore bianco*, un particolare tipo di rumore spaziale, caratterizzato dall'assenza di periodicità nel tempo e da un'ampiezza costante su tutto lo spettro delle frequenze. Un descrittore del rumore spaziale ampiamente utilizzato in letteratura ([7]) é il comportamento statistico dei valori di intensità nella componente η : vedendone i valori come una variabile casuale z , si descrive statisticamente il rumore con la relativa densità di probabilità $p(z)$.

Nella pratica, vengono molto spesso utilizzati modelli di rumore gaussiani, descritti da

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

dove μ e σ rappresentano rispettivamente la media e la deviazione standard delle intensità z . Tale distribuzione gode di numerose proprietà che rendono questo modello conveniente, anche in situazioni in cui non rappresenta la miglior scelta.

Altri modelli che descrivono il rumore nell'imaging medico sono modelli di

rumore impulsivo bipolare, rappresentati da

$$p(z) = \begin{cases} P_a & z = a \\ P_b & z = b \\ 0 & \text{altrimenti} \end{cases}$$

e di rumore esponenziale, dati da

$$p(z) = \begin{cases} ae^{-az} & z \geq 0 \\ 0 & z < 0 \end{cases}$$

In particolare, se $0 < P_a, P_b < 1$ e se $a \ll b$, il rumore impulsivo é meglio noto come *rumore sale-e-pepe*, perché sulle immagini appare come un disturbo granulare bianco e nero.

Per misurare l'entità del rumore e quantificarne quindi il disturbo, sono state definite diverse grandezze. La piú popolare é il *rapporto segnale-rumore*, indicato con l'acronimo SNR (dall'inglese *signal to noise ratio*) ed espresso come

$$SNR = \frac{\|b + \eta\|_2^2}{\|\eta\|_2^2}$$

Tale quantità indica la potenza del segnale rispetto a quella del rumore e fornisce una misura del livello di informazione del dato reale, al livello della potenza del rumore. Teniamo infatti presente che, nelle applicazioni reali, non é mai noto il termine esatto ma solo la quantità $b^{exact} + \eta$, che d'ora in poi indicheremo sempre come termine noto b .

1.3 Metodi diretti di regolarizzazione

Nel capitolo precedente abbiamo visto che, trattando problemi mal posti, la principale difficoltà che si incontra é la presenza di valori singolari molto piccoli, dovuti al kernel K : risulta quindi necessario incorporare ulteriori informazioni, relative alla soluzione desiderata, in modo da stabilizzare il problema e ottenere un output stabile e veramente utilizzabile. Questo é l'obiettivo

12 1. La regolarizzazione per la risoluzione di problemi lineari inversi

della *regolarizzazione*.

In linea teorica, nei casi continui esistono moltissime tipologie di informazioni con cui integrare la soluzione f , ma l'approccio alla regolarizzazione dominante in letteratura utilizza la funzione di norma residua

$$\rho(f) = \left\| \int_0^1 K(s, t)f(t)dt - g(s) \right\|_2$$

e implementa uno dei seguenti schemi:

1. minimizzare $\rho(f)$ col vincolo che la soluzione f appartenga a una specifica classe di funzioni, $f \in S_f$;
2. minimizzare $\rho(f)$ col vincolo che una misura $\omega(f)$ della dimensione della soluzione sia minore una prefissata soglia δ ;
3. minimizzare $\omega(f)$ col vincolo che $\rho(f) < \alpha$, dove α é una soglia prefissata;
4. minimizzare una combinazione lineare delle due quantità $\rho(f)^2$ e $\omega(f)^2$, nella forma $\min\{\rho(f)^2 + \lambda^2\omega(f)^2\}$ dove λ é un parametro di peso prefissato.

In tale contesto, δ, α, λ sono chiamati *parametri di regolarizzazione*, mentre la funzione $\omega(f)$ é detta *smoothing norm*. Come vedremo piú avanti, la scelta di tali variabili é di particolare rilievo nella determinazione di un buon output. Da un punto di vista statistico, introdurre la regolarizzazione riduce la dimensione della matrice di covarianza della soluzione ma aumenta il bias della f stessa.

Per risolvere numericamente il problema discreto $Ax = b$ che ne deriva, occorre discretizzare anche i metodi di regolarizzazione. L'approccio piú frequente prevede di regolarizzare il problema continuo e di discretizzare subito dopo, poiché questo consente di controllare meglio l'aggiunta di informazioni della regolarizzazione stessa. Trattando equazioni integrali, si

utilizzano i metodi di quadratura o i metodi di Galerkin, che forniscono approssimazioni \tilde{x} e \tilde{b} rispettivamente dei vettori x, b , così il sistema discreto ha forma

$$A\tilde{x} = \tilde{b} \quad (1.8)$$

quadrata o sovradeterminata. Considerando i quattro schemi possibili per la regolarizzazione, occorre determinare un opportuno spazio $S_{\tilde{x}}$ o la misura $\Omega(\tilde{x})$: per esempio, il quarto schema diventa

$$\min\{\|A\tilde{x} - \tilde{b}\|_2^2 + \lambda^2\Omega(\tilde{x})^2\}$$

e, fissata $\Omega(\tilde{x}) := \|L\tilde{x}\|_2$ dove L é un operatore matriciale, questo metodo prende il nome di *regolarizzazione discreta di Tikhonov*.

Questi modelli possono essere ulteriormente raffinati in diversi modi. Per esempio, se si conosce una stima a priori x^* della soluzione ottimale, tale informazione può essere accorpata nella funzione Ω definendo

$$\Omega(\tilde{x}) := \|L(\tilde{x} - x^*)\|_2 \quad (1.9)$$

il che renderá l'output finale piú simile alla stima x^* .

I metodi che vedremo in seguito sono esprimibili come filtro per la soluzione \tilde{x} , cioè sottoforma di matrice diagonale $\Phi = \text{diag}(\Phi_1, \Phi_2, \dots, \Phi_n)$ tale per cui

$$\tilde{x} = \Phi A^{-1}b$$

Questa particolaritá può essere sfruttata sia per l'implementazione dei metodi, che per la loro descrizione e visualizzazione in forma breve e compatta.

Quelli che analizzeremo sono solo i metodi diretti di regolarizzazione: esistono anche diversi metodi iterativi, in cui ogni vettore di iterazione x^k può essere pensato come un'approssimazione della soluzione e l'indice d'iterazione k gioca quindi il ruolo del parametro di regolarizzazione. Inoltre, questi metodi non necessitano di operazioni computazionalmente costose (come la svd), quindi sono preferibili quando le dimensioni del sistema sono elevate. Poiché

14 1. La regolarizzazione per la risoluzione di problemi lineari inversi

avremo a che fare con problemi a piccole dimensioni, non ne approfondiamo l'analisi.

1.3.1 Metodo di Tikhonov

Entriamo ora nello specifico del metodo di Tikhonov, pensato nel secolo scorso da questo matematico russo.

L'idea che sta alla base di questo metodo é di incorporare al modello di partenza alcune assunzioni fatte a priori, sia sulla dimensione che sul livello di smoothness della soluzione desiderata: considerando già il caso discreto, questo si riflette nella scelta della funzione di smoothing, Ω .

Fissiamo ora un generico problema discreto mal posto, $Ax = b$ dove A é una matrice quadrata di dimensione n . La regolarizzazione di Tikhonov nella sua versione piú generale ha allora la seguente formulazione:

$$\min\{\|Ax - b\|_2^2 + \lambda^2\|Lx\|^2\} \quad (1.10)$$

dove il parametro λ indica quanto incide il peso della minimizzazione del termine di regolarizzazione, rispetto alla minimizzazione della norma del residuo del sistema originale.

La formulazione 1.10 é equivalente a

$$\min \left\| \begin{pmatrix} A \\ \lambda L \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2. \quad (1.11)$$

Da quest'ultima si capisce meglio che se gli spazi *null* di A e di L hanno intersezione triviale, ossia $\mathcal{N}(A) \cup \mathcal{N}(L) = \{0\}$, allora la soluzione minimizzante é unica.

Risolvendo invece il problema 1.10, mediante la ricerca dello zero del gradiente della quantità da minimizzare, si riformula il problema come

$$(A^t A + \lambda^2 L^t L)x = A^t b \quad (1.12)$$

e si ricava quindi una formulazione esplicita della soluzione, ossia

$$x_{L,\lambda} = (A^t A + \lambda^2 L^t L)^{-1} \cdot A^t b$$

Dal punto di vista computazionale, il sistema lineare 1.12 non si risolve mai col calcolo diretto della matrice inversa $(A^t A + \lambda^2 L^t L)^{-1}$, ma si utilizzano metodi quali la fattorizzazione di Cholesky o l'algoritmo di bidiagonalizzazione, in base alle caratteristiche e alla dimensione del sistema stesso.

Per precisione, si osserva che per $\lambda = 0$ il metodo di Tikhonov si riduce al metodo dei minimi quadrati, poiché $x_{L,\lambda} = x_{LS} = A^t A \cdot x = A^t b$, mentre per $\lambda \rightarrow \infty$ si ottiene la soluzione nulla, per l'eccessivo smooth.

Se si effettua la svd generalizzata per (A, L) , é semplice mostrare che la regolarizzazione di Tikhonov si può convertire in un filtro, di fattori Φ_i , da applicare alla ricostruzione diretta della f . Nel caso più semplice, in cui $L := I_n$, si hanno

$$\Phi_i := \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \quad \forall i = 1, \dots, n$$

ossia

$$\Phi_i \approx \begin{cases} 1 & \sigma_i \gg \lambda \\ \frac{\sigma_i^2}{\lambda^2} & \sigma_i \ll \lambda \end{cases}.$$

Analogamente, per $L \neq I_n$, vale

$$\Phi_i := \frac{\gamma_i^2}{\gamma_i^2 + \lambda^2} \quad \forall i = 1, \dots, n.$$

1.3.2 Metodo della TSVD

Dalla scomposizione ai valori singolari, é possibile ricavare un metodo di regolarizzazione molto semplice e intuitivo.

Consideriamo il sistema $Ax = b$ dove la matrice A , $n \times n$, sia numericamente rank-deficient e sia, per tanto, r_ϵ il suo rango numerico, mentre $rank(A) = n$ sia massimo. L'idea alla base del metodo é di ignorare le componenti della SVD associate a quei valori singolari talmente piccoli e tali per cui $r_\epsilon < n$: poiché i valori singolari sono in ordine decrescente, questo equivale a utilizzare tutti e soli i σ_i per $i \leq r_\epsilon$. In altri termini, la soluzione regolarizzata sarà

16 1. La regolarizzazione per la risoluzione di problemi lineari inversi

data da

$$x_{tsvd} = \sum_{i=1}^{r_\epsilon} \frac{u_i^t b}{\sigma_i} v_i \quad . \quad (1.13)$$

Si pensi alla soluzione di un sistema ai minimi quadrati, la cui norma é

$$\|x_{LS}\|_2^2 = \sum_{i=1}^n \left(\frac{u_i^t b}{\sigma_i} \right)^2$$

Questo valore é sempre molto elevato per la presenza di valori σ_i molto bassi al denominatore, a meno che non si verifichi $|u_i^t b| < \sigma_i$ per tutti gli ultimi $n - r_\epsilon$ valori singolari. In realtà, tale condizione é raramente verificata, per la presenza di errori nel termine noto b : sono gli ultimi valori che necessariamente rovinano la ricostruzione con la svd e la TSVD tiene proprio conto di questo limite.

Poiché l'intero r_ϵ non é esattamente calcolabile, l'equazione 1.13 viene implementata come

$$x_x = \sum_{i=1}^k \frac{u_i^t b}{\sigma_i} v_i \quad (1.14)$$

dove k indica l'indice dopo il quale non vengono piú considerate le componenti per la ricostruzione della soluzione. D'altra parte, si suppone che sia $k = r_\epsilon$, poiché $k < r_\epsilon$ comporta la perdita delle informazioni associate a valori singolari alti ($\sigma_{k+1}, \dots, \sigma_{r_\epsilon}$), mentre $k > r_\epsilon$ significa includere alcune componenti instabili nella soluzione.

Da un punto di vista geometrico, la TSVD corrisponde a sostituire la matrice mal posta A con la matrice $A_k := \sum_{i=1}^k u_i \sigma_i v_i^t$, che rimane di dimensione $n \times n$ ma ha rango (analitico e numerico) pari a k . Il problema ora può essere riformulato ai minimi quadrati come $\min \|A_k x - b\|_2$ e la soluzione di minima norma, cioè vincolata ad essere $x_k = \operatorname{argmin} \|x\|_2$, é unica ed esprimibile con l'equazione 1.14.

Osserviamo infine che la TSVD può essere vista come variante filtrata della ricostruzione con la SVD: definendo infatti i fattori di filtro

$$\Phi_i = \begin{cases} 1 & i \leq k \\ 0 & k < i \leq n \end{cases},$$

la soluzione x_k equivale a $x_k = \sum_{i=1}^n \Phi_i \frac{u_i^t b}{\sigma_i} v_i$.

1.4 Scelta del parametro di regolarizzazione

Nessun algoritmo di regolarizzazione è completo se non contempla un metodo che determini il miglior parametro regolarizzante all'interno di un insieme di possibili valori.

Per la TSVD, il parametro ottimale k_{opt} va obbligatoriamente scelto tra gli interi compresi tra 1 e la dimensione n del sistema, quindi si lavora con un insieme discreto e limitato di valori; al contrario il metodo di Tikhonov necessita di un valore reale positivo, $\lambda_{opt} \in \mathbb{R}^+$, quindi il range è continuo e illimitato.

I metodi che vedremo sono indipendenti da queste caratteristiche dei metodi diretti di risoluzione: ciò permette di affrontare la trattazione dei prossimi algoritmi facendo riferimento, per esempio, alla sola variabile continua λ . Infatti, sotto certe condizioni, è dimostrabile che per ogni parametro k di troncamento per la TSVD esiste un valore λ_k per il metodo di Tikhonov tale per cui le due soluzioni regolarizzate sono molto vicine. Ovviamente a valori decrescenti di k corrispondono valori crescenti di λ , poiché questi corrispondono ai casi in cui si chiede maggior smooth sulla soluzione regolarizzata e viceversa.

Analisi preliminare dei metodi per la scelta del parametro

Consideriamo ora il dato esatto b^{exact} a cui corrisponde la soluzione esatta x^{exact} . Indichiamo poi con x_{reg} la soluzione regolarizzata con un qualsiasi

18 1. La regolarizzazione per la risoluzione di problemi lineari inversi

metodo diretto di risoluzione, applicato al sistema $Ax = b$: risulta

$$x_{reg} = x_{reg}(\lambda).$$

Fissiamo anche un metodo per scegliere il parametro ottimale, λ_{opt} . Ci si chiede allora quanto velocemente $x_{reg}(\lambda)$ converge a x^{exact} , per $\|e\|_2 := \|b - b^{exact}\|_2 \rightarrow 0$ e $\lambda \rightarrow 0$. L'obiettivo ovviamente é di sviluppare metodi la cui convergenza sia piú rapida possibile e questo porta a studiare l' "ordine di ottimalitá" dei metodi di regolarizzazione.

Un altro aspetto molto importante per i metodi di scelta del parametro é analizzare l'errore totale e^{tot} , che si commette avendo fissato il λ , per verificare se il parametro scelto sia effettivamente l'ottimale. Piú precisamente, detta $A^\#$ la matrice complessiva tale per cui il metodo di regolarizzazione calcola $x_{reg} = A^\#b$, risulta

$$\begin{aligned} e^{tot} &= x^{exact} - x_{reg} = A^{-1}b^{exact} - A^\#b = \\ &= (A^{-1} - A^\#)b^{exact} - A^\#e \end{aligned} \tag{1.15}$$

Da questa formulazione si distinguono bene le due componenti dell'errore totale, ossia la parte $(A^{-1} - A^\#)b^{exact}$ dovuta alla sola regolarizzazione e quella $A^\#e$ data dalle perturbazioni presenti nel termine noto b del sistema originale. A questo punto si pretende che il metodo di scelta del parametro restituisca il parametro λ_{opt} che riesca a bilanciare meglio l'errore di regolarizzazione con l'errore di perturbazione, in modo che la norma dell'errore totale sia minima. Concretamente, ci si accontenta di algoritmi che restituiscano un λ in grado di approssimare accuratamente il parametro ottimale.

Ad ogni modo, risolvere il problema

$$\lambda_{opt} := \operatorname{argmin}\{\|e^{tot}\|_2\} = \operatorname{argmin}\{\|x^{exact} - x_{reg}(\lambda)\|_2\}$$

non é banale, soprattutto per sistemi con matrice mal-condizionata, dove entrambe le componenti dell'errore totale sono funzioni di λ con variazioni molto lente.

Se l'errore di regolarizzazione é ben calcolabile a partire dal metodo scelto, ciò non é sempre vero per l'errore di perturbazione. Si distinguono dunque due classi principali di metodi, in base alle assunzioni fatte sulla $\|e\|_2$:

1. metodi basati sulla conoscenza (esatta o almeno ben stimata) di $\|e\|_2$;
2. metodi che non richiedono di sapere la quantità $\|e\|_2$ a priori, ma cercano di estrapolarla dal termine noto b .

Nella prima categoria, sicuramente piú vincolata dell'altra, rientra il metodo basato sul principio di discrepanza, che vedremo in seguito, mentre nella seconda si collocano il metodo della L-curva e il Cross-Validation che sono di utilizzo piú frequente.

Un risultato fondamentale per l'efficacia degli algoritmi é dato dal seguente teorema:

Teorema 1.4.1.

Se l'operatore é compatto e infinito-dimensionale, allora non é possibile trovare una strategia convergente per la scelta del parametro, che non rischierà esplicitamente la conoscenza di $\|e\|_2$.

Nei casi che affronteremo, la finitezza dei dati ci assicura la convergenza

$$\lim_{\|e\| \rightarrow 0} x_{reg} = x^{exact} \quad (1.16)$$

Concludiamo questo quadro introduttivo per i metodi di scelta del parametro di regolarizzazione, definendo alcune funzioni che torneranno utili successivamente.

Fissato un metodo di regolarizzazione, si considerino i fattori di filtro che lo caratterizzano e si pone

$$\rho(\lambda) = \begin{cases} \sum_{i=1}^n \Phi_i & L = I_n \\ n - p + \sum_{i=1}^p \Phi_i & L \neq I_n \end{cases} \quad (1.17)$$

come somma dei fattori Φ_i che dipendono dal parametro regolarizzante.

Si conviene nel porre anche

$$\mathcal{T}(\lambda) = \text{trace}(I_m - A A^\#) = m - (n - p) - \sum_{i=1}^p \Phi_i = m - \rho(\lambda) \quad (1.18)$$

20 1. La regolarizzazione per la risoluzione di problemi lineari inversi

dove $m \leq n$: tale quantità rappresenta il numero effettivo di gradi di libertà del problema e serve per definire

$$\mathcal{V}(\lambda) = \frac{\|A x_\lambda - b\|_2^2}{\mathcal{T}(\lambda)} \quad (1.19)$$

che può essere utilizzata per stimare alcune grandezze relative all'errore e .

1.4.1 Principio di discrepanza

Si considera il sistema $A x = b$ con A mal-condizionata e consistente, cioè tale per cui sia $A x^{exact} = b^{exact}$. Si suppone di sapere $\|e\|_2$ almeno con una certa precisione, quindi si fissa δ_e come sovrastima (calcolata a priori) dell'errore stesso, ossia $\delta_e \geq \|e\|_2$.

Poiché per definizione vale $\|e\|_2 := \|b - b^{exact}\|_2$, il metodo per la scelta del parametro di regolarizzazione basato sul principio di discrepanza consiste nel determinare λ tale per cui la norma del residuo sia uguale a δ_e ; in altri termini:

$$\lambda \quad \text{tale che} \quad \|A x_\lambda - b\|_2 = \delta_e, \quad \text{con } \delta_e \geq \|e\|_2 \quad (1.20)$$

Per parametri discreti di regolarizzazione, il metodo si traduce nella scelta del più piccolo intero k per cui valga $\|A x_k - b\|_2 \leq \delta_e$.

In entrambi i casi, il principio della discrepanza si basa sulla scelta di mantenere un certo livello di residuo, il che porta sempre a una soluzione regolarizzata più smooth rispetto alla soluzione esatta. Come già accennato, soprattutto nei casi in cui non si conosce il vero valore $\|e\|_2$, è importante avere una buona stima di tale quantità: un sovrastima abbondante porta infatti a uno smooth ancora più accentuato e, di conseguenza, eccessivo, mentre una sottostima, seppur piccola, rende la norma residua talmente bassa che la soluzione finale rischia di avere norma elevata, essendo troppo poco regolarizzata.

Principio della discrepanza compensato

Per evitare uno smoothing eccessivo, se si suppone di avere un rumore bianco, é possibile migliorare la formulazione del principio di discrepanza mostrato nell'equazione 1.20.

Assumiamo che e abbia media non nulla m e sia $\sigma_0^2 I_m$ la matrice di covarianza.

Il principio della discrepanza compensato si esprime come segue:

$$\begin{aligned} \|A x_\lambda - b\|_2 &= (\|e\|_2^2 - \sigma_0^2 \operatorname{trace}(A A^\#))^{1/2} \\ &= \sigma_0 (m - \rho(\lambda))^{1/2} = \sigma_0 \mathcal{T}(\lambda)^{1/2}. \end{aligned} \quad (1.21)$$

La "compensazione" deriva dal fatto che $\sigma_0^2 \sqrt{m}$ é una valida stima di $\|e\|_2$, ma non della norma residua. Entrambi i valori possono comunque essere determinati con lo studio della funzione $\mathcal{V}(\lambda)$ definita nell'equazione 1.19: plottandola rispetto λ , si nota una parte piana la cui ordinata corrisponde a σ_0^2 . Si ricava quindi $\|e\|_2^2 = m\sigma_0^2$.

1.4.2 L-curva

Uno dei metodi piú largamente diffusi per la ricerca di un buon parametro di regolarizzazione si basa sull'osservazione della cosiddetta *L-curva*. Questa altro non é che il grafico di una curva rappresentante tutti i possibili parametri di regolarizzazione ed é, pertanto, discreta se applicata alla TSVD, oppure continua, se applicata al metodo di Tikhonov (almeno a livello teorico, per ovvie ragioni numeriche).

Considerando tutte le soluzioni che si ottengono al variare del parametro di regolarizzazione, cioé le $x_{reg}(\lambda) = A^\# b$, si calcolano le norme del residuo

$$\|r(\lambda)\|_2 = \|(A^\#)^{-1} x_{reg}(\lambda) - b\|_2 \quad \forall \lambda > 0$$

e della soluzione

$$\|\psi(\lambda)\|_2 = \|x_{reg}(\lambda)\|_2 \quad \forall \lambda > 0$$

che sono viste come funzioni del parametro stesso.

Plottando la funzione $\log(\psi)$ rispetto a $\log(r)$, si ottiene poi una curva dalla

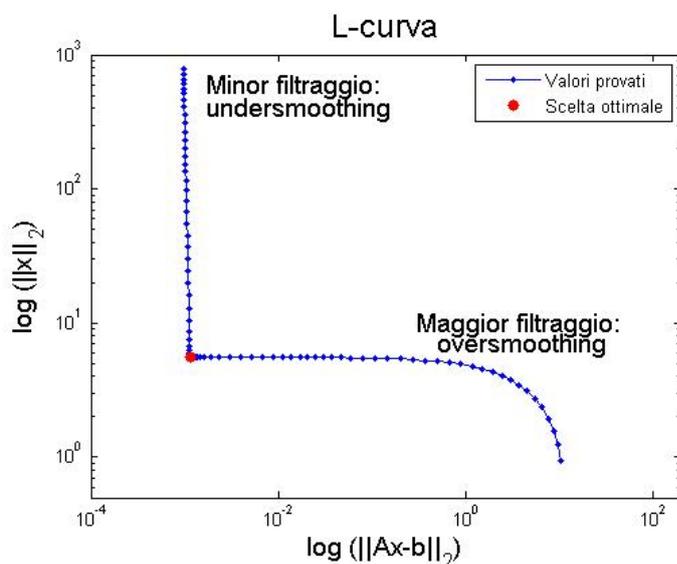


Figura 1.2: Rappresentazione della L-curva

forma a "L" (da cui il nome del metodo stesso). Tale plot esplicita, sul piano bidimensionale, il compromesso tra le minimizzazioni delle due quantità che caratterizzano la regolarizzazione. Soprattutto per problemi mal posti, si preferisce plottare la curva dei valori logaritmici, perché questo rende la forma a L più evidente e marcata, come mostrato nella figura 1.2.

Ora, una regolarizzazione molto piccola produce soluzioni poco corrette, perché viene attribuita troppa importanza alla minimizzazione della norma residuo: in questo caso si parla di *undersmoothing*. Inoltre, utilizzando parametri sotto una certa soglia e tendenti a zero, il valore di $\|(A^\#)^{-1} x_{reg}(\lambda) - b\|_2$ non viene più significativamente ridotto, mentre avanza la propagazione delle componenti affette da rumore, che alzano la norma della soluzione $\|x_{reg}(\lambda)\|_2$: si crea per questo motivo il tratto verticale della L-curva.

Al contrario, valori di $\lambda \rightarrow \infty$ (o comunque elevati) provocano l'*oversmoothing*: la regolarizzazione è talmente forte che prevale sulla "risoluzione" del sistema assegnato, quindi il residuo rimane alto ma la norma della soluzione sarà molto piccola. In questo modo si crea il tratto orizzontale della curva.

Risulta così evidente la comodità di plottare la curva in "log-log scale": di-

venta meglio localizzabile il vertice della L, che é il punto in cui le due norme si compensano al meglio, e il parametro a cui corrisponde questa coppia di norme é quindi il parametro ottimale da utilizzare. Infatti, in molti casi la forma della L non é precisa, quindi al posto di un vertice netto si trova una curva: l'algoritmo restituirá un punto prossimo al vertice "ideale".

Nel caso specifico del metodo di Tikhonov, é possibile raffinare l'algoritmo della L-curva, sfruttando il ruolo che l'operatore L riveste nel metodo. É infatti coerente con l'equazione 1.10 definire la norma di η come

$$\|\eta(\lambda)\|_2 = \|L x_{reg}(\lambda)\|_2$$

cioé considerare anche L nello smooth che contrasta con la minimizzazione del residuo. Vale inoltre il seguente teorema: a ogni coppia di valori (r, η) della L-curva é associato univocamente un λ , reale positivo, a cui corrisponde la soluzione $x_{reg}(\lambda)$ che rende norma η e residuo r . Questo risultato non é ovviamente valido per la TSVD, in cui la L-curva é rappresentata da un set finito di punti discreti e il parametro restituito dall'algoritmo é quello che corrisponde al punto piú vicino al vertice.

Aspetti computazionali della L-curva

Da un punto di vista strettamente teorico, il metodo della L-curva é molto semplice e intuitivo. Dal punto di vista computazionale, invece, presenta alcune caratteristiche meno positive.

Innanzitutto, per plottare l'intera L-curva é necessario risolvere il sistema $(A^\#)^{-1} x_{reg}(\lambda) - b$ per ogni valore prefissato del parametro λ e questo ha un costo computazionale non indifferente, soprattutto se il sistema ha dimensione elevata.

L'altro passaggio complicato é la localizzazione del vertice sulla curva: soprattutto nel caso della TSVD, in cui la curva consiste in n punti, trovare il punto di massima curvatura non é banale. La strategia piú diffusa prevede di approssimare la L-curva con una spline cubica, ma non é detto che questa

24 1. La regolarizzazione per la risoluzione di problemi lineari inversi

goda delle proprietà di smooth necessarie per il buon esito dell'algoritmo, soprattutto se i punti non sono disposti con una certa regolarità sulla "traccia" di una L.

Curve e metodi alternativi alla L-curva

Negli ultimi anni sono stati suggeriti alcuni metodi alternativi alla L-curva, preferibili qualora localizzarne il vertice risulti molto complicato o, addirittura, la curva non abbia l'andamento desiderato.

Miller suggerì di assumere a priori che la soluzione ottimale debba essere tale per cui

$$\|A x_{reg} - b\|_2 \leq \delta \quad \text{e} \quad \|x_{reg}\|_2 \leq \alpha$$

così

$$\bar{\lambda} = \frac{\delta}{\alpha}$$

produce una soluzione $x_{reg}(\bar{\lambda})$ molto vicina a quella data dal vertice della L-curva, se i valori δ e α sono buone stime di $\|e\|_2$ e $\|x^{exact}\|_2$.

Nel caso specifico della TSVD applicata a un sistema di dimensione media-bassa, in cui la L-curva è costituita da pochi punti, [8] propone di considerare anche le curve date dai punti

$$(\log(\|r_k\|), \log(k)) \tag{1.22}$$

e

$$(\log(\|r_k\|), \log(k(A_k))) \tag{1.23}$$

dove $k(A_k)$ è il numero di condizione della matrice A_k , calcolato come $k(A_k) = \frac{\sigma_1}{\sigma_k}$. Anche queste curve, dette rispettivamente *curva residuale* e *curva condizionale*, hanno andamenti a forma di "L": gli indici k_r e k_c , che corrispondono ai due vertici, sono possibili parametri di regolarizzazione, perché rappresentano un compromesso tra due caratteristiche importanti del problema e della regolarizzazione, come era per la L-curva classica.

In [9], Regińska propone altri due metodi, che non prevedono l'individuazione di alcun vertice. Vengono infatti definite le funzioni

$$\Phi_\nu(k) = \|r_k\| \cdot \|x_k\|^\nu \quad \text{con } \nu > 0 \quad (1.24)$$

e

$$\Psi_p(k) = \|r_k\|^{\frac{1}{p}} \cdot \|x_k\|^p \quad \text{con } p > 0 \quad (1.25)$$

ed é stato mostrato che minimizzare queste funzioni restituisce altre due stime del parametro k ottimale per la regolarizzazione. In particolare, Regińska ha dimostrato che tale minimizzazione é equivalente a determinare il vertice della L-curva formata dalle stesse norme $(\|x_k\|, \|r_k\|)$.

1.4.3 Generalised Cross Validation

Il Generalised Cross Validation (GCV) é un metodo molto potente e usato, per determinare il parametro di regolarizzazione. Essendo indipendente da una conoscenza a priori di $\|e\|_2$, si basa su considerazioni di ordine statistico e agisce in maniera "predittiva", poiché cerca di minimizzare il valore del presunto errore ai minimi quadrati $\|A x_\lambda - b^{exact}\|_2$. Dal momento che b^{exact} non é noto, il GCV definisce la funzione

$$\mathcal{G}(\lambda) = \frac{\mathcal{V}(\lambda)}{\mathcal{T}(\lambda)} = \frac{\|A x_\lambda - b\|_2^2}{\text{trace}(I_m - AA^\#)^2} \quad (1.26)$$

dove $\mathcal{V}(\lambda)$ e $\mathcal{T}(\lambda)$ sono state definite rispettivamente nelle equazioni 1.19 e 1.18. Il metodo GCV restituisce il parametro che minimizza la funzione $\mathcal{G}(\lambda)$, ossia

$$\lambda_{GCV} = \text{argmin}(\mathcal{G}(\lambda)) \quad (1.27)$$

Si può dimostrare che se la condizione discreta di Picard é soddisfatta e siamo in presenza di solo rumore bianco, allora questo λ_{GCV} é molto vicino al λ_{opt} che minimizza la norma $\|A x_\lambda - b^{exact}\|_2$ tanto che risulta $\lambda_{GCV} = \lambda_{opt}(1 + o(1))$, dove

$$o(1) \xrightarrow{m \rightarrow \infty} 0.$$

26 1. La regolarizzazione per la risoluzione di problemi lineari inversi

Teoremi di convergenza del GCV, applicato sia al metodo di Tikhonov che alla TSVD, sono stati dimostrati dall'americano Wahba e dal tedesco Vogel. In generale, si osserva che $\mathcal{T}(\lambda)$ é una funzione monotona crescente per $\lambda \geq 0$, i cui valori appartengono all'intervallo $[m-n, m-n+p]$. Minimizzare $\mathcal{G}(\lambda)$ corrisponde allora a trovare il punto di transizione in cui la funzione $\mathcal{V}(\lambda)$ passa dall'aver un andamento a variazione molto lenta, all'aver un incremento molto rapido.

Capitolo 2

Le TC perfusionali

2.1 Le neoplasie e la tomografia computerizzata

In patologia, un *tumore* (dal latino tumor, "rigonfiamento") o neoplasia (dal greco néos, "nuovo", e plásis, "formazione") é, nella definizione dell'oncologo R.A. Willis, "una massa anormale di tessuto che cresce in eccesso e in modo scoordinato rispetto ai tessuti normali, e persiste in questo stato dopo la cessazione degli stimoli che hanno indotto il processo". Nei tessuti normali le cellule si riproducono dividendosi, in modo da sopperire alle varie necessità dell'organismo, come far fronte a un maggiore fabbisogno in termini di quantità di cellule nell'organismo o di "qualità di cellule" per differenziamento tissutale, oppure per rimpiazzare le cellule morte o quelle che, danneggiate, vanno incontro a un processo di morte programmata. Nei tumori, invece, questo delicato equilibrio, che normalmente é governato dai messaggi chimici inviati da una cellula all'altra e dai geni che si trovano nel loro DNA, é compromesso da alterazioni geniche, dette *mutazioni*. Tali anomalie si sviluppano seguendo quattro fasi:

1. l'acquisizione dell'autonomia moltiplicativa e della capacità di formare nuovi vasi sanguigni per fornire ossigeno e nutrimento alle cellule tu-

- morali (angiogenesi);
2. l'acquisizione della capacità di replicazione illimitata;
 3. la perdita delle inibizioni da contatto e da densità;
 4. la riduzione e la perdita della possibilità di andare incontro, come le cellule nei tessuti normali, al processo di morte programmata (apoptosi).

In base a determinate caratteristiche, i tumori si suddividono in:

benigni che sono costituiti da cellule che mantengono in parte le loro caratteristiche morfologiche e funzionali, pur esibendo autonomia moltiplicativa. Essi sono caratterizzati da un tipo di sviluppo detto espansivo, perché comprime i tessuti vicini senza distruggerli, quindi presentano una crescita lenta e normalmente non si riformano dopo l'asportazione chirurgica;

maligni detti anche "*cancro*", che sono invece costituiti da cellule che appaiono morfologicamente e funzionalmente diverse dalle corrispondenti cellule sane. L'atipia morfologica, che è tanto maggiore quanto più indifferenziato è il tumore, si manifesta con mancanza di uniformità nella forma e nelle dimensioni delle cellule (polimorfismo) e degli organi cellulari, in particolare del nucleo. I tessuti vicini al tumore maligno vengono invariabilmente infiltrati e distrutti dalle cellule tumorali che si sostituiscono a quelle normali: questo fenomeno, tipico della malignità, è noto col termine "*invasività neoplastica*" ed è caratterizzato da una notevole rapidità di crescita e dall'elevata capacità di sviluppare metastasi;

borderline che presentano un comportamento intermedio fra la malignità e la benignità, o risultano inclassificabili (come i tumori a comportamento biologico incerto). In questi casi, generalmente la frequenza delle metastasi è molto bassa e il decorso è lento.

2.1.1 Diagnosi e terapie

Nella fase iniziale la crescita di un tumore é lenta, ma accelera quando raggiunge la dimensione di 1 cm^3 . A questo punto la massa tumorale puó essere diagnosticata mediante esami clinici radiologici, integrati da dati di laboratorio e dalle tecniche di diagnostica per le immagini. É però necessario un esame istologico dei tessuti neoplasici, per conoscerne la natura e il grado di malignitá.

La chirurgia é il principale metodo per affrontare la maggior parte dei tumori solidi e isolati ed é il trattamento di elezione nel caso di neoplasie benigne. Costituisce, inoltre, una parte fondamentale del processo diagnostico perché rende tangibile la massa tumorale e permette di eseguire biopsie. Nei casi di neoplasie benigne o di tumori confinati, la chirurgia generalmente tenta di resecare l'intera massa insieme, in alcuni casi, ai linfonodi locali. Per alcuni tipi di neoplasie l'intervento chirurgico é insufficiente a sconfiggere la malattia, quindi bisogna ricorrere a diverse tecniche terapeutiche, che possono essere utilizzate anche combinate tra loro. Per esempio:

- la radioterapia, che consiste nell'utilizzo di radiazioni ionizzanti per "bombardare" le cellule maligne, danneggiandone il DNA. Le cellule tumorali, infatti, sono scarsamente capaci di riparare i propri danni e quindi, se danneggiate, vanno incontro a morte cellulare. Per risparmiare tessuti sani, come la pelle o gli organi che la radiazione deve superare per arrivare al tumore, i fasci delle radiazioni vengono sagomati e rivolti da diverse angolazioni, intersecandosi nel centro della zona da trattare, dove perciò vi sará un quantitativo totale di dose assorbita superiore rispetto alle zone tissutali adiacenti;
- la chemioterapia, che miscela farmaci con capacitá di distruzione in maniera quanto piú mirata e selettiva possibile: é pertanto una terapia personalizzata, per ogni paziente e per la tipologia del suo tumore;
- l'ormonoterapia, che é una terapia ormonale capace di creare un ambiente ostile alla proliferazione cellulare. La si sceglie per curare certi tipi

di tumori, detti "ormono-dipendenti", poiché la loro crescita é stimolata da ormoni, quali estrogeni e androgeni, prodotti dall'organismo. La sua efficacia é proporzionale alla sensibilità delle cellule tumorali agli ormoni. Si utilizza principalmente contro il cancro della mammella, dell'utero, della prostata o della tiroide;

- l'immunoterapia, che prevede la somministrazione di molecole ad azione immunostimolante, prodotte dal sistema immunitario, che hanno la capacità di stimolare i leucociti affinché aggrediscano le cellule cancerogene, distruggendole;
- l'ipertermia, che consiste nel riscaldamento mediante onde radio dell'organo ammalato: le cellule neoplastiche sottoposte a temperature elevate (43-44 °C) subiscono un danno irreversibile, che riduce l'efficacia dei sistemi riparativi normalmente presenti a livello cellulare, giungendo quindi all'apoptosi.

2.1.2 La tomografia computerizzata

Analizziamo ora i mezzi di indagine diagnostica piú utilizzati.

In radiologia la tomografia computerizzata, indicata con l'acronimo TC o CT (dall'inglese "*computerized tomography*"), é una metodica diagnostica per immagini, che sfrutta radiazioni ionizzanti (i raggi X) per riprodurre sezioni o strati (le vere e proprie tomografie) del corpo del paziente ed effettuare elaborazioni tridimensionali. Per ottenere delle immagini é poi necessario un elaboratore di dati.

L'apparecchiatura é sostanzialmente costituita dal cosiddetto *gantry* (ossia dal tubo radiogeno con detettori e un sistema di controllo delle emissioni delle radiazioni ionizzanti), dal lettino dove viene posto il paziente e dalla workstation alla quale lavorano il tecnico e il radiologo. Nelle apparecchiature di tipo tradizionale (TAC, tomografie assiali computerizzate), il tubo radiogeno ruota attorno al soggetto che deve essere esaminato ed emette un fascio di radiazioni ionizzanti: man mano che le radiazioni emesse dal tubo

radiogeno attraversano il corpo del soggetto, subiscono un'attenuazione e una perdita di energia. Dopo l'attraversamento del corpo, le radiazioni incidono su specifici sensori, detti detettori: tali sensori misurano l'attenuazione che le radiazioni ionizzanti hanno subito, dopo l'attraversamento del corpo. I sensori, a questo punto, effettuano numerose misurazioni relative alle dosi di radiazioni che sono state assorbite in tutti i punti della zona scansionata; le misurazioni (i cosiddetti "*raw data*", cioè "*dati grezzi*") vengono quindi inviate all'elaboratore di dati che, attraverso complessi algoritmi di tipo matematico ricostruisce l'immagine tomografica. Terminata la scansione, il tubo radiogeno torna alla posizione originale e il lettino, dove é posizionato il paziente, si sposta; si passa quindi alla scansione successiva.

Da diversi anni si é passati all'utilizzo di apparecchiature TAC a spirale, piú sofisticate, che consentono di ottenere dati relativi a un volume corporeo e non a uno strato, come accadeva con le tecniche tradizionali. Nella TAC a spirale, sia la rotazione del tubo radiogeno che l'emissione delle radiazioni ionizzanti e gli spostamenti del lettino porta-paziente avvengono in contemporanea. Le apparecchiature TAC a spirale non sono tuttavia l'ultimo capitolo di questa tecnica diagnostica: un ulteriore passo avanti é rappresentato dalle TAC multistrato. Tali apparecchiature consentono, infatti, una qualità delle immagini notevolissima, una considerevole riduzione dei tempi per ciascun esame (é possibile valutare un distretto corporeo, come per esempio l'addome, in poco meno di 20 secondi) e la possibilità di acquisire l'intero volume corporeo in un'unica sequenza. Interessante é la possibilità, con le nuove apparecchiature TAC, di ottenere immagini tridimensionali e informazioni relative alla funzionalità del settore che é stato analizzato.

Nella maggior parte dei centri diagnostici viene poi rilasciato al paziente, insieme all'esito dell'esame, un cd contenente tutte le fasi dell'esame, cosicché lo specialista o il medico curante possano analizzare il tutto con l'utilizzo di un normale personal computer.

Poiché le immagini sono di tipo digitale, il corpo studiato viene suddiviso in una serie discreta di elementi di volume (voxel), ai quali corrisponde

un elemento unico d'immagine (pixel), rappresentato secondo una scala dei grigi: quanto piú piccolo é il volume rappresentato da un pixel, tanto maggiore sará la risoluzione spaziale dell'immagine di output. L'attenuazione qui riportata é direttamente proporzionale alla densitá elettronica dei tessuti presenti nel voxel: il suo valore é detto "densitometrico". Un voxel con alta densitá viene infatti rappresentato con una gradazione di grigio piú chiara. L'unitá di misura della densitá elettronica é l'UH (unitá di Hounsfield - HU), la cui scala comprende 2001 diverse tonalitá di grigio, che vanno dal nero (per rappresentare basse densitá elettroniche) al bianco (per rappresentare alte densitá elettroniche).

Inoltre, le dimensioni di una singola immagine sono normalmente di 512×512 pixel, per una profonditá di 16 bit/pixel. Infine, i dati elaborati sono rappresentati sul monitor sotto forma di sezioni, non necessariamente contigue, di spessore pre-impostato.

2.1.3 La TC perfusionale

La tomografia computerizzata perfusionale (TCp) é un tipo di esame TC, che necessita la somministrazione di un mezzo di contrasto iodato per via endovenosa e che permette di ottenere informazioni sulla funzionalitá dell'organo analizzato. Attraverso l'analisi dinamica della concentrazione del mezzo di contrasto nelle diverse regioni del tessuto esaminato, é possibile effettuare la stima di numerosi parametri emodinamici a livello di microvasi capillari, poiché si evidenzia un elevato apporto di sangue nelle zone tumorali, dove la densitá dei nuovi vasi é maggiore.

La TCp si basa su due requisiti tecnici indispensabili. Innanzitutto sulla ripetizione nel tempo delle scansioni (denominate scansioni dinamiche o cinetiche): le acquisizioni, infatti, vengono effettuate prima, durante e dopo la somministrazione del mezzo di contrasto per poter apprezzare le variazioni di densitá nel tempo. Inoltre, si osserva che i valori di densitá misurati dalla TC all'interno di ogni singolo voxel sono direttamente proporzionali alla quantitá del mezzo di contrasto presente all'interno dei vasi e all'espansione

dello spazio extra-vascolare o cellulare, di quel voxel. L'altra prerogativa di una TCp é la possibilità di selezionare un input arterioso: il posizionamento di una *region of interest* (ROI) su un vaso arterioso, che nutre il tessuto da esaminare, é indispensabile per calcolare i parametri di perfusione.

A causa dei fenomeni di neovascolarizzazione, nelle lesioni neoplastiche si verifica un aumento della perfusione che rende i parametri perfusionali buoni indicatori dell'angiogenesi tumorale. Pertanto, sebbene trovi il suo maggiore campo di applicazione nella patologia cerebro-vascolare, la TC perfusionale é applicabile con successo anche nello studio della patologia neoplastica.

I parametri perfusionali

Per quanto appena detto, vediamo ora i parametri perfusionali nello specifico.

In medicina, la *perfusione* é la somministrazione in cospicua quantità di una soluzione medicamentosa o di sangue per via venosa, o piú raramente (qualora si intenda esporre un organo o distretto corporeo, come un arto, a una piú alta concentrazione di un farmaco) per via arteriosa (perfusione regionale selettiva) e anche la protratta irrorazione (con sangue, soluzioni nutritizie o equilibratrici del pH) di un organo isolato dall'organismo, sia a scopo sperimentale sia, in cardiocirurgia, nel corso di interventi in circolazione extra-corporea.

Siccome la TCp é in grado di misurare la variazione di densità generata nei tessuti dall'apporto del mezzo di contrasto, questa tecnica diagnostica viene utilizzata in campo oncologico per valutare il processo di angiogenesi e lo stadio di evoluzione dei tumori, calcolando la perfusione come il trasporto di sangue all'unità di volume del tessuto nell'unità di tempo. Possiamo quindi calcolare alcuni parametri che quantificano la perfusione del tessuto analizzato, quali:

flusso sanguigno BF espresso in ml/ 100 g di tessuto/minuto, rappresenta il flusso di sangue che attraversa il letto vascolare del tumore: un valore maggiore puó indicare la presenza di un numero elevato di shunt

artero-venosi (comunicazione diretta fra un'arteria e una vena, tipica nei tessuti tumorali) e quindi di una massa tumorale;

volume sanguigno BV espresso in ml/100 g di tessuto, misura il volume del sangue che finisce nei vasi sanguigni del tessuto in esame e non di quello stagnante, per cui un valore alto ci rivela un anomalo processo di angiogenesi;

tempo di transito medio MTT espresso in secondi, ci indica il tempo medio impiegato dal sangue per attraversare le strutture micro vascolari dall'estremità arteriosa a quella venosa e, in questo caso, un valore basso è indice della presenza di una massa tumorale;

superficie di permeabilità PS espressa in ml/100 g di tessuto/minuto, misura il flusso unidirezionale del mezzo di contrasto dal compartimento vascolare a quello interstiziale: l'aumento della permeabilità dell'endotelio è indice dell'anomala presenza di vasi neoformati.

Le mappe colorimetriche Per effettuare un'analisi qualitativa dei parametri di perfusione, negli ultimi anni si è diffuso l'utilizzo di mappe colorimetriche che attribuiscono, ad ogni pixel dell'immagine TC, un colore rappresentante il valore numerico del parametro perfusionale calcolato per quel pixel. La scala dei colori, scelta dall'operatore, permette di visualizzare al meglio la distribuzione della perfusione all'interno di tutto il volume esaminato e di averne una rappresentazione dei diversi livelli, mettendo in evidenza le aree dove il tumore è maggiormente radicato. Inoltre le mappe colorimetriche permettono di vedere i margini della neoplasia e questa indagine è utilissima prima di un eventuale intervento chirurgico e dopo i cicli di terapie, per controllarne l'efficacia. Nel capitolo quarto di questa tesi vedremo numerose mappe colorimetriche, relative al parametro BF.

2.1.4 Considerazioni generali

Ovviamente, oltre alla TCp esistono altre tecniche, non invasive, di imaging che permettono di monitorare le caratteristiche vascolari dei tumori, come le tecniche dinamiche ad ultrasuoni (DCE-US), la tomografia ad emissione di positroni (PET) e la risonanza magnetica dinamica (DCE-MR). Tuttavia la TCp é preferibile per i seguenti motivi:

1. é adatta a tutti i pazienti, bambini compresi;
2. ha un ampio spettro di patologie indagabili;
3. permette una rapida esecuzione;
4. ha un'elevata risoluzione spaziale e temporale;
5. ha un costo esecutivo relativamente basso;
6. é in grado di effettuare misurazioni esatte e ripetibili delle lesioni neoplastiche.

D'altra parte, ci sono alcuni problemi ancora irrisolti: da un lato non é stato ancora accertato quali siano i parametri perfusionali piú importanti per la stima delle variazioni vascolari nel tumore in risposta a trattamenti specifici, dall'altro non sono ancora stati chiariti il ruolo del filtraggio spazio-temporale dei dati e le modalitá del posizionamento delle ROI; infine devono essere ulteriormente sviluppati metodi che consentano di confrontare tra loro risultati ottenuti con metodi analitici diversi.

Torneremo su questi "problemi aperti" nel quarto capitolo.

2.2 Il modello della TCp

2.2.1 La curva di concentrazione nel tempo, TCC

Come precedentemente accennato, dai dati ottenuti da una TCp é possibile selezionare una slice di tessuto per analizzarne i valori di attenuazione

prima, durante e dopo la somministrazione del mezzo di contrasto. Per ogni pixel e per ogni specifica porzione di tessuto é possibile riportare in un grafico i valori di concentrazione assunti dal segnale nel tempo, generando la curva della concentrazione nel tempo, TCC (dall'inglese *Time Concentration Curve*), avente in ascissa l'asse temporale e in ordinata i valori relativi al segnale proveniente dal pixel in analisi. Nella figura 2.1 ne é riportata un esempio: in rosso sono indicati i valori di concentrazione del segnale nel tempo, mentre in blu é riportata la curva che interpola i valori assunti dal segnale tissutale.

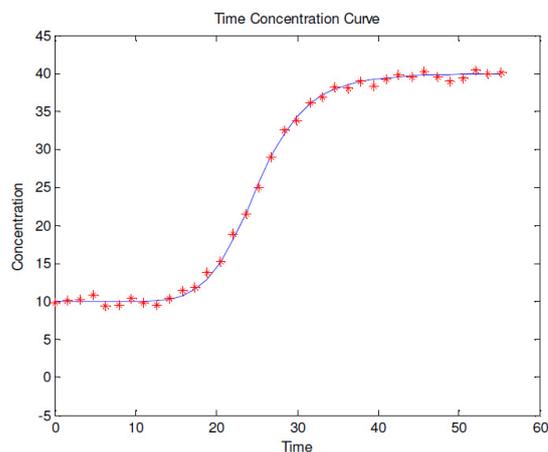
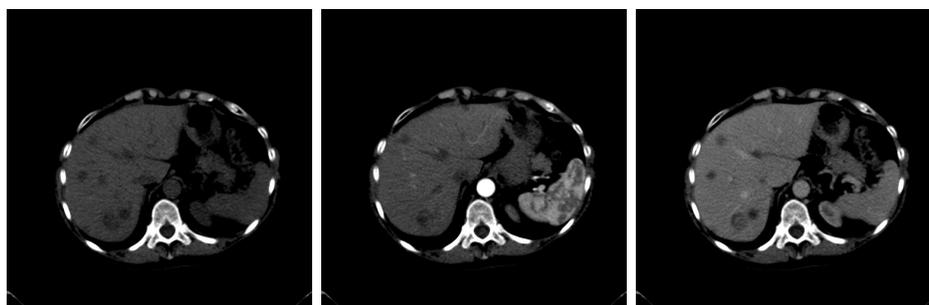


Figura 2.1: Esempio di TCC

La curva é di forma sigmoideale e nel suo andamento si distinguono principalmente tre fasi. La prima é detta "*baseline*", perché piuttosto stabile e riguarda il periodo di tempo precedente la somministrazione del mezzo di contrasto, mentre la seconda é la fase di "*enhancement*", che evidenzia l'aumento di valori di concentrazione tissutale provocato dal mezzo di contrasto. Infine si ha l'"*outflow*", relativo al periodo di tempo in cui il tracciante fuoriesce dal tessuto in esame. Nell'immagine 2.2 sono riportate le immagini della stessa slice di tessuto nelle tre diverse fasi.



(a) Fase di baseline (b) Fase di enhancement (c) Fase di out flow

Figura 2.2: Slice di tessuto nelle diverse fasi dell'esame

Se sottraiamo, alla sequenza di immagini rappresentata nella TCC della prima figura, l'immagine di baseline, possiamo focalizzare il valore di enhancement. Tale valore può essere utilizzato per valutare la funzionalità professionale e consente di tracciare la TAC (*Time Attenuation Curve*), di cui l'immagine 2.3 ci offre un esempio.

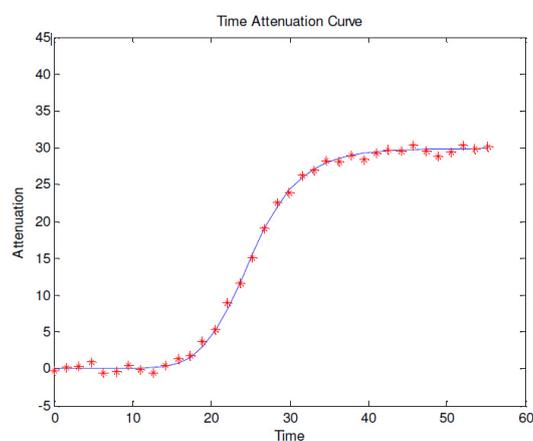


Figura 2.3: Una TAC, Time Attenuation Curve

2.2.2 Principii della tecnica

Per ottenere un buon enhancement tissutale serve un mezzo di contrasto iodato ad alta concentrazione (370-400mg/l), con tecniche di somministrazione

diverse, a seconda dei diversi tipi di tessuto da esaminare. Una volta entrato in circolo nel corpo, il tracciante inizia ad entrare ed uscire dai vari tessuti seguendo il flusso sanguigno. Il periodo di tempo compreso tra l'iniezione del mezzo di contrasto e l'istante di tempo che precede il secondo passaggio del tracciante all'interno di un determinato tessuto é denominato "*fase di first pass*". Il periodo di tempo successivo é la "*fase interstiziale*".

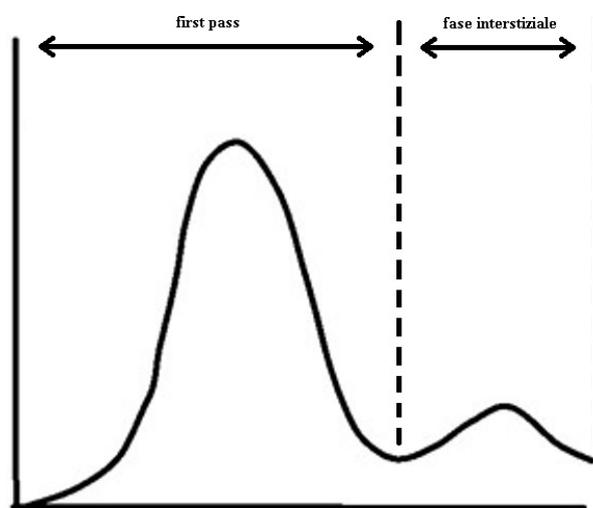


Figura 2.4: Andamento del tracciante

A causa dell'alta soluzione risoluzione temporale necessaria durante la fase di first pass e della necessità di dover ripetere la scansione dello stesso volume corporeo nel tempo, per misurare variazioni di enhancement, la TC dovrà essere effettuata tramite una scansione assiale con il lettino in posizione stazionaria. Tali condizioni pongono un limite alle dimensioni del volume che può essere studiato: lungo l'asse z l'altezza del volume analizzato varierá in base al numero di detettori presenti nella macchina TC usata per la scansione (ad esempio nella TC a 16 slice sará pari a 20mm, mentre nella TC a 64 slice sará pari a 40mm).

Le scansioni effettuate prima dell'iniezione del mezzo di contrasto hanno come unico scopo quello di individuare l'area da sottoporre ad analisi per-

fusione, quindi, non essendo necessario disporre di un'elevata risoluzione temporale, né di slice particolarmente sottili, è possibile acquisire le immagini utilizzando una dose bassa di radiazioni. Nel caso in cui un tumore abbia volume maggiore rispetto a quello che la scansione perfusione è in grado di analizzare, è necessario identificare la sezione nella quale l'area del tumore visualizzata è maggiore e utilizzare quella slice come centro volumetrico dello studio di perfusione.

Le scansioni effettuate durante il first pass, come si è già detto, richiedono un'elevata risoluzione temporale variabile in base al modello cinetico scelto: 1 s per il metodo di deconvoluzione e di 3-5 s per il modello cinetico bicomportamentale. In questa fase della TCp si raccolgono dati che servono per calcolare la maggior parte dei parametri perfusionali (BF, BV, MTT). Per ridurre il quantitativo di radiazioni a cui si sottopone il paziente, è preferibile mantenere livelli di voltaggio e di amperaggio piuttosto bassi. Inoltre per mantenere un buon equilibrio tra risoluzione spaziale e rapporto segnale rumore delle scansioni, si deve mantenere lo spessore delle slice superiore a 5mm.

Nella fase interstiziale della TCp, in cui vengono calcolati altri parametri come la permeabilità superficiale, la lunga durata della scansione rende alto il numero di radiazioni somministrate al paziente, ma tale dose può essere contenuta grazie alla risoluzione temporale delle scansioni, che può essere più bassa rispetto a quella necessaria durante il first pass.

Gli artefatti Al fine di ottenere risultati affidabili è necessario fare il possibile per evitare che si creino degli artefatti nelle immagini. Molto frequenti sono gli artefatti da movimento del paziente, che influiscono sia sull'esecuzione dell'esame che sulla sua valutazione clinica, ma possono comunque essere ridotti prendendo precauzioni durante la scansione. Ad esempio, per regioni anatomiche molto soggette ad artefatti, come il basso torace e la parte alta dell'addome (che tendono a muoversi per lo spostamento del diaframma durante la normale respirazione), si può usare una cintura addominale,

oltre a chiedere al paziente di trattenere il respiro durante le fasi piú delicate dell'esame. Anche i movimenti peristaltici e la deglutizione possono produrre artefatti da movimento, ma possono essere ridotti somministrando al paziente degli agenti inibitori del movimento. Infine, per tutti gli artefatti da movimento che non si possono evitare, esistono degli appositi software di correzione che rendono piú affidabile l'analisi TCp.

Inoltre esistono artefatti causati dalla macchina e anche quelli prodotti per effetto dei volumi parziali, che sono spesso presenti, per esempio, a livello dell'area di contatto tra polmone ed emi-diaframma, due strutture di densità diverse fra loro. L'unico modo per ridurre l'artefatto in questo caso é quello di ricorrere a slice piú sottili, che consentano di campionare meglio le varie strutture, ma bisogna prestare attenzione alla risoluzione dell'immagine, perché riducendo lo spessore della slice si ottiene la riduzione del rapporto segnale rumore.

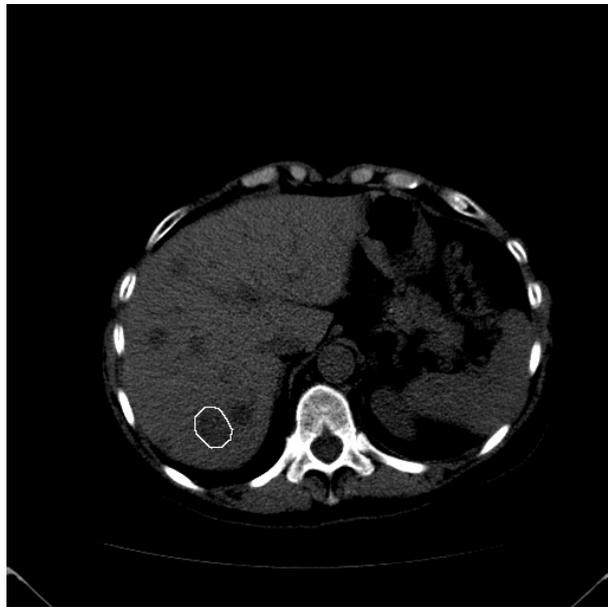
2.2.3 Il post-processing

Una volta effettuate le scansioni, si passa al calcolo dei parametri perfusionali, effettuato con appositi software semi-automatici. Infatti la parte di analisi matematica viene eseguita in maniera automatica, anche se richiede che un operatore posizioni le ROI opportune, che ottimizzino l'intero processo. Come mostra l'immagine 2.5, le ROI da posizionare sono due: una va messa all'interno di un vaso arterioso che nutra il tumore e l'altra in una zona tissutale all'interno del volume di scansione.

Dalla prima ROI é possibile estrarre l'input arterioso (un parametro che é specifico per ogni paziente e anche per ogni esame effettuato sullo stesso paziente a distanza di tempo), dato che serve per evitare artefatti da movimento e da volume parziale. La seconda ROI ci consente di ottenere il segnale a partire dal quale verranno calcolati i parametri di perfusione.



(a) Contorno della ROI arteriosa



(b) Contorno della ROI tissutale

Figura 2.5: ROI necessarie per definire il modello

2.2.4 Problemi legati alla mancanza di un protocollo d'acquisizione standard

Nonostante l'uso di macchine TC di ultima generazione, spesso nello studio perfusionale si ottengono risultati poco attendibili per la scarsa riproducibilità intra-paziente ed inter-operatore, soprattutto a causa della mancanza di standardizzazione nell'uso delle macchine stesse.

Ad esempio in letteratura non è ancora stato definito con precisione quale tracciante debba essere usato, quale debba essere la sua concentrazione o con quale velocità debba essere iniettato in vena. Anche per i valori di voltaggio, di amperaggio e di spessore delle slice vengono proposti dei range di valori da adottare, anziché dei valori ottimali precisi con cui impostare la macchina TC. Non esiste nemmeno una definizione standard di first pass: un errore nella sua identificazione potrebbe interferire sul calcolo dei parametri perfusionali e, sebbene alcuni software siano in grado di individuare la fine del first pass in modo automatico, i risultati non sempre sono completamente affidabili. Ulteriori problemi di standardizzazione riguardano il posizionamento delle ROI. Per esempio, non è stato ancora definito su quale lato del tumore (quello ipso-laterale o quello contro-laterale) sovrapporre la ROI tissutale, mentre per la ROI di input arterioso, è ancora incerta la scelta del vaso sanguigno su cui posizionarla: scegliendo un vaso di grandi dimensioni diminuisce la turbolenza del flusso sanguigno e di conseguenza diminuiscono gli artefatti da essa provocati, mentre scegliendo un vaso di dimensioni ridotte che nutra il tumore, si può ottenere una stima più realistica dell'apporto di sangue alla massa tumorale. Infine per la ROI di input arterioso rimane ancora incerta l'influenza della scelta di un vaso nella parte sinistra o destra del corpo, che potrebbe introdurre un certo grado di variabilità legato alle differenze anatomiche tra i due lati del corpo umano stesso.

Anche il posizionamento delle ROI nel tempo costituisce un ulteriore fattore di variabilità. Ad oggi infatti sono stati proposti tre metodi per seguire gli spostamenti delle ROI, relative alle lesioni, durante gli istanti della scan-

sione.

Nel metodo del *fixed track* si disegna una ROI sulla regione individuata in una delle immagini dell'esame, poi la si mantiene fissa per tutte le altre immagini senza modificarne la posizione. Questo metodo non tiene però in considerazione la possibilità che il tessuto in esame si sposti, per movimenti del paziente o movimenti fisiologici degli organi: in tal caso infatti il segnale, prelevato in diversi istanti di tempo da uno specifico pixel della ROI, non proviene dalla stessa porzione tissutale umana, come sarebbe necessario.

Il secondo metodo utilizza una ROI tissutale, tracciata manualmente dal radiologo, che viene traslata sulle immagini della sequenza *fixed track* e ciò consente di seguire meglio gli spostamenti del tessuto nel tempo.

Il terzo metodo, denominato *best track*, prevede ancora che il radiologo scelga una slice di riferimento su cui tracciare la ROI della lesione, ma questa viene poi applicata sulle slice più significative, una per ogni istante temporale della scansione, che non necessariamente appartengono alla sequenza standard del *fixed track*.

Ulteriore causa di non-standardizzazione intra-paziente è la vena scelta per l'iniezione del mezzo di contrasto: l'utilizzo della stessa vena può ridurre la variabilità degli studi longitudinali sul paziente, ma ci sono degli effetti collaterali perché, incannulando frequentemente la stessa vena, questa perde di elasticità e di resistenza alle forti pressioni e non è più adatta ad uno studio perfusionale.

2.2.5 Lo studio delle neoplasie tramite TCp

Attualmente l'utilizzo della TCp nello studio dei tumori non è stato ancora validato per tutte le neoplasie, per esempio del polmone e del pancreas. Per quanto riguarda il fegato, alcuni studi hanno dimostrato la capacità della TCp di monitorare l'effetto delle terapie anti-angiogenetiche/vascolari, ma non le lesioni secondarie epatiche in pazienti sottoposti a chemioterapia. Non sono ancora stati fatti studi in merito ad un eventuale ruolo della TCp nella

predizione della risposta alle terapie nei tumori dei reni, mentre nel caso delle neoplasie del retto ne é già stata dimostrata la capacità nel monitorare gli effetti delle terapie anti-angiogenetiche.

2.3 Il modello cinetico deconvoluzionale

Il metodo deconvoluzionale é uno dei modelli cinetici maggiormente utilizzati nello studio dei parametri perfusionali e si basa su un operatore di deconvoluzione matematica per confrontare un input arterioso con un input tissutale, ottenendo poi la funzione dell'impulso residuo, cioè la curva teorica che descrive la concentrazione, nel tempo, del mezzo di contrasto presente nel tessuto analizzato con la ROI. Da questa si estrapolano successivamente i parametri perfusionali, in primis il BF.

Chiaramente ha senso analizzare le scansioni relative alla fase di *first pass*: nel caso di TCp epatiche ciò corrisponde a considerare solo quelle effettuate nei primi 45/50 secondi successivi all'iniezione del mezzo di contrasto. In questa fase possiamo infatti assumere che il tracciante sia presente nel solo compartimento vascolare, per cui é possibile calcolare il flusso di sangue con una buona precisione.

2.3.1 Formulazione matematica

Il modello deconvoluzionale si basa sul *principio del volume centrale*, che considera le reti vascolari come piccoli volumi isolati, ciascuno dotato di un proprio input arterioso e di un output venoso, tale per cui tutto il mezzo introdotto viene espulso dalla rete, in tempi piú o meno brevi. Altro tassello fondamentale del metodo é la convoluzione, che modella come l'input arterioso $C_a(t)$ sia legato all'output venoso $C_v(t)$, grazie alla funzione "impulso" o "di trasporto" $h(t)$. Piú precisamente, si pone

$$C_v(t) = h(t) * C_a(t) = \int_0^t h(t - \tau) C_a(\tau) d\tau \quad (2.1)$$

intendendo proprio h come funzione densità di probabilità dei tempi di transito. L'operatore così definito è un operatore memoria, poiché si considera la curva tempo-concentrazione arteriosa come una sequenza di boli ideali, di mezzo di contrasto, scalati e ritardati nel tempo. Inoltre, così facendo, è possibile vedere la risposta del sistema come somma di una successione di funzioni di risposta, a loro volta scalate e ritardate. In realtà, all'interno della rete vascolare, le particelle di ogni bolo seguiranno diversi percorsi e quindi i loro tempi di transito dipenderanno anche dalla struttura vascolare stessa: ne segue che la funzione $h(t)$ deve dipendere sia dalle caratteristiche anatomiche della rete locale, sia dal comportamento del tracciante.

Mentre la funzione h si riferisce all'input arterioso che poi viene espulso, noi siamo maggiormente interessati a vedere il processo da un'altra ottica, cioè quella dell'eliminazione dello stesso mezzo di contrasto, che viene direttamente descritto dalla funzione residua $R(t)$ ed espressa formalmente come

$$R(t) = 1 - \int_0^t h(\tau) d\tau \quad (2.2)$$

Poiché h è una densità di probabilità, $R(t)$ è una funzione non-negativa e decrescente nel tempo, con valori che variano tra l'uno e lo zero.

In particolare, si pone $R(t_0) = 1$ per l'istante in cui il mezzo di contrasto viene iniettato, perché si suppone che tutto il tracciante sia all'interno della rete vascolare e che l'iniezione sia istantanea. Invece, dopo un certo tempo t_N , tutto il mezzo sarà uscito dalla rete e quindi $R(t_N) = 0$, come idealmente per tutti i tempi a seguire. In realtà, si verifica il fenomeno del ricircolo all'interno della rete vascolare, quindi il residuo non sarà più nullo; ad ogni modo il modello è valido per la fase di first-pass.

Ora vediamo come vengono definiti i parametri perfusionali, dal punto di vista matematico.

Il tempo di transito medio MTT è modellizzato a partire dalla funzione di

trasporto come

$$MTT = \frac{\int_{-\infty}^{\infty} \tau h(\tau) d\tau}{\int_{-\infty}^{\infty} h(\tau) d\tau} \quad , \quad (2.3)$$

mentre il volume sanguigno si definisce considerando sia l'input arterioso che l'output venoso, relativi ad un determinato volume tissutale, ossia con

$$BV = \frac{\int_{-\infty}^{\infty} C(\tau) d\tau}{\int_{-\infty}^{\infty} C_a(\tau) d\tau} \quad . \quad (2.4)$$

Sfruttando il principio del volume centrale, infine, si esprime il flusso sanguigno mediante i primi due parametri, come

$$BF = \frac{BV}{MMT} \quad (2.5)$$

Dalle immagini di un esame TCp, si ricavano le TAC $C_a(t)$ di input arterioso e le TAC di output tissutale di un volume di interesse $C(t)$. Poiché la funzione impulso non é nota, non é possibile calcolare il parametro MTT in maniera diretta, sfruttando l'equazione 2.3 : é quindi necessario riformulare la TAC $C(t)$ in funzione del parametro BF. Sfruttando le equazioni 2.1 e 2.2, si arriva alla formulazione

$$C(t) = BF C_a(t) * R(t) = BF \int_0^t C_a(\tau) R(t - \tau) d\tau \quad (2.6)$$

Ora, dato che R é definita tramite h , non si conosce neppure la vera funzione residuo a priori: per ricavare BF si deve pertanto effettuare una deconvoluzione, per risolvere il sistema inverso che deriva dall'equazione 2.6. Il metodo piú implementato é il *modello deconvoluzionale non parametrico*, che prevede di calcolare la quantità $BF \cdot R(t)$, per poi ricavare il parametro perfuzionale sfruttando l'equazione 2.2: poiché nell'istante t_0 il residuo ha valore unitario, risulta infatti

$$BF = BF \cdot R(t_0) \quad (2.7)$$

In altri termini, dal risultato della deconvoluzione ricavo il valore di BF e la vera forma della funzione residua, da cui posso poi calcolare la funzione h e i parametri mancanti, con le formule sopra elencate.

Esistono anche modelli parametrici di deconvoluzione, ma questi necessitano di ipotesi aggiuntive che, per quanto realistiche, limitano il modello con le loro supposizioni. Per questo motivo si preferisce il modello non parametrico.

2.3.2 Formulazione matematica discreta

Approfondiamo ora l'aspetto deconvoluzionale, esplicitandone gli aspetti algebrici, col fine di ottenere una formulazione matriciale del problema.

Iniziamo campionando la variabile temporale: siano t_1, t_2, \dots, t_N gli istanti a cui si eseguono le scansioni, equidistanziati col passo di campionamento Δt .

Si discretizza la convoluzione dell'equazione 2.6 come

$$C(t_i) = BF \cdot \int_0^{t_i} C_a(\tau)R(t_i - \tau)d\tau \approx BF \cdot \Delta t \cdot \sum_{j=1}^i C_a(t_j)R(t_i - t_j) \quad (2.8)$$

e si definiscono i vettori

$$C = \begin{pmatrix} C(t_1) \\ C(t_2) \\ \vdots \\ C(t_N) \end{pmatrix}, \quad R = \begin{pmatrix} R(t_1) \\ R(t_2) \\ \vdots \\ R(t_N) \end{pmatrix} \quad (2.9)$$

e la matrice

$$C_a = \begin{pmatrix} C_a(t_1) & 0 & \cdots & 0 \\ C_a(t_2) & C_a(t_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ C(t_N) & C_a(t_{N-1}) & \cdots & C_a(t_1) \end{pmatrix} \quad (2.10)$$

Così, l'equazione matriciale che descrive il processo della TCp é

$$C = BF \Delta t C_a \cdot R \quad (2.11)$$

da cui si vede che per ricavare $BF R$ é necessario calcolare la deconvoluzione tra la matrice C_a di kernel e il vettore noto C , moltiplicando per un fattore scalare pari a $\frac{1}{\Delta t}$.

Avendo discretizzato un modello continuo, si sono inevitabilmente commessi degli errori. In letteratura, tuttavia, é stato proposto un metodo per limitarne gli effetti: é sufficiente infatti riscrivere la matrice C_a con gli elementi

$$\alpha_{i,j} = \begin{cases} C_a(t_{i-j-1}) + 4C_a(t_{i-j}) + C_a(t_{i-j+1}), & \text{se } 1 \leq j \leq i \\ 0, & \text{altrove} \end{cases}$$

Capitolo 3

Applicazione al modello simulato

Prima di affrontare il caso reale di deconvoluzione applicata alla TCp, si testano i metodi implementati su un problema simulato. Tale problema si sviluppa a partire dal modello e dalle formule esposte nel precedente capitolo e viene poi risolto applicando i metodi matematici di risoluzione trattati nel primo capitolo.

L'obiettivo di questo capitolo é strettamente quella di verificare e confrontare tra loro i risultati forniti dai metodi di regolarizzazione, della TSVD e di Tikhonov, al variare del parametro scelto, con i metodi della L-curva o della GCV, per la ricostruzione del parametro perfusionale BF. Il confronto si esegue tenendo conto del valore assegnato inizialmente a tale parametro e utilizzato poi per la costruzione del sistema di simulazione. Inoltre, si testano anche due metodi suggeriti in letteratura per la scelta del parametro di regolarizzazione nel caso specifico della ricostruzione su tomografie.

Nei programmi utilizzati per gli output riportati in seguito, si é fatto uso del pacchetto *regu*, un package Matlab per l'analisi e la risoluzione di problemi discreti mal posti, sviluppato dal professore Per Christian Hansen e scaricabile dal sito <http://www2.imm.dtu.dk/~pcha/Regutools/>. Per una

descrizione dettagliata delle function, si rimanda a [3] e al manuale scaricabile da www.mathworks.com/matlabcentral/fileexchange.

Per questa tesi, sono state ampiamente sfruttate le funzioni `picard` e `svd`, per l'analisi preliminare dei problemi, `tsvd`, `tikhonov` e `gcv` per la risoluzione dei problemi stessi, mentre con `corner` si é cercato il vertice di ogni L-curva, sfruttandone l'algoritmo di pruning adattivo.

3.1 Creazione del problema test

Idealmente viene fissato un voxel e si studia il problema ad esso relativo. Si inizia attribuendo dei valori ai parametri che descrivono il problema. Ad esempio fissano i parametri temporali come segue

```
t0=20;
dt=1;
sec_time_global=[0:dt:199];
```

quindi considereremo duecento istanti temporali, equidistanti una unitá temporale, e nel tempo `t0` si ha l'iniezione del mezzo di contrasto.

La curva C_a viene verosimilmente generata con il modello della funzione gamma variata, perché meglio simula il comportamento della TAC arteriosa per uno schema di iniezione standard. Assume pertanto l'espressione

$$C_a(t_i) = \begin{cases} 0 & t_i \leq t_0 \\ a \cdot (t_i - t_0)^b \cdot e^{-\frac{(t_i - t_0)}{c}} & t_i > t_0 \end{cases}$$

dove vengono prefissti i valori $a=1$; $b=3$; e $c=1.5$;

Il risultato é mostrato nella figura 3.1.

Vengono anche settati i parametri perfusionali

```
MTT=9;
BF=1;
```

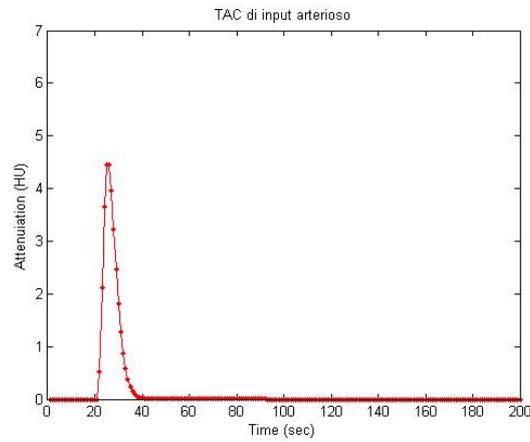


Figura 3.1: Funzione di input arterioso

e successivamente é costruita la funzione residuo, con il modello esponenziale, ossia come

$$R(t_i) = \exp\left(-\frac{t_i}{MTT}\right).$$

In piena concordanza con l'equazione 2.2, la funzione residuo ha l'andamento mostrato nella figura 3.2. A questo punto, sfruttando l'approssimazione

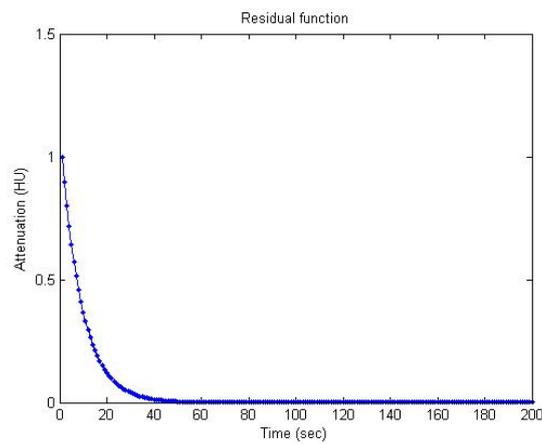
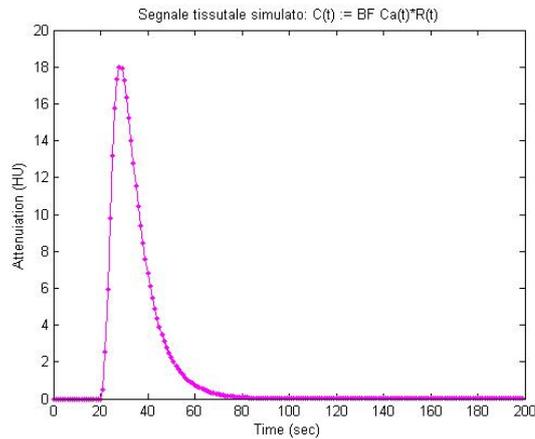
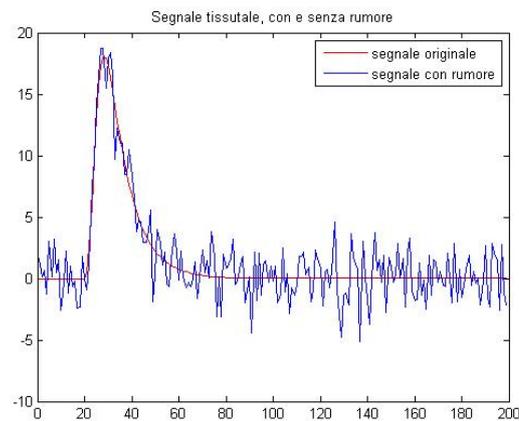


Figura 3.2: Funzione residua teorica

riportata nell'equazione 2.6, viene calcolata la convoluzione dei due segnali e si ottiene la funzione di segnale tissutale, illustrata nella figura 3.3(a).



(a) Funzione di segnale tissutale

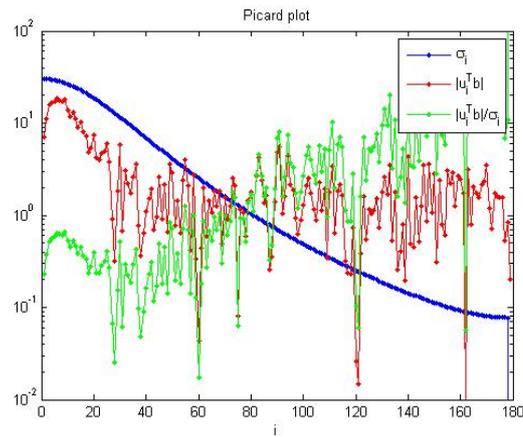


(b) Funzione di segnale tissutale con rumore

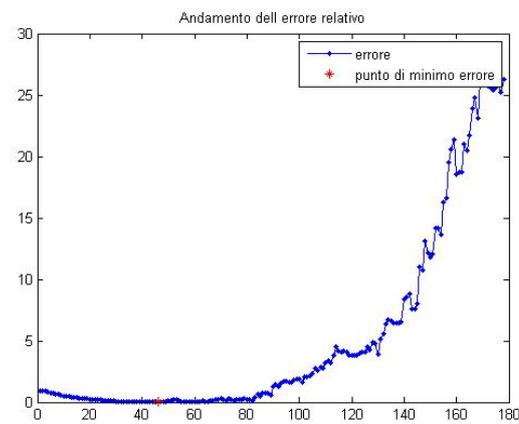
Figura 3.3: Funzioni di segnale tissutale, pulito e con rumore aggiunto

Per rendere il modello piú veritiero, al segnale tissutale é stato aggiunto del rumore, generato come rumore gaussiano di media entitá, in cui il SNR é settato a 10. Il disturbo cosí provocato é osservabile dalla curva blu dell'immagine 3.3(b).

I grafici riportati nella figura 3.4 ci permettono di analizzare il sistema, prima di risolverlo: l'immagine (a) grafica la condizione di Picard vista nel primo capitolo, mentre la (b) mostra l'andamento dell'errore relativo, compiuto dal calcolo di BF mediante la SVD, ad ogni passo iterativo. Considerando l'irregolarità della curva rossa, ci aspettiamo una regolarizzazione molto forte che, per la TSVD, corrisponde a un valore di k almeno inferiore all'80. La curva (b), d'altra parte, poiché ha minimo per $k=46$, suggerisce una regolarizzazione ancora più incisiva.



(a) Grafico di Picard, per SNR=10



(b) Andamento degli errori della TSVD

Figura 3.4: Grafici per analizzare preventivamente il sistema da risolvere

3.2 Analisi degli output BF

Abbiamo testato i metodi di regolarizzazione della TSVD e di Tikhonov, usando la L-curva o il GCV come algoritmo di scelta del parametro regolarizzante. Per la precisione, come operatore L per il metodo di Tikhonov sono state utilizzate la matrice identità e le matrici di approssimazione delle derivate prima e seconda, alle differenze finite centrali, di seguito riportare:

$$\nabla = \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \ddots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 & -2 \end{pmatrix}, \quad \nabla^2 = \begin{pmatrix} 4 & -1 & 0 & \cdots & 0 \\ -1 & 4 & -1 & \ddots & 0 \\ 0 & -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 4 \end{pmatrix} \quad (3.1)$$

Inoltre, per la scelta del parametro k per la TSVD, abbiamo anche considerato due varianti suggerite da [11] che consistono nel prendere l'indice tale per cui si conservino informazioni fino al 4% o 20% del primo valore singolare σ_1 : indicheremo tali metodi rispettivamente con `perc4` e `perc20`. Come si può capire anche dalla figura 3.5, con questa tecnica il parametro della regolarizzazione dipende esclusivamente dalle caratteristiche della matrice del sistema, mentre il termine noto non ne influenza affatto la decisione.

Ora, poiché BF è settato pari a 1, la curva dell'immagine 3.2 corrisponde alla curva $BF \cdot R(t_i)$ che si desidera ricostruire, supponendo note solo l'input arterioso e l'output tissutale, con la deconvoluzione applicata alla formula 2.11. Sarà pertanto la curva di riferimento, per valutazioni successive.

Si testano ora le otto varianti dei classici metodi risolutivi, più le quattro alternative per la L-curva.

Per evitare osservazioni troppo specifiche su un singolo caso, tutti i metodi sono stati eseguiti diverse volte, in ognuna delle quali il rumore aggiunto all'output tissutale è stato creato usando il comando `randn('seed', 0)`; e un "seed" differente. In questo modo il rumore additivo è sempre di tipo gaussiano, ma consiste in una sequenza di numeri pseudo-random generati

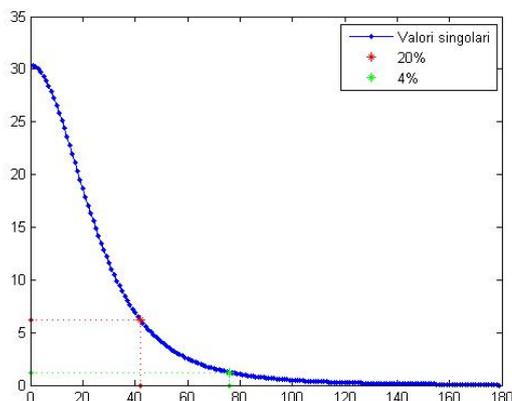


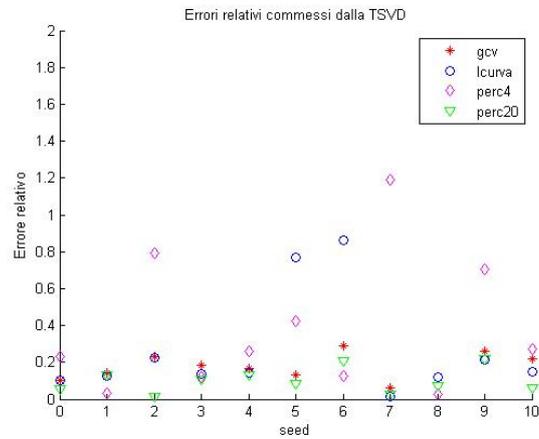
Figura 3.5: Grafico dei valori singolari della matrice, con evidenziati quelli relativi al 20% e 4% del valore singolare piú grande

a partire da valori variabili, i cosiddetti *seed*. Questo porta a trattare con segnali tissutali mai uguali tra loro, ma le cui differenze sono limitate e quindi gli output sono ben paragonabili.

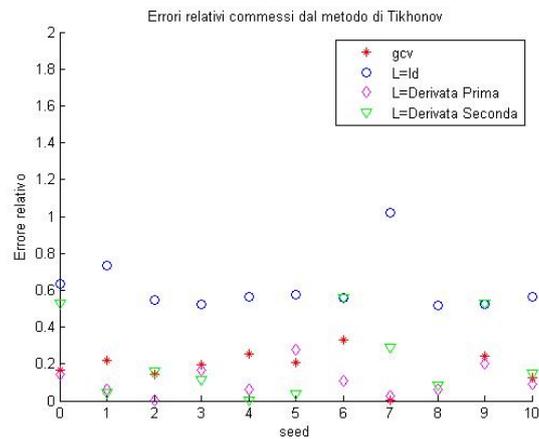
Eseguendo i programmi, si osserva che in nessun caso viene restituito il valore esatto $BF=1$. Valutiamo allora l'efficienza dei metodi in base a quanto le loro ricostruzioni si sono avvicinate al valore sperato, definendo l'errore commesso come norma euclidea della distanza tra l'1 e l'output. Le due immagini della figura 3.6 riportano gli errori relativi commessi dai due metodi di risoluzione, al variare del *seed*.

Confrontando i risultati, si evince che l'algoritmo migliore é quello di Tikhonov con la derivata prima come operatore e il parametro scelto dalla L-curva: si é calcolato infatti che commette un errore medio di $1.01 \cdot 10^{-1}$. Per precisione, nei metodi di Tikhonov vengono sempre testati i parametri λ contenuti nel vettore `lambda=[0.1:0.1:0.9, 1:50]`: tali parametri generano (con "*seed* = 0") la L-curva riportata in figura 3.7 che, come si può vedere, ha una buona forma a "*L*". Inoltre il vertice viene correttamente localizzato.

Guardando al grafico (a) si capisce che anche i metodi basati sulle percentuali fisse di informazioni sono affidabili: soprattutto con l'opzione "*perc20*"



(a) Errori ottenuti con la TSVD



(b) Errori ottenuti con Tikhonov

Figura 3.6: Errori relativi per SNR=10 e seed variabile

la soluzione é molto buona e, osservandone anche una corrispondente curva del residuo, si nota che questa é parecchio regolarizzata, a differenza della curva data da "perc4" dove il rumore comincia già a propagarsi eccessivamente (si veda la figura 3.8). Tenendo presente che $k \in \{1, 2, \dots, 179\}$ (poiché $n = 179$ é la dimensione del problema), si osserva che i parametri di regolarizzazione individuati da "perc4" e "perc20" sono rispettivamente $k=76$ e $k=42$, indipendentemente dal seed del rumore: le proprietà delle curve riportate nella figura 3.8 sono in accordo con quanto osservato grazie al grafico di

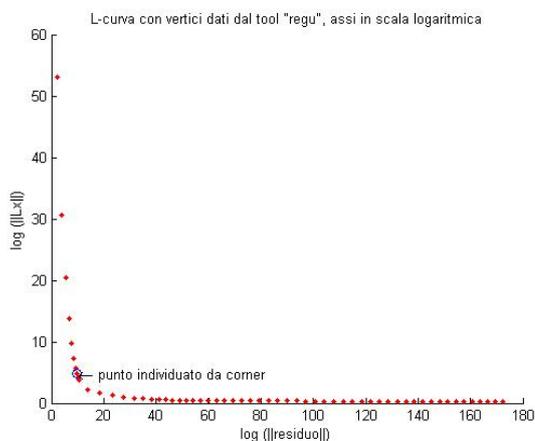


Figura 3.7: L-curva ottenuta per il metodo di Tikhonov con operatore di derivata prima

Picard.

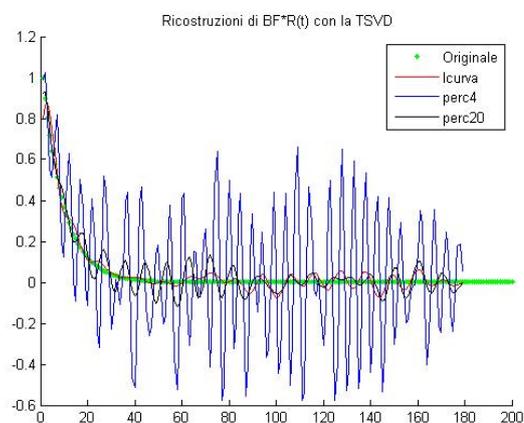
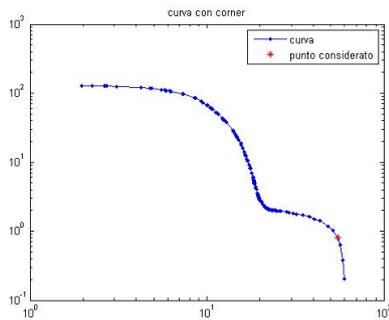


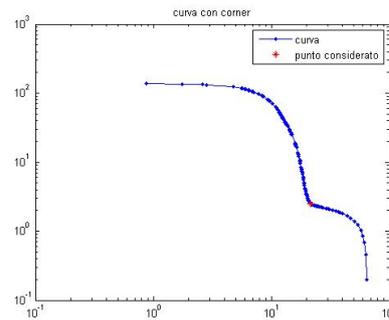
Figura 3.8: Ricostruzioni ottenute con la TSVD, della funzione residuo

Analizzando ancora la figura 3.6, si vede che la TSVD con la L-curva non é particolarmente efficiente solo in alcuni casi (mentre negli altri é molto valida). Plottando alcune L-curva, si osserva che queste volte a sbagliare é la function che cerca il vertice della curva: quando questo non viene correttamente localizzato, infatti, viene utilizzato un k troppo piccolo, che rende

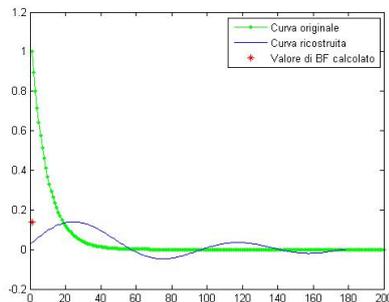
la soluzione eccessivamente regolare e, di conseguenza, un valore di BF distante dal valore atteso perché piú basso. Le immagini (a) e (c) della figura 3.9 sono relative all'esecuzione con "*seed* = 6" e mostrano quanto appena osservato: il punto considerato é distante dal vertice e corrisponde a $k=4$, quindi la curva residuo calcolata in questo caso é talmente smoothed che restituisce $BF=0.1388$. Al contrario, quando il corner viene ben individuato, come nell'immagine (b) che determina $k=36$, la curva soluzione esegue un miglior fitting la prima parte della vera curva residua e la propagazione del rumore rovina solo l'ultimo tratto, con oscillazioni tollerabili: il parametro perfusionale calcolato in questo caso é $BF=0.9829$ che ben si avvicina al valore unitario atteso.



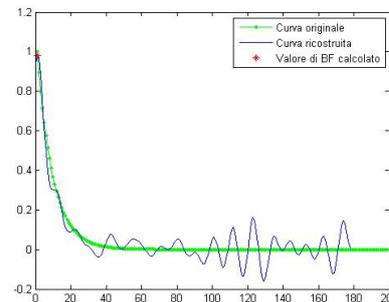
(a) L-curva con vertice errato



(b) L-curva con vertice corretto



(c) Ricostruzione del residuo ottenuta col vertice errato



(d) Ricostruzione del residuo ottenuta col vertice corretto

Figura 3.9: Esempi di L-curves con diversi vertici trovati e relative soluzioni

Dall'immagine (a) della figura 3.9 si osserva anche che la forma della "L" non é infatti delle migliori: l'andamento dei punti presenta un lungo tratto orizzontale che si abbassa lentamente fino a formare il tratto verticale della "L" e nel vertice si concentrano moltissimi punti. Al contrario, nella curva (b) la forma a "L" é piú marcata.

Vengono allora testate le varianti della L-curva, spiegate nel primo capitolo, nella speranza di ottenere curve e parametri di regolarizzazione migliori. La figura 3.10 riporta quanto ottenuto: in alto abbiamo le curve generate dalle coppie di punti $(\mathbf{k}, \|\mathbf{res}\|_k)$ e $(\text{cond}(\mathbf{A}_k), \|\mathbf{res}\|_k)$, mentre sotto sono graficate le funzioni di Reginska, definite nelle equazioni 1.24 e 1.25. Successivamente, la figura 3.11 mostra gli errori relativi dei valori ottenuti per BF, al variare dei seed.

Da quest'ultimo grafico, si nota che per questo problema il metodo di Reginska non é affidabile, in nessuna delle due formulazioni, perché restituisce un parametro perfusionale sempre distante dal valore sperato. Osservando anche l'andamento delle due funzioni, si capisce che il minimo viene spesso assunto per $\mathbf{k}=1$: un parametro di regolarizzazione cosí basso spiega l'entità dell'errore.

Inoltre, anche le curve coi residui non sono generalmente preferibili alla classica L-curva, perché commettono errori relativi sempre compresi tra lo 0.3 e lo 0.7, mentre la L-curva ha spesso errori inferiori allo 0.2. D'altra parte, quando quest'ultima commette errori piú elevati, la costanza degli errori compiuti dalle curve residue le rende migliori della L-curva stessa.

Per approfondire l'analisi dei metodi utilizzati, guardiamo anche come variano gli output per rumori piú o meno intensi. Innanzitutto, osserviamo l'evoluzione della L-curva al crescere del rumore, dalla figura 3.12.

Quando il SNR é alto, la curva ha una discreta forma a "L": man mano che il rapporto diminuisce, il vertice si sposta verso destra, indicando che é necessaria una maggior regolarizzazione, coerentemente col fatto che il ru-

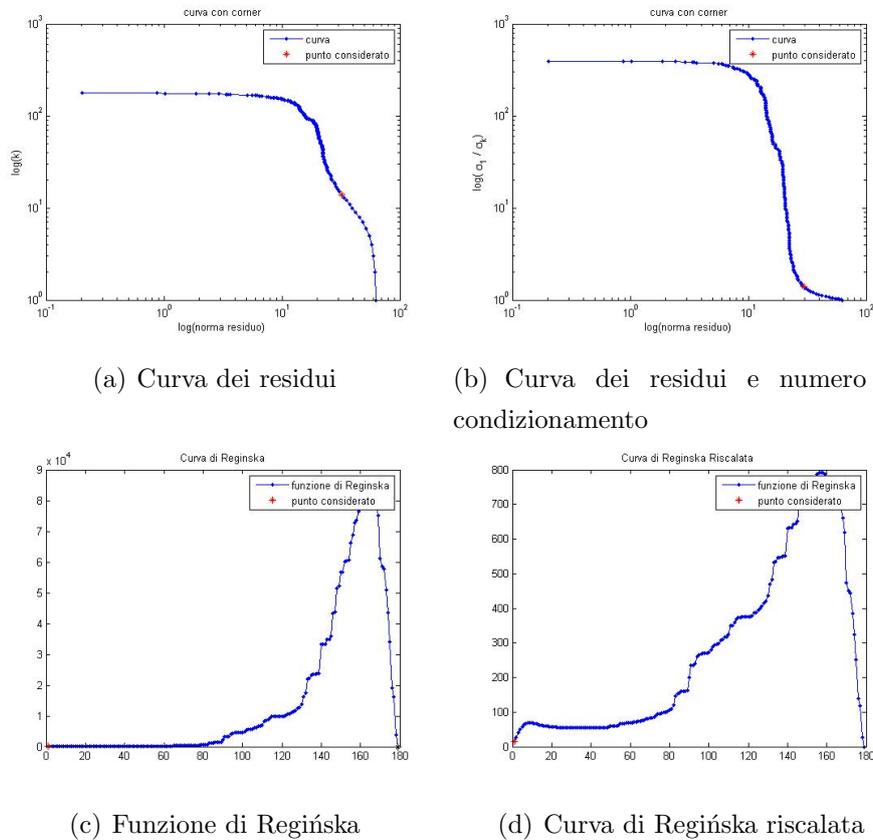


Figura 3.10: Varianti della L-curva, per determinare un buon parametro di regolarizzazione

more è più intenso. Quando però il SNR scende sotto il 10, la "L" è talmente smussata che il vertice forma una sorta di "gradino" e il tratto verticale, meno regolare, mostra lievissime convessità, in cui potrebbe cadere l'algoritmo per la localizzazione del vertice. Infine, per $\text{SNR}=2,5$ la caratteristica forma a "L" sparisce completamente: ciò nonostante, nel caso riportato nella figura (d), il punto evidenziato è verosimilmente buono.

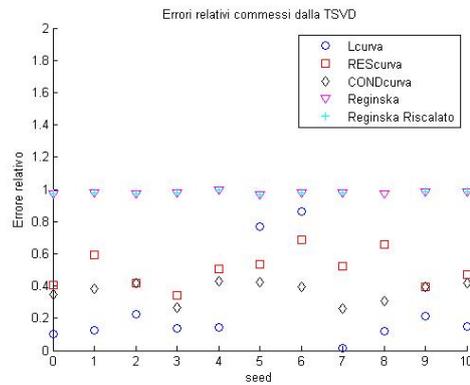


Figura 3.11: Errori relativi commessi dalla TSVD con parametro scelto con diverse curve e funzioni

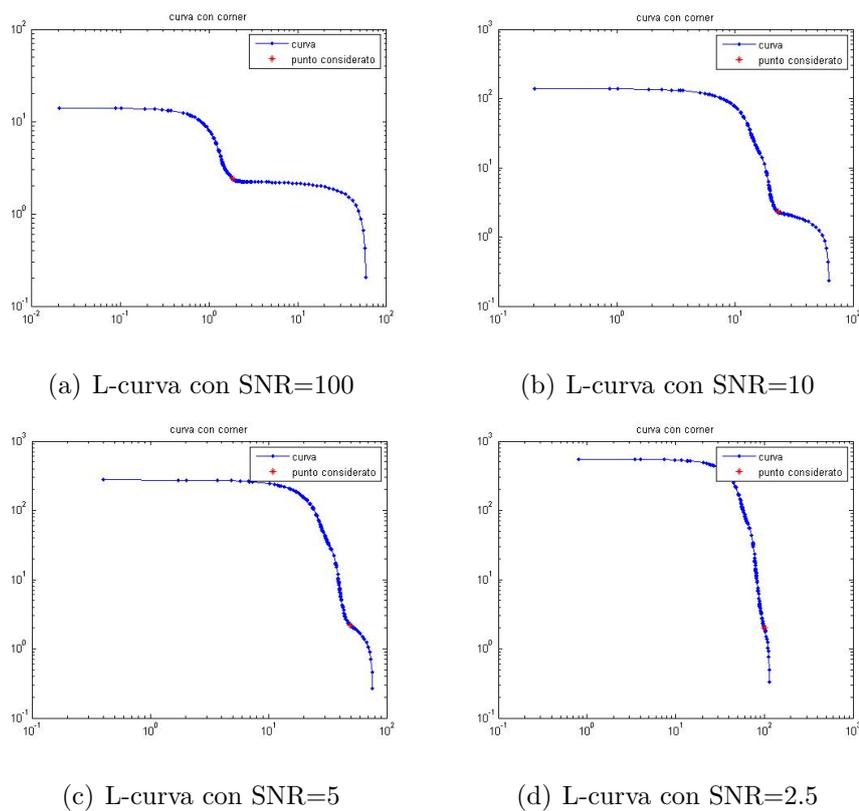


Figura 3.12: L-curve della TSVD al crescere del rumore

Vediamo ora le tabelle 3.1 e 3.2, dove sono riportati gli errori relativi commessi sul calcolo di BF dai metodi della TSVD e di Tikhonov, al variare del SNR. Nei casi con limitato rumore, i metodi basati sulla TSVD sono sempre molto validi, perché commettono errori dell'ordine di 10^{-2} mentre alcuni metodi con Tikhonov fanno già errori attorno al $5 \cdot 10^{-1}$. Solo l'algoritmo che unisce Tikhonov alla GCV sbaglia molto di più. Ad ogni modo, il valore di perfusione calcolato non è mai esatto.

Quando il rumore è maggiore, invece, i migliori risultati si ottengono con Tikhonov, sfruttando $L=Id$ e $L=\nabla^2$ come operatori: nonostante il SNR abbia valori molto bassi, gli errori registrati si aggirano attorno al $3 \cdot 10^{-1}$ e sono addirittura inferiori ai casi con minor rumore.

In generale, si nota la presenza di una "soglia critica" attorno al $SNR=5$, perché sotto tale valore i metodi rispondono in maniera diversa rispetto ai casi con rumore meno intenso.

Dal punto di vista strettamente computazionale, è opportuno guardare anche ai tempi di esecuzione di ciascun metodo, soprattutto in previsione del "caso reale", dove la ROI applicata alle TCp ci porterà a risolvere qualche centinaia di sistemi come questo. In realtà, i sistemi reali avranno dimensione talmente ridotta che il peso computazionale dei calcoli sarà complessivamente limitato. Ad ogni modo, già in questo caso con un solo sistema di dimensione 200×200 , le diverse velocità di calcolo sono significative: si veda a tal proposito la tabella 3.3, che riporta i secondi impiegati dal calcolatore per ciascuna deconvoluzione. Come ci si poteva aspettare, i metodi di Tikhonov che richiedono la L-curva sono i più lenti: devono infatti risolvere un sistema per ogni valore preimpostato di λ (in questo caso, quindi, 157 sistemi) e ciascuna risoluzione è più onerosa di una TSVD. Infatti, sebbene la TSVD necessiti di 200 prove per poter graficare la L-curva, ogni risoluzione è più veloce perché consiste in un solo prodotto matrice per matrice.

I metodi che non richiedono la L-curva sono ovviamente più veloci. Addirittura le TSVD con "perc4" e "perc20" richiedono una sola decomposizione ai valori singolari della matrice del sistema e decretano il parametro di regolar-

izzazione osservando semplicemente l'andamento dei σ_i : questo si concretizza in un tempo di esecuzione pari a un quarto del tempo impiegato dalla TSVD con la L-curva e, addirittura, pari a un trentacinquesimo del tempo servito dal metodo di Tikhonov piú rapido (che sfrutta comunque la L-curva).

SNR	TSVD			
	gcv	lcurva	perc4	perc20
100	0,0430	0,0401	0,0412	0,0849
50	0,1086	0,0353	0,0301	0,0817
10	0,1032	0,1032	0,2283	0,0556
5	0,1366	0,2597	0,7994	0,0231
2.5	0,4198	0,4023	2,5993	0,0420
2	0,5601	0,4548	3,4993	0,0796
1.5	0,5505	1,9736	4,9992	0,1867

Tabella 3.1: Errori relativi compiuti sul calcolo di BF, al variare del SNR, con la TSVD

SNR	TIKHONOV			
	gcv	$L=Id$	$L=\nabla$	$L=\nabla^2$
100	1,6997	0,5598	0,0622	0,5299
50	4,3995	0,5595	0,0618	0,5296
10	0,1487	0,5570	0,0262	0,5268
5	0,2333	0,5207	0,0739	0,5233
2.5	0,4328	0,4355	7,2862	0,4444
2	0,5490	0,4268	9,3702	0,3515
1.5	0,6312	0,3647	12,827	0,3346

Tabella 3.2: Errori relativi compiuti sul calcolo di BF, al variare del SNR, con Tikhonov

TSVD				TIKHONOV			
lcurva	perc4	perc20	gcv	$L=Id$	$L=\nabla$	$L=\nabla^2$	gcv
0.1966	0.0455	0.0444	0.1823	1.6131	1.7942	2.0685	0.3671

Tabella 3.3: Tempo necessario al calcolatore per eseguire ciascun metodo, nella simulazione

Capitolo 4

Applicazione al caso reale

I metodi di risoluzione matematica, visti nel primo capitolo e testati in quello precedente, vengono ora applicati al problema reale presentato nel capitolo 2. Le immagini tomografiche utilizzate sono state fornite dall'Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori, IRST, con cui il centro ARCES di Bologna collabora da diversi anni.

Il nostro obiettivo é calcolare il parametro perfusionale BF di una massa tumorale epatica e rappresentarne la mappa colorimetrica, da mostrare successivamente ai medici radiologi dell'istituto.

ISTITUT
SCIENTIFIC
ROMAGNOL
PER LO STUDI E LA CURA
DEI TUMORI



4.1 Definizione del problema

La figura 4.1 riporta alcune slide della sequenza *best track* dello stesso esame TCp. In particolare, l'immagine (a) é la prima dell'intera scansione e riporta la situazione a riposo, la (b) é relativa al quindicesimo istante temporale e si nota che il mezzo di contrasto iniettato sta circolando nei vasi del paziente, mentre la (c) corrisponde alla fase finale dell'esame e il tracciante é in ricircolo nel fegato. In totale, la sequenza é formata da 36 immagini perfusionali.

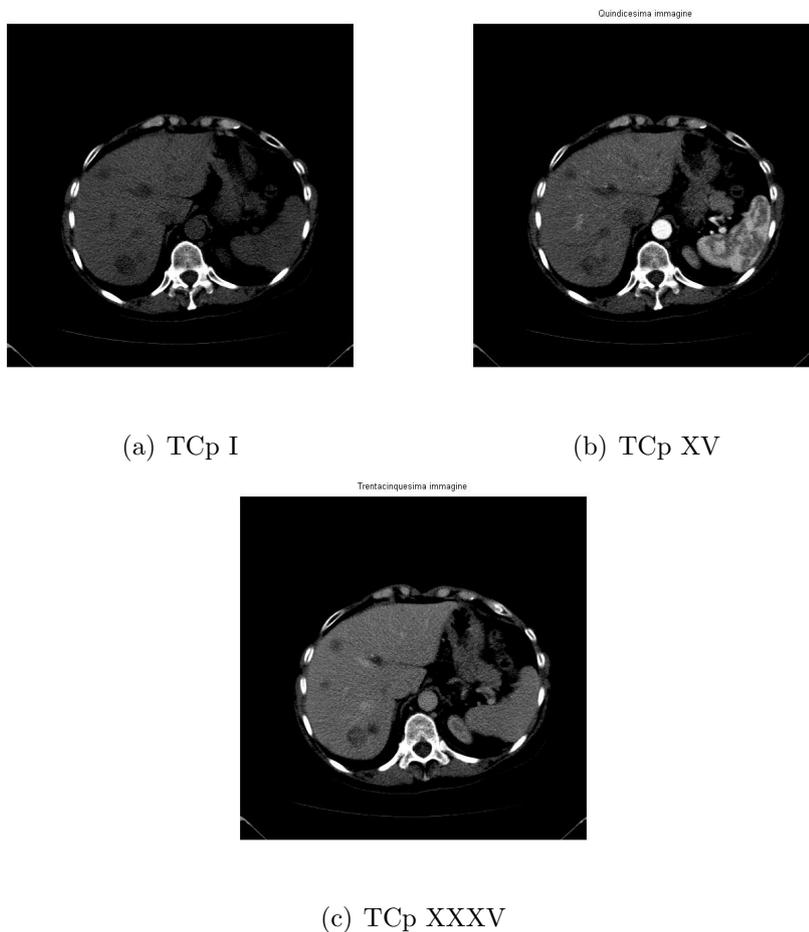


Figura 4.1: Alcune scansioni di TCp epatiche, relative allo stesso esame

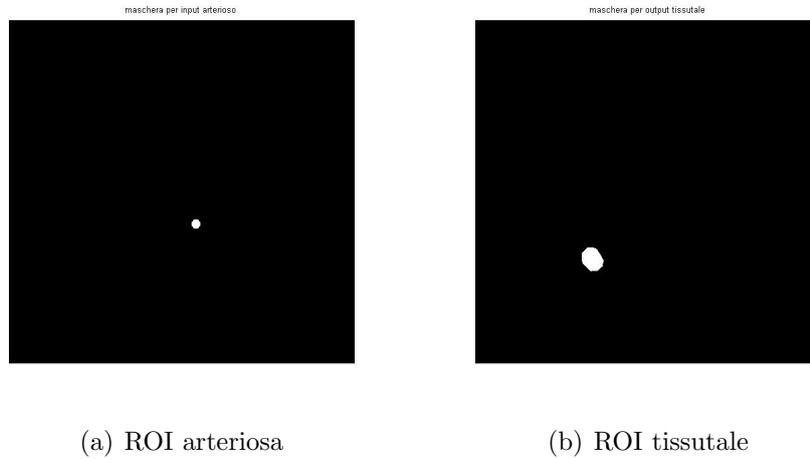
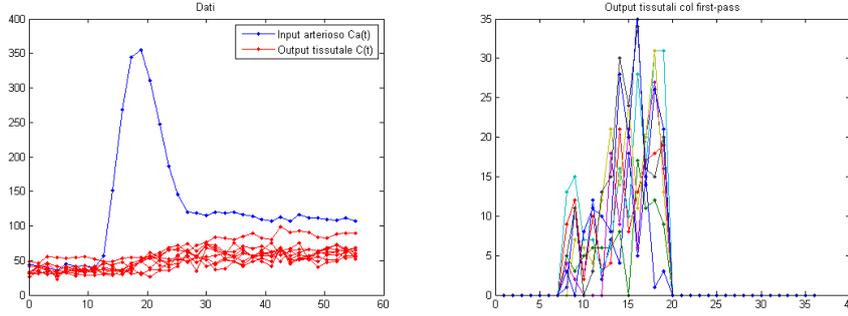


Figura 4.2: ROI utilizzate per ricavare i dati

Vengono poi utilizzate le ROI mostrate nella figura 4.2, per ricavare dalle TCp i valori di input. Con la ROI (a), si determina un unico vettore C_a di input arterioso, 36-dimensionale, con un'operazione di media fra tutti i pixel della ROI stessa. Al contrario, da ognuna delle 36 immagini, viene estrapolato il segnale relativo ad ogni pixel della ROI tissutale (b), quindi si ottengono 864 curve temporali C_i , $\forall i = 1, \dots, 864$, che costituiscono il termine noto di altrettanti sistemi che dovranno essere risolti e regolarizzati. La figura 4.3(a) mostra in blu l'input arterioso e in rosso alcuni di questi output tissutali. In verità, come abbiamo già spiegato nel secondo capitolo, è necessario individuare la fase di *first-pass* del mezzo di contrasto, per evitare di considerare quegli istanti di tempo in cui il tracciante ricircola nei vasi sanguigni e le TCp ci forniscono dati non attendibili per il calcolo dei parametri perfusionali. In questo caso, si è calcolato che l'iniezione del tracciante è avvenuta durante l'ottava scansione, mentre il ricircolo ha avuto inizio dopo la diciannovesima: il *first-pass* dura pertanto solo 12 scansioni. Di conseguenza, in tutti i vettori tissutali vengono azzerati i termini che non corrispondono a questa fase e l'immagine 4.3(b) riporta alcuni output così elaborati.

Anche i sistemi che andremo a risolvere devono essere ridimensionati,



(a) Curve ottenute di input arterioso e (b) Output tissutali della sola fase di di output tissutale first pass

Figura 4.3: Dati del problema

considerando i soli dati (ossia gli elementi vettoriali) compresi tra gli ottavi e i diciannovesimi termini: questo significa però definire problemi mal-posti di dimensione $\tilde{N} = 12$. Lavorare con sistemi a dimensione cosá bassa non é conveniente, soprattutto pensando al ruolo che riveste n nella scelta del parametro di regolarizzazione per la TSVD. Per questo motivo e per trattare in maniera piú robusta l'eventuale ritardo che si registra tra l'input e gli output, [11] suggerisce di implementare per i sistemi una formulazione matriciale lievemente diversa da quella mostrata nelle equazioni 2.9 e 2.10.

Posti

$$q = 864$$

come numero di punti della maschera tissutale, nonché numero di sistemi da risolvere, e

- $G^i = (G_1^i, G_2^i, \dots, G_{\tilde{N}}^i) = (C_i(t_8), C_i(t_9), \dots, C_i(t_{19})) \quad \forall i = 1, \dots, q$
- $H = (H_1, H_2, \dots, H_{\tilde{N}}) = (C_a(t_8), C_a(t_9), \dots, C_a(t_{19}))$

come vettori di soli dati relativi al *first-pass*, ora si vuole passare dalla dimensione \tilde{N} alla dimensione $n = 2\tilde{N} - 1$, con lo zero-padding di tutti i vettori G^i , arrivando a una formulazione a blocchi circolanti. Si definiscono pertanto i

vettori

$$b_i = \begin{pmatrix} G_1^i \\ G_2^i \\ \vdots \\ G_{\tilde{N}}^i \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \forall i = 1, \dots, q, \quad (4.1)$$

mentre la matrice n-dimensionale A viene costruita mantenendo tutte le informazioni del vettore H e ripetendole integralmente in ogni colonna:

$$A = \begin{pmatrix} H_1 & 0 & \cdots & 0 & H_{\tilde{N}} & \cdots & H_2 \\ H_2 & H_1 & \cdots & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 & \vdots & \ddots & H_{\tilde{N}} \\ H_{\tilde{N}} & H_{\tilde{N}-1} & \cdots & H_1 & 0 & \cdots & 0 \\ 0 & H_{\tilde{N}} & \cdots & H_2 & H_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_{\tilde{N}} & H_{\tilde{N}-1} & \cdots & H_1 \end{pmatrix} \quad (4.2)$$

In questo modo vengono simultaneamente definiti tutti i problemi da risolvere, 23-dimensionali, formulati come:

$$A \cdot x_i = b_i, \quad \forall i = 1, \dots, q \quad (4.3)$$

dove $x_i = BF \Delta t \cdot R$, in analogia col sistema matriciale 2.11 Il parametro perfusionale BF del pixel i –esimo viene poi ricavato come

$$BF = \max_{j=1, \dots, n} \left(\frac{x_{i,j}}{\Delta t} \right) \quad (4.4)$$

ossia come componente massima della corrispondente soluzione, riscalata per un fattore dovuto alla discretizzazione temporale del processo di acquisizione delle immagini tomografiche.

4.1.1 Preprocessing

Prima di risolvere tutti questi problemi, cerchiamo di analizzarli per poter valutare al meglio i metodi di risoluzione e di regolarizzazione che verranno applicati.

La figura 4.4 riporta alcune informazioni relative alla sola matrice A , comune a tutti i sistemi. Il grafico (a) mostra infatti l'andamento dei valori singolari: i primi sono molto alti, ma velocemente la curva si abbassa e dal decimo punto giace sotto quota 60. La maggior parte dell'informazione é quindi contenuta nelle prime 9 componenti. Il grafico (b) mostra invece l'andamento del numero di condizione delle matrici che ricostruiscono A per diadi, ossia il numero $k(A_k) = \frac{\sigma_1}{\sigma_k}$: coerentemente con quanto appena osservato dal primo grafico, le prime nove matrici non sono particolarmente mal consizionate, mentre le successive lo sono sempre di piú. Possiamo quindi pensare che una

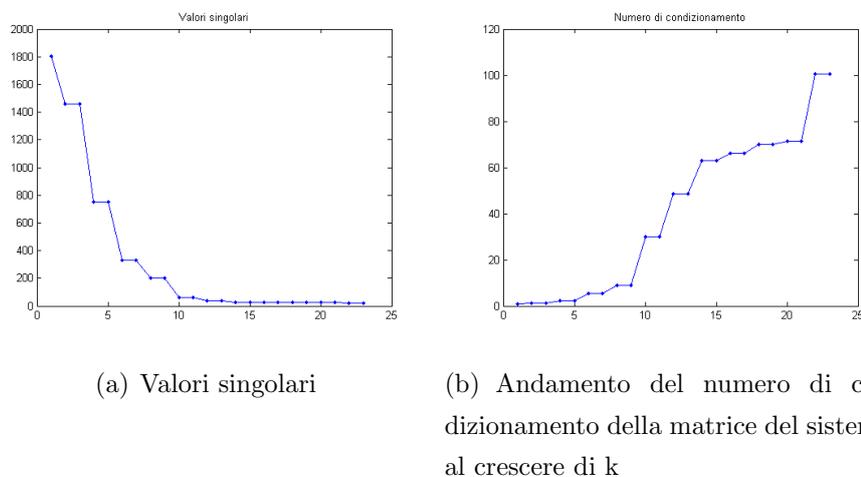
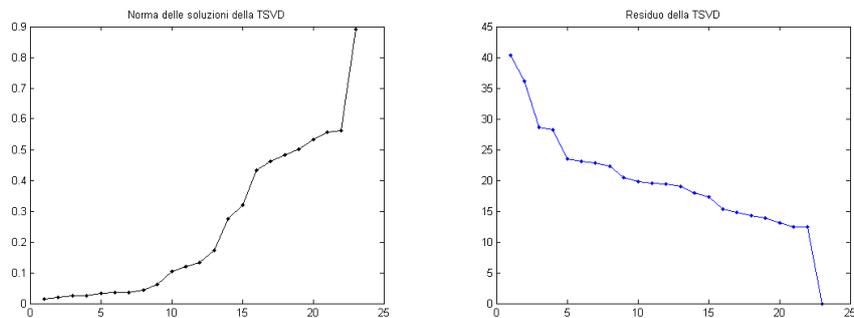


Figura 4.4: Grafici per lo studio preliminare del problema

buona TSVD non debba utilizzare parametri di regolarizzazione superiori a 9, cosí come il metodo di Tikhonov non debba servirsi di λ bassi.

Vediamo ora la figura 4.5, che é relativa a un problema fissato (per precisione, quello per $i = 500$), simile alla maggioranza degli altri sistemi.

Con la TSVD, per ogni valore possibile del parametro di regolarizzazione, sono state calcolate le soluzioni del problema e il grafico (a) ne riporta le norme, mentre il (b) mostra le norme dei residui corrispondenti. Se la norma delle x_k rimane bassa fino alla tredicesima prova, tanto da rimanere quasi costante nelle prime nove, la norma residua si riduce rapidamente nei primi cinque passaggi, si dimezza verso il decimo, poi non subisce particolari variazioni fino all'ultima prova, dove quasi si annulla. Questa discordanza, seppure limitata, potrebbe produrra L-curve dal vertice poco pronunciato.



(a) Andamento della norma della soluzione, al crescere di k (b) Andamento della norma del residuo, al crescere di k

Figura 4.5: Grafici per lo studio preliminare del problema

Un'ultima osservazione riguarda le curve di output tissutali, mostrate nella figura 4.3(b).

Paragonandole alla curva tissutale della simulazione (figura 3.3), é evidente che il rumore sui dati é molto elevato. Per questa ragione viene eseguita un'analisi del rumore, riportata dalla figura 4.6: di ogni curva tissutale originale C_i , si determina una curva approssimante C'_i (rossa) calcolando la media mobile relativa a 3 punti e centrata in ogni valore del first-pass, in modo da attenuare i picchi di intensitá dovuti al rumore. Precisamente, tali curve

vengono calcolate come

$$C'_i(t_j) = \begin{cases} \frac{1}{3} \sum_{\gamma=-1}^1 C_i(t_{j+\gamma}) & \forall j = 8, \dots, 19 \\ 0 & \text{altrimenti} \end{cases}, \forall i = 1, \dots, q$$

Le curve C'_i dei grafici in figura 4.6 sono infatti meno irregolari delle corrispondenti blu, pur mantenendone gli andamenti: si osservi per esempio il caso (b), dove il segnale originale mostra un'ampia oscillazione (inaspettata) attorno al quindicesimo valore e la curva di media smussa notevolmente tale irregolarità.

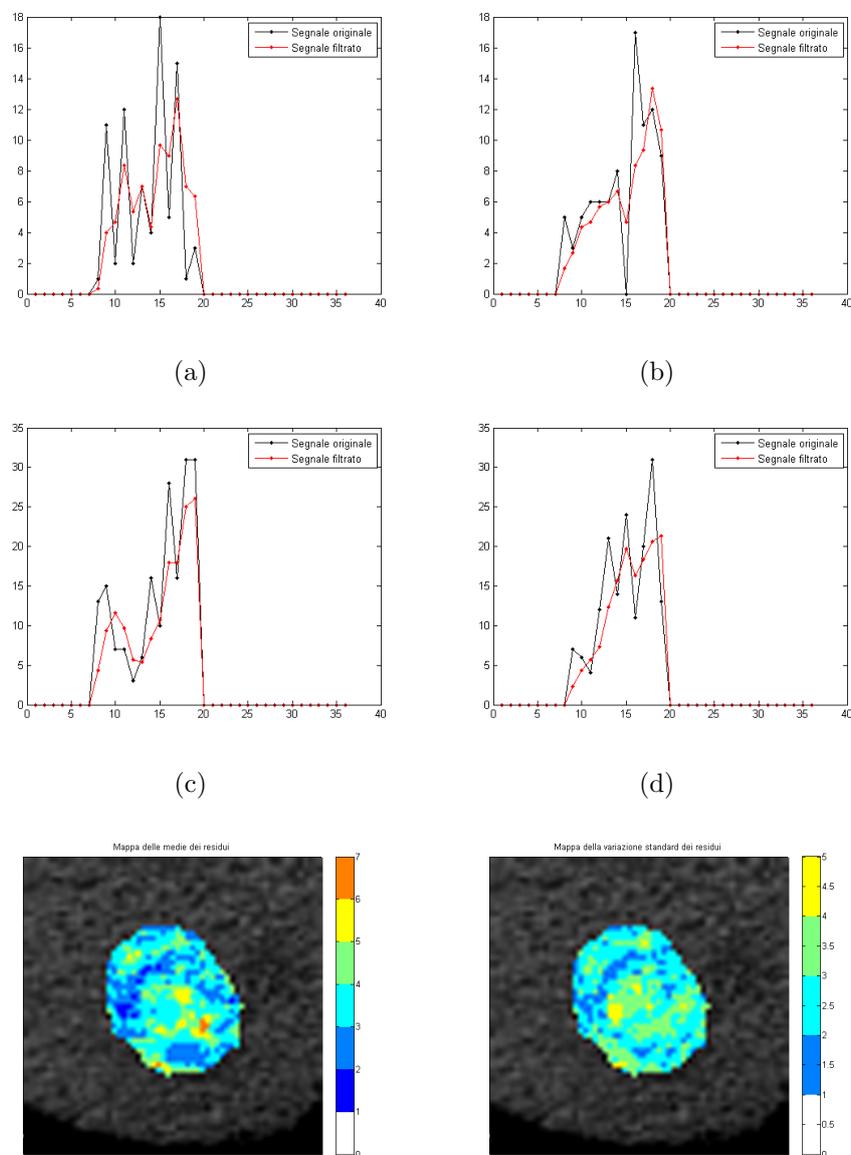
Successivamente si calcolano le curve residue

$$R_i(t_j) = \begin{cases} |C_i(t_j) - C'_i(t_j)| & \forall j = 8, \dots, 19 \\ 0 & \text{altrimenti} \end{cases}, \forall i = 1, \dots, q$$

e si utilizzano i valori statistici

$$\begin{cases} \mu_i = \frac{1}{12} \sum_{j=8}^{19} R_i(t_j) \\ \sigma_i^2 = \frac{1}{11} \sum_{j=8}^{19} (\mu_i - R_i(t_j))^2 \end{cases} \quad \forall i = 1, \dots, q \quad (4.5)$$

per riassumere tutte queste informazioni. Le mappe relative a queste due grandezze sono riportate nelle figure 4.6(e) e 4.6(f). Vengono considerati "buoni" i problemi relativi ai pixel i – *esimi* dove sia μ_i che σ_i^2 sono bassi: in questi casi, infatti, le curve C'_i sono simili alle rispettive C_i , quindi il rumore é limitato. Se, invece, μ_i é alto ma non σ_i^2 , allora i valori residui sono ovunque elevati e i risultati della curva in esame non sono particolarmente attendibili. Infine, alti valori sia di μ_i che di σ_i^2 corrispondono ai casi in cui C'_i si distanzia solo in alcuni punti dalla curva originale, quindi i risultati saranno buoni solo a tratti. Guardando alle due mappe della figura 4.6, capiamo allora che i punti nell'area centrale del tumore possono restituire risultati diversi, anche discordanti, al variare del metodo di risoluzione applicato, mentre dai punti nella fascia superiore ci aspettiamo maggior stabilità.



(e) Mappa coi valori di media dei residui (f) Mappa coi valori di standard deviation dei residui

Figura 4.6: Esempi di alcune curve, di output tissutale, mediate con un filtro di media mobile 3×3 e relative mappe colorimetriche dei residui, descritti tramite gli operatori di media e deviazione standard

4.2 Risultati

Vediamo ora i risultati ottenuti risolvendo tutti i sistemi coi metodi regolarizzati. Le mappe riportate corrispondono ai valori calcolati per il BF, opportunamente scalati per essere espressi in mL/s .

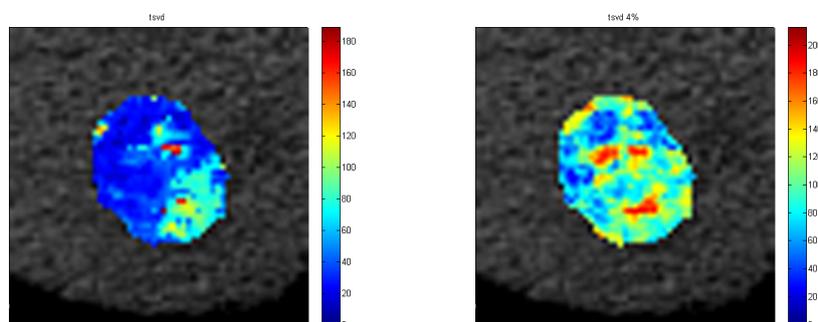
A differenza di quanto avveniva nel capitolo 3, ora non si possiede neppure una stima numerica di un buon output, ma ci si affida esclusivamente ai pareri forniti dai radiologi dell'IRST. Per questo esame epatico, tutti i medici consultati hanno confermato la diagnosi di maggior perfusione nella zona centrale della lesione, dove sembrano esserci due piccole masse piú dense. I contorni di tali masse non sono però definite con chiarezza: a volte i radiologi le descrivono piuttosto vicine tra loro, mentre osservando altre immagini sembrano essere piú distanti. Solo un'esame anatomico-patologico sarebbe in grado di fornire una diagnosi certa, ma per ora possiamo solo fidarci del parere medico.

D'altra parte, da un punto di vista strettamente matematico, é possibile calcolare indici e definire features capaci di dare un'indicazione della bontá delle TCC.

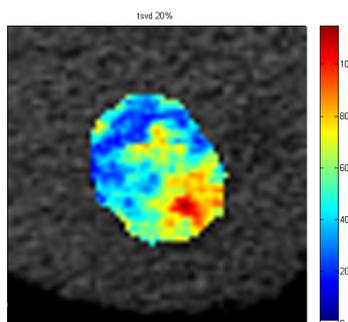
4.2.1 Risultati con la Tsvd

Cominciamo analizzando i risultati forniti dalla TSVD, riportati nella figura 4.7.

Innanzitutto osserviamo che il range dei valori assunti dal parametro perfunzionale é sempre accettabile. Nei casi (a) e (b) si hanno le soluzioni relative alla TSVD con parametro k , dato rispettivamente dalla L-curva o dal metodo *perc4*: qui BF arriva a valori vicini al 200, ma la maggioranza dei pixel indica livelli di perfusione piú bassi. Nella soluzione (c), invece, relativa alla TSVD con regolarizzazione di *perc20*, BF ha valori nell'intervallo $[0, 120]$ che é migliore dei due precedenti.



(a) Mappa ottenuta con la TSVD e la L-curva
(b) Mappa ottenuta con la TSVD e perc4



(c) Mappa ottenuta con la TSVD e perc20

Figura 4.7: Output con la TSVD classica

I metodi *perc4* e *perc20*, poiché decidono il parametro di troncamento dai valori della sola matrice dei sistemi, fissano indici k costanti per tutto il programma: in particolare, *perc4* determina $k = 9$ mentre *perc20* usa $k = 5$, che corrisponde al parametro stimato precedentemente, dall'osservazione delle figure 4.4 e 4.5.

Le loro mappe risultano effettivamente buone: nella (c) si distingue la massa tumorale ad alta perfusione, giacente nella parte inferiore della neoplasia, mentre la seconda non è stata individuata. Al contrario, la figura (b)

evidenzia entrambe le masse, anche se quella superiore viene sconnessa in due macchie molto vicine.

Al contrario, la mappa (a) non costituisce un buon risultato perché presenta molto blu, cioè valori inferiori a 40, e nessuna massa ad alta perfusione: evidentemente la L-curva ha fornito indici molto piccoli, che hanno eccessivamente regolarizzato gli output. A questo punto è interessante guardare la figura 4.8, che riporta alcune L-curve realmente ottenute. Nonostante siano molto diverse tra loro, in entrambi i casi i vertici delle "L" sono facilmente rintracciabili a occhio, ma non lo sono sempre per gli algoritmi di automatizzazione: nella curva (b), più regolare della prima, il punto evidenziato giace infatti nel tratto più orizzontale della curva.

D'altra parte, le curve mostrate nella figura 4.9 non sono certo migliori.

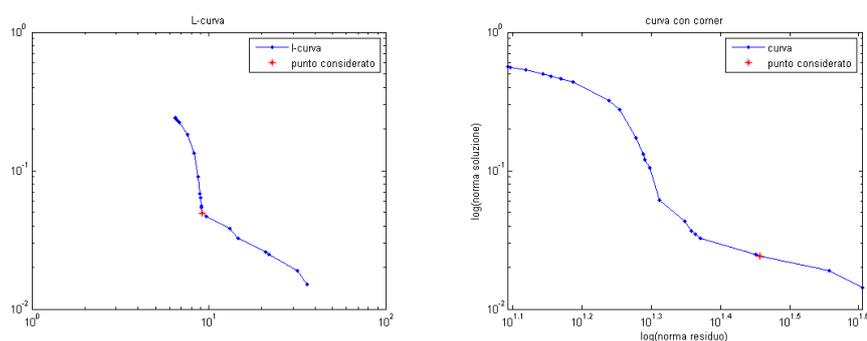


Figura 4.8: Esempi di L-curve della TSVD

Quelle riportate in (a) e (b) sono rispettivamente le curve residuali e condizionali dello stesso sistema avente per L-curva il grafico 4.8(b): la forma a "L" non è affatto evidente nella prima, mentre nella seconda la presenza di numerosi "gradini" non rende palese (neppure all'occhio umano) dove si trovi il presunto vertice. Per gli altri pixel, le curve sono analogamente poco utili.

Come si può notare direttamente dalle curve (c) e (d), che graficano le funzioni di Regińska del medesimo problema, anche queste alternative per la L-curva non possono restituire risultati affidabili: viene infatti fissato $k = 1$

in entrambi i casi, per la presenza dei valori minimizzanti agli estremi del dominio. Tali funzioni, inoltre, sono molto simili per tutti gli 864 problemi.

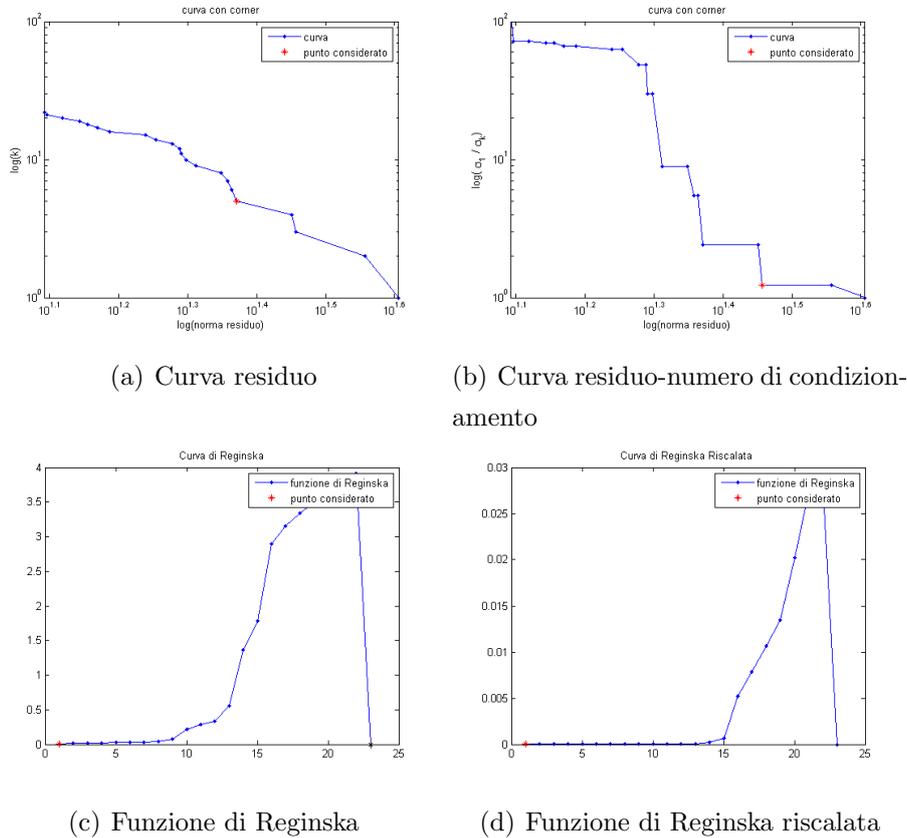


Figura 4.9: Curve alternative alla L-curve

Risultati con Picard

Anche se non é tipicamente illustrata come un metodo per la scelta del parametro di regolarizzazione, la condizione discreta di Picard (vista nel primo capitolo) puó essere sfruttata per determinare l'indice di troncamento per la TSVD. La figura 4.10 mostra i grafici di Picard di due problemi: chiaramente la curva dei valori singolari é la medesima ma, variando il termine noto, cambiano i coefficienti per le soluzioni. Nel primo caso, le oscillazioni della curva rossa indicano la presenza di un rumore piú intenso sul segnale

tissutale: già dal terzo punto, la successione non è più decrescente, mentre dal settimo in poi l'andamento è totalmente irregolare. Nel secondo caso, invece, la decrescita dei valori $|u_i^t b|$ è maggiormente contenuta: fino al tredicesimo valore, i coefficienti $\frac{|u_i^t b|}{\sigma_i}$ possono ricostruire una soluzione accettabile. Alla luce di queste osservazioni e di quelle effettuate nella prima parte di questo capitolo, è stata testata la TSVD con parametri di regolarizzazione costanti: la figura 4.11 ne mostra alcune mappe.

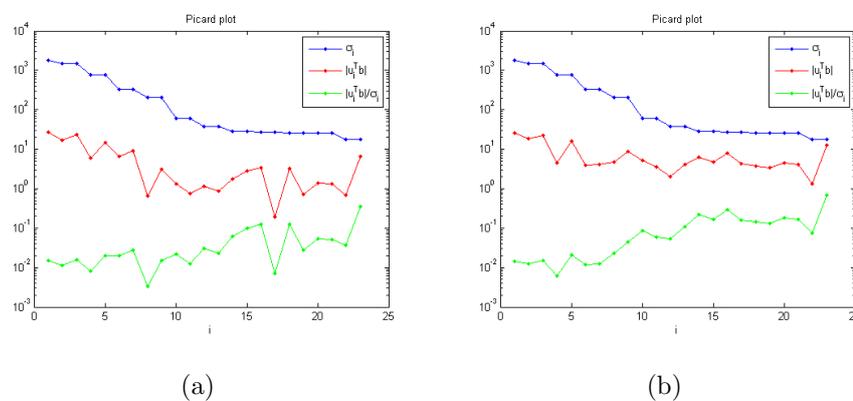
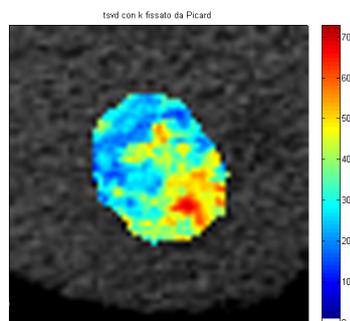
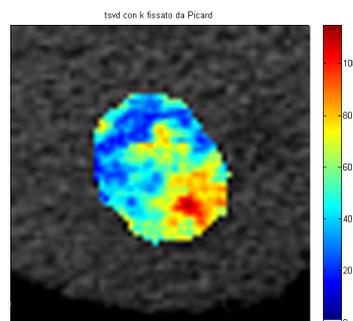
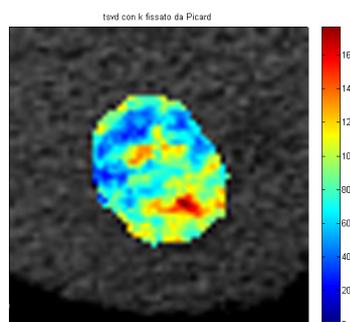
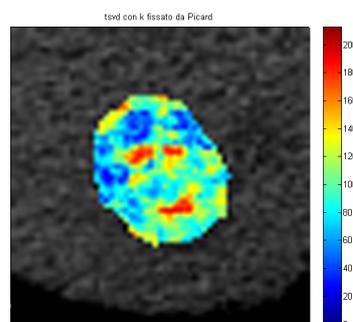
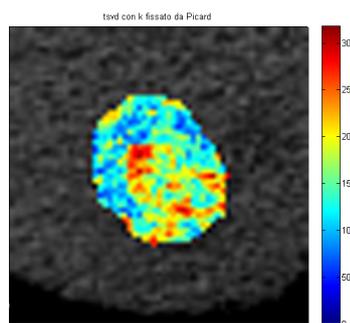
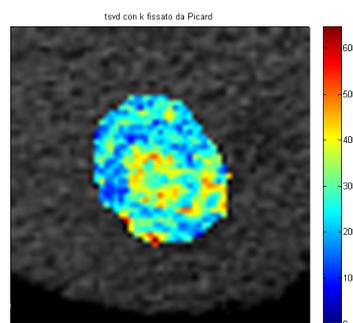


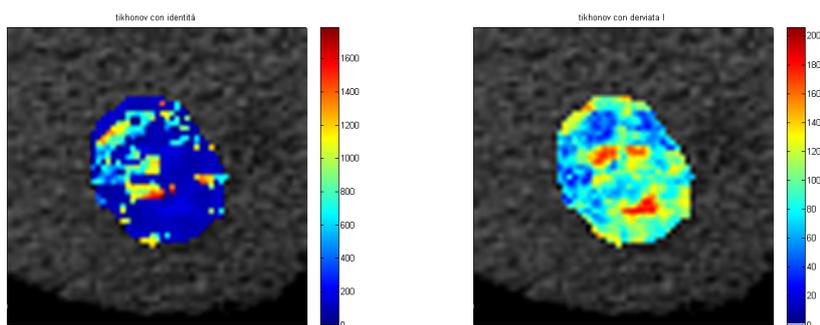
Figura 4.10: Grafici di Picard

Osserviamo che al crescere di k si registra una progressiva espansione del range dei valori per BF, che concorda con la calante regolarizzazione imposta. Inoltre, nelle prime due mappe, la massa superiore ad alta perfusione è quasi impercettibile e l'inferiore non è ben definita, mentre nelle ultime due non sono più distinguibili e la distribuzione dei colori appare molto casuale. Al contrario, le mappe (c) e (d) riportano output più soddisfacenti, perché almeno hanno localizzato le due masse nelle giuste posizioni. In particolare, questi due risultati corrispondono alle esecuzioni della TSVD dove il parametro di regolarizzazione è stato fissato rispettivamente a $k = 7$ e $k = 9$, valori osservati come possibili soglie, dall'analisi dei grafici precedenti: anche le analisi dell'andamento dei valori singolari e della condizione di Picard possono essere tenute in forte considerazione, per la scelta del parametro di regolarizzazione.

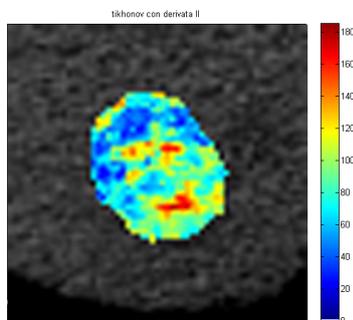
(a) Mappa ottenuta per $k=3$ (b) Mappa ottenuta per $k=5$ (c) Mappa ottenuta per $k=7$ (d) Mappa ottenuta per $k=9$ (e) Mappa ottenuta per $k=11$ (f) Mappa ottenuta per $k=13$ Figura 4.11: Output della TSVD con parametro k di regolarizzazione fissato

4.2.2 Risultati con Tikhonov

Passiamo ora all'analisi degli output dati dalla regolarizzazione di Tikhonov. La figura 4.12 mostra le tre mappe ottenute utilizzando rispettivamente gli operatori matriciali $L=Id$, $L=\nabla$ e $L=\nabla^2$, definiti nell'equazione 3.2 e di dimensione $n \times n$.



(a) Mappa ottenuta con Tikhonov e $L=Id$ (b) Mappa ottenuta con Tikhonov e $L=\nabla$



(c) Mappa ottenuta con Tikhonov e $L=\nabla^2$

Figura 4.12: Output con il metodo di Tikhonov

Osserviamo subito che l'output (a) contiene parecchi punti ad elevatissima perfusione, dove $BF \in [800, 1800]$, e solo quelli rappresentati in blu hanno valori in un range accettabile. Inoltre, la presenza dei punti ad alti valori

provoca il ridimensionamento della Colorbar, che rende impossibile capire se almeno per i pixel blu il calcolo del parametro perfusionale sia corretto. D'altra parte, vista la cospicua quantità di punti in cui la regolarizzazione non é stata sufficiente, l'intero output viene considerato non attendibile. Le immagini (b) e (c), al contrario, costituiscono buoni output: in entrambi i casi, infatti, BF ha valori in un range accettabile e le due masse a maggior perfusione sono ben evidenti, staccandosi dallo sfondo azzurro e blu con tonalità arancioni e rosse. Come per gli output della TSVD, la massa inferiore viene individuata meglio dell'altra, perché la superiore ha contorni sempre molto incerti. Quest'ultima, infatti, non viene determinata come unica massa connessa: nei punti dove i radiologi hanno spiegato trovarsi tale massa, gli algoritmi hanno semplicemente trovato una maggiore perfusione.

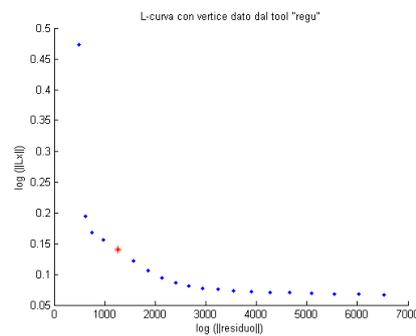
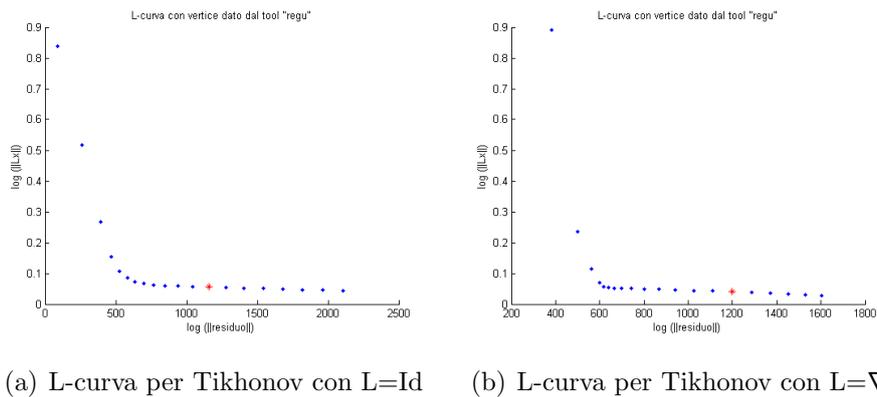


Figura 4.13: L-curve relative al metodo di Tikhonov

I valori di λ provati per la regolarizzazione sono i multipli di 10 compresi tra 10 e 200: la figura 4.13 riporta una L-curva significativa, per ciascuna delle tre varianti del metodo di Tikhonov. L'andamento delle "L" é buono, anche se non sempre vengono utilizzati i valori corrispondenti al vero vertice delle curve: questa é la principale causa di errore per il metodo che utilizza la matrice identità come operatore, poiché λ viene scelto a volte troppo basso e altre volte troppo alto. Nel caso di $L=\nabla$, invece, i valori considerati cadono spesso nel tratto orizzontale della "L", probabilmente perché la distribuzione dei punti non é omogenea lungo la curva e pochi appartengono al tratto verticale: ad ogni modo, questo impone una regolarizzazione molto forte che, come abbiamo visto, produce buoni output finali.

Il terzo caso é analogo, ma la presenza di un vertice "ampio" porta mediamente a usare valori piú bassi di λ .

Non ne vengono riportate le mappe finali, ma sono state provate alcune varianti del metodo di Tikhonov, basate sulla formulazione espressa in 1.9. Come soluzioni di riferimento x^* , sono state testate le curve ottenute come soluzioni dei sistemi meglio risolti con la TSVD (cioé quelle relative sia a *perc4* che a *perc20*), ma le mappe finali non sono buone. Ottenere una buona stima complessiva dei BF su tutta la massa tumorale, in questo caso non significa avere una soluzione regolare per ciascun problema inverso, perché il parametro perfusionale é estrapolato dalla soluzione del sistema come mostrato in 4.4 e questo non é un metodo particolarmente robusto. Le curve proposte come x^* non sono infatti regolari, ma presentano oscillazioni e picchi a cui le altre soluzioni non devono necessariamente tendere. Se ne conclude che, in questo caso, chiedere che due soluzioni siano simili non é una strategia vincente.

4.2.3 Osservazioni

In letteratura non sono ancora stati definiti con precisione degli indici di qualità, con cui valutare le singole mappe colorimetriche. D'altra parte,

é possibile ricavare delle informazioni dal confronto degli output ottenuti, ad esempio tramite l'analisi di correlazione delle mappe stesse. La tabella 4.1 riporta i coefficienti di correlazione che legano tra loro i risultati precedentemente mostrati.

	TSVD			TIKHONOV			PICARD			
	lcurva	perc4	perc20	L=Id	L= ∇	L= ∇^2	k=5	k=7	k=9	k=11
lcurva	1.0000	0.8925	0.9488	0.4504	0.8987	0.9119	0.9488	0.9247	0.8925	0.8614
perc4	0.8925	1.0000	0.9549	0.5572	0.9985	0.9928	0.9549	0.9804	1.0000	0.9533
perc20	0.9488	0.9549	1.0000	0.5344	0.9602	0.9710	1.0000	0.9848	0.9549	0.9427
L=Id	0.4504	0.5572	0.5344	1.0000	0.5539	0.5444	0.5344	0.5408	0.5572	0.6682
L= ∇	0.8987	0.9985	0.9602	0.5539	1.0000	0.9934	0.9602	0.9855	0.9985	0.9549
L= ∇^2	0.9119	0.9928	0.9710	0.5444	0.9934	1.0000	0.9710	0.9859	0.9928	0.9522
k=5	0.9488	0.9549	1.0000	0.5344	0.9602	0.9710	1.0000	0.9848	0.9549	0.9427
k=7	0.9247	0.9804	0.9848	0.5408	0.9855	0.9859	0.9848	1.0000	0.9804	0.9528
k=9	0.8925	1.0000	0.9549	0.5572	0.9985	0.9928	0.9549	0.9804	1.0000	0.9533
k=11	0.8614	0.9533	0.9427	0.6682	0.9549	0.9522	0.9427	0.9528	0.9533	1.0000

Tabella 4.1: Coefficienti di correlazione delle mappe migliori

Questi dati confermano che le soluzioni piú discordanti sono quelle ottenute dal metodo di Tikhonov con $L=Id$ e dalla TSVD con la sola L-curva o con il troncamento fissato da indici medio-alti ($k = 11$). Come già si capiva "a occhio", le altre mappe sono molto simili tra loro, poiché hanno coefficienti superiori allo 0,95. Da un punto di vista statistico, questi valori lasciano sperare che la concordanza delle mappe sia dovuta alla bontá delle soluzioni e dei valori ricavati, ma non se ne ha la certezza matematica.

Un altro metodo per valutare gli output é osservare l'istogramma dei valori calcolati per BF. Tale analisi é indipendente dalla descrizione, fornitaci dai radiologi, circa la presenza e la collocazione delle masse ad elevata perfusione, perché valuta l'attendibilitá delle mappe guardandone esclusivamente la distribuzione dei valori. Al contrario, la sola analisi della correlazione tra le mappe non é in grado di stabilire la bontá dei risultati: se la maggior parte dei risultati sbagliasse, anche solo in una precisa regione, e se un solo metodo fornisse una buona ricostruzione della realtà, i coefficienti relativi

a quest'ultimo output sarebbero tutti molto bassi, al punto che il metodo robusto verrebbe scartato, poiché "anomalo".

Consideriamo ora la figura 4.14, nella quale sono riportati gli istogrammi relativi ai metodi classici visti per la TSVD e per Tikhonov: la bontà delle mappe, si riflette nelle proprietà degli istogrammi. Nei casi (b), (e) e (f), i cui output sono stati ritenuti migliori già in precedenza, gli istogrammi approssimano una distribuzione gaussiana, con valori compresi nel giusto range di $[0, 200]$. Al contrario, negli altri grafici tale equilibrio non è presente: se la TSVD con la L-curva assume pochissime volte valori superiori a 100, troppi punti registrano un BF inferiore a 40. Nel caso estremo della risoluzione col metodo di Tikhonov e $L=Id$, si capisce bene che molti pixel oltrepassano la soglia 200 e la distribuzione è sproporzionata. Infine, il caso (c) rappresenta una situazione intermedia: non ci sono punti anomali e, come già osservato, il range è molto limitato, ma la distribuzione dei valori è fin troppo uniforme nell'intervallo $[20, 80]$.

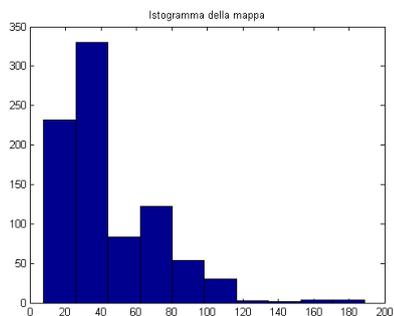
Tuttavia, non si conosce una distribuzione ottimale per la soluzione di questi problemi, quindi nessuna tecnica di *histogram matching* può riconoscere il risultato migliore. D'altra parte, risulta evidente che istogrammi non regolari corrispondono a soluzioni non affidabili: questo risultato può essere utilizzato per scartare output non validi, con una certa "sicurezza matematica".

Nei casi in cui gli istogrammi sottolineano la presenza di pochi punti a valori particolarmente elevati, la letteratura suggerisce di effettuare un *clipping* di tali presunti "errori sporadici", eliminando l'ultimo 5 percentile dei risultati stessi e sostituendoli con un valore neutro, come lo zero o il massimo valore considerabile. Tale tecnica è stata testata nei casi della TSVD con la L-curva e di Tikhonov con operatore $L=Id$, viste le caratteristiche dei relativi istogrammi: i risultati sono riportati nella figura 4.15.

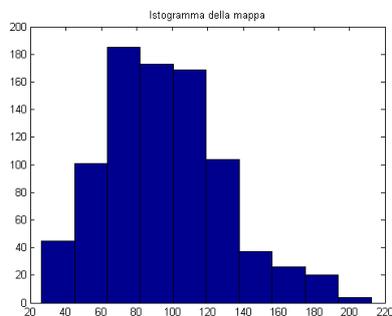
Confrontando queste mappe con le originali (figure 4.7(a) e 4.12(a)), si apprezza il ridimensionamento della colorbar, che permette una visualizzazione migliore dei valori perfusionali. D'altra parte, le soluzioni non sono ancora

buone, perché non vengono evidenziate le due masse ad alta perfusione che ci aspettiamo di localizzare.

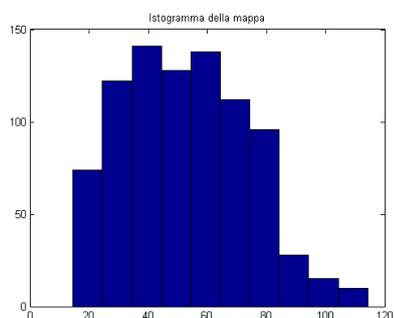
Guardando agli istogrammi di queste due nuove mappe, notiamo ancora una distribuzione irregolare dei valori graficati. Nel secondo caso sarebbe opportuno un clipping piú incisivo, ma eliminare i valori superiori a 200 (soglia suggerita dall'osservazione dell'istogramma (d)) significa eliminare troppi dati. Analogamente, manipolare ulteriormente la mappa della TSVD non porterebbe a nessun risultato utile. Si ritiene pertanto che se il metodo di regolarizzazione non é stato appropriato per un piccolo ma significativo cluster di punti, come in questi due casi, allora anche i risultati che fornisce per gli altri pixel non siano attendibili.



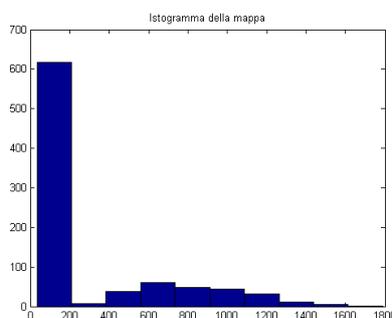
(a) Istogramma relativo alla mappa della TSVD con la L-curva



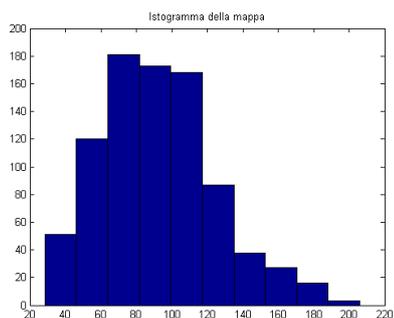
(b) Istogramma relativo alla mappa della TSVD con $perc_4$



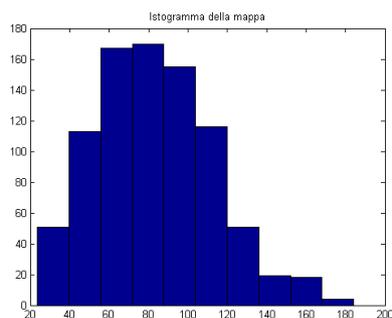
(c) Istogramma relativo alla mappa della TSVD con $perc_{20}$



(d) Istogramma relativo alla mappa del metodo di Tikhonov con $L=Id$

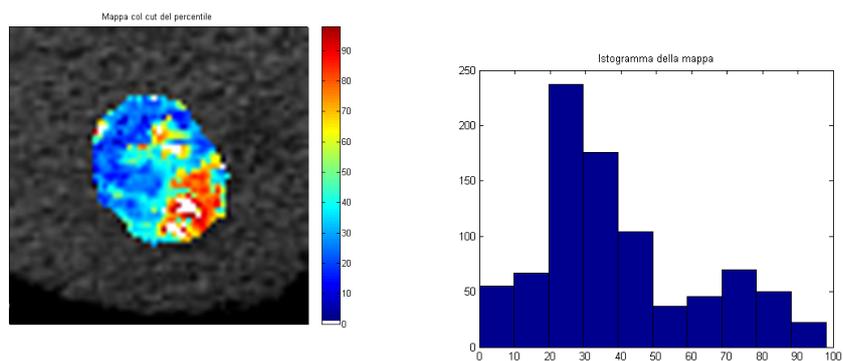


(e) Istogramma relativo alla mappa del metodo di Tikhonov con $L=\nabla$



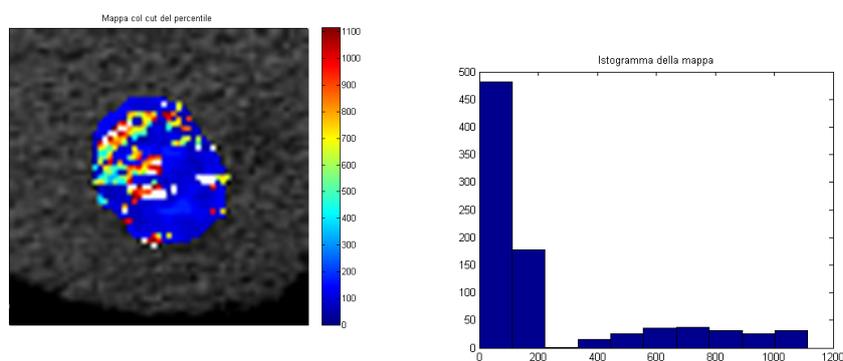
(f) Istogramma relativo alla mappa del metodo di Tikhonov con $L=\nabla^2$

Figura 4.14: Istogrammi delle mappe riportate nelle figure 4.7 e 4.12



(a) Nuova mappa della TSVD con la L-curve

(b) Istogramma della mappa (a)



(c) Nuova mappa del metodo di Tikhonov con $L=Id$

(d) Istogramma della mappa (c)

Figura 4.15: Mappe e relativi istogrammi, ottenuti cancellando l'ultimo 5 percentile

Conclusioni

In questa tesi sono stati analizzati e confrontati i metodi della TSVD e di Tikhonov, di regolarizzazione per problemi discreti inversi e mal posti, assieme ad alcune tecniche per la scelta del parametro di regolarizzazione. Al primo metodo sono stati applicati diversi metodi per la scelta del parametro, al secondo esclusivamente il criterio della L-curva, ma é stato variato l'operatore che esso prevede.

La regolarizzazione della TSVD non é stata sempre soddisfacente: sfruttando la L-curva, infatti, i risultati non sono stati di buoni. Al contrario, sono stati ottenuti validi output quando la scelta del parametro di regolarizzazione era basata sull'analisi della matrice dei sistemi o della condizione discreta di Picard. In questi casi, infatti, la regolarizzazione dipende esclusivamente dalle caratteristiche intrinseche del problema di partenza, quindi non considera in alcun modo la propagazione del rumore che domina sui dati e velocemente passa sulle soluzioni enfatizzandosi.

La regolarizzazione col metodo di Tikhonov é stata efficiente quando alla soluzione veniva imposta una *smoothing norm* non banale, tramite gli operatori di derivata prima e seconda (in forma matriciale). Anche in questo modo, infatti, si affronta l'instabilitá dovuta al forte rumore.

Ad ogni modo, siamo riusciti ad automatizzare la scelta del parametro di regolarizzazione in maniera positiva e a irrobustire cosí tutta la parte di deconvoluzione che ci era stata affidata dai ricercatori dell'ARCES bolognese.

Sicuramente l'intero modello bio-medico puó essere raffinato, curando maggiormente le fasi di *pre-* e di *post-processing*. Da un lato, infatti, sarebbe

opportuno studiare meglio la componente del rumore, con indagini statistiche, e definire un protocollo di acquisizione delle immagini tomografiche perfusionali, per poter standardizzare e ottimizzare i parametri relativi a questa prima fase. Di conseguenza, sarebbe anche possibile migliorare la simulazione del problema, rendendola piú veritiera e utile per il test dei modelli matematici, e lavorare con dati migliori.

D'altro lato, occorre determinare dei parametri numerici, delle caratteristiche o dei test che possano valutare la qualità degli output, in maniera autonoma e concordante rispetto al parere dei radiologi, o almeno permettere di allenare dei *support vector machines* per costruire algoritmi validi che generino nuove mappe. Quest'ultimo punto é di fondamentale importanza per la ricerca, che si prefigge di automatizzare i processi diagnostici, seppure compiendo piccoli passi per volta, per affiancare e sostenere i medici in questo complicato compito.

Bibliografia

- [1] P. C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, SIAM, Philadelphia, 1997.
- [2] P. C. Hansen, *Discrete inverse problems: insight and algorithms*, SIAM, Philadelphia, 2010.
- [3] P. C. Hansen, *Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numerical Algorithms, n. 6, pp. 1-35, 1994.
- [4] P. C. Hansen, *Regularization Tools Version 4.0 for Matlab 7.3*, Numerical Algorithms, 46, pp. 189-194, 2007.
- [5] <http://www2.imm.dtu.dk/~pcha/Regutools/>.
- [6] J. Hadamard, *Lectures on Cauchy's problem in Linear Partial Differential Equations*, Yale University Press, New Haven, CT, 1923.
- [7] R. C. Gonzales e R. E. Woods, *Digital Image Processing, 3rd edition*, Pearson Education, 2008.
- [8] L.Reichel e G.Rodriguez, *Old and new parameter choice rules for discrete ill-posed problems*, Numer. Algor. DOI 10.1007/s11075-012-9612-8.
- [9] T. Regińska, *A regularization parameter in discrete ill-posed problems*, SIAM, J. Sci. Comput, 17, pagg 740-749, 1996.

-
- [10] Silvia Malavasi, *Studio e sviluppo di algoritmi di elaborazione di immagini da TC per analisi perfusionali*, CdS Ing. Biomedica, sessione prima, A.A. 2011-2012
- [11] I. He, B. Orten, S. Do, W. C. Karl, A. Kambadakone, D.V. Sahani e H. Pien, *A spatio-temporal deconvolution method to improve perfusion CT quantification*, IEEE Transaction on Medical Imaging, vol. 29, n. 5, pagg 1182-1191, 2010.
- [12] K. A. Miles, *Perfusion CT for the assessment of tumour vascularity: which protocol?*, The British Journal of Radiology, vol. 76, pagg S36-S42, 2003.
- [13] K. A. Miles e R. Griffiths, *Perfusion CT: a worthwhile enhancement?*, The British Journal of Radiology, vol. 76, pagg 220-231, 2003
- [14] T. S. Koh, C. H. Thng, P. S. Lee, S. Hartono, H. Rumpel, B. C. Goh e S. Bisdas, *Hepatic Metastases: in vivo assessment of perfusion parameters at dynamic contrast-enhanced MR imaging with dual-input two-compartment traces kinetics model*, Radiology, vol. 249, n.1, 2008.
- [15] G. Petralia, L. Preda, G. D'Andrea, S. Viotti, L. Bonello, R. De Filippi e M. Bellomi *CT perfusion in solid-body tumors. Part I: technical issues*, Radiol Med, vol. 115, pagg. 843-857, 2010.
- [16] M. Bellomi, S. Viotti, L. Preda, G. D'Andrea, L. Bonello e G. Petralia, *CT perfusion in solid-body tumors, part II. Clinical applications and future development*, Radiol Med, vol. 115, pagg. 858-874, 2010.

Ringraziamenti

Un grazie di cuore alla Prof. Elena Loli Piccolomini, per la disponibilità mostrata nei miei confronti in questo biennio magistrale, perché per me è stata un rassicurante punto di riferimento e di confronto.

Grazie anche alla Prof. Germana Landi per la sua gentilezza e passione, oltre che per l'aiuto fornitomi per questa tesi.

Grazie all'Ing. Silvia Malavasi, per la pazienza con cui mi ha istruita e la professionalità con cui mi ha guidata nei mesi della tesi, aiutandomi a realizzare questo progetto.

Grazie a Gabri e a Fra, che più di tutti hanno condiviso con me questi 5 anni, su quaderni Monocromo e riassunti colorati, a parlare di GMG o di giochi dell'IPad, su panchine all'aria aperta o in fredde stanze seminterrate, davanti a kebab, insalate o piadine-smezzate-del-post-esame.

Infine, grazie alla matematica, che in questi 5 anni è entrata nella mia testa e l'ha cambiata.

Grazie alla matematica, che è stata il primo argomento di conversazione con uno sconosciuto informatico... che da 5 mesi è mio marito.

Grazie alla matematica, per questa tesi così coinvolgente che mi ha offerto. Ora che la sto terminando, mi rendo conto che sono davvero entrata nel problema matematico, come mai mi sarei aspettata: i problemi numerici sono diventati i miei problemi, ma non le mie ansie da studentessa; mi sono

sentita vulnerabile come la norma della soluzione x , alla propagazione degli errori... anzi, mi sentivo proprio la soluzione x ! Ma poi ho sentito anche di avere in mano un'arma, la spada della regolarizzazione, che mi difende e con cui affronto quei diavoletti piccoli e insidiosi dei σ_i troppo bassi... con la mia lama, taglio e anniento le componenti delle k troppo alte, che vogliono entrare in me e impossessarsi della mia forma graziosa e regolare. Picard e la L-curva sono diventati i miei saggi consiglieri in questa buffa battaglia... No, non sono diventata pazza! Sono solo entrata nella parte... gli attori bravi si riconoscono così, no?!