



UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

Continuazione analitica della funzione self-energia nel
metodo GW

Relatore

Prof. Paolo Umari

Laureando

Andrea Lizzit

Anno Accademico 2021/2022

Continuazione analitica della funzione self-energia nel metodo GW

Andrea Lizzit

Abstract

La fisica dei sistemi debolmente correlati, sia molecolari sia solidi, è ben studiabile con il metodo della teoria del funzionale della densità (DFT). Tale metodo permette di esprimere il problema di N elettroni interagenti in quello di N elettroni non interagenti che risentono di un potenziale dipendente dalla sola densità elettronica totale. Nonostante la DFT permetta il calcolo di proprietà di stato fondamentale con grande accuratezza, essa riproduce le proprietà di eccitazione (le energie di ionizzazione o le affinità elettroniche) solo qualitativamente. Per ovviare a tale limite è stato introdotto il metodo GW rigorosamente basato sulla teoria delle perturbazioni a molti-corpi in cui la funzione self-energia viene espressa come prodotto della funzione di Green G non-perturbata e della interazione Coulombiana schermata W . In pratica i livelli elettronici DFT vengono corretti risolvendo una semplice equazione auto-consistente uni-dimensionale in cui appare il valore di aspettazione della self-energia calcolato per l'orbitale DFT. Tale valore è quindi una funzione dell'energia. Per semplicità di calcolo è conveniente calcolare prima tale funzione lungo l'asse delle energie immaginarie e poi effettuare numericamente la continuazione analitica di tale funzione lungo l'asse reale. L'obiettivo della tesi è analizzare in dettaglio vari metodi per tale operazione. Lo studente per prima cosa si familiarizzerà con il metodo GW. Poi, investigherà l'efficacia di procedure standard di continuazione analitica quali il fit non lineare con una formula a multi-poli e con gli approssimanti di Padè. A tal fine lo studente, svilupperà dei codici capaci di leggere le funzioni di self-energia calcolate (sull'asse immaginario) dal software *Quantum ESPRESSO*. Infine, lo studente analizzerà una strategia apparsa recentemente in letteratura basata sull'uso di reti neurali.



Capitolo 1

Introduzione

Lo studio della struttura elettronica di materiali debolmente correlati è di grande interesse per lo sviluppo di materiali con proprietà vantaggiose e per le potenziali applicazioni tecnologiche di alcuni loro comportamenti. La complessità dei sistemi a molti corpi rende necessaria l'introduzione di approssimazioni che rendano il problema risolvibile con metodi numerici. La density functional theory (DFT) è spesso utilizzata per studi di struttura elettronica perchè consente di ricondursi ad un problema a singola particella per il quale si possono facilmente calcolare autofunzioni e autovalori di energia. A causa delle notevoli approssimazioni introdotte dalla DFT, essa non descrive accuratamente gli stati eccitati, tuttavia gli autovalori ottenuti tramite la DFT possono essere corretti applicando il metodo GW, che si è rivelato estremamente efficace e sorprendentemente accurato. Uno dei passaggi più delicati in tutto il procedimento è il calcolo della funzione self-energia, che è ben nota lungo l'asse immaginario ma deve essere calcolata lungo l'asse reale del piano complesso. Bisogna quindi effettuare una continuazione analitica della self-energia dall'asse immaginario all'asse reale. Benchè in teoria la continuazione analitica della self-energia sia unica[1], in pratica la precisione numerica finita e le risorse computazionali limitate fanno in modo che la soluzione non sia ben definita; pertanto risulta necessario sviluppare dei metodi che riescano a riprodurre il comportamento corretto della funzione sull'asse reale. In questo lavoro di tesi è stato sviluppato il software *Quantum-SEF* che implementa diversi metodi descritti in letteratura per effettuare la continuazione analitica a partire dai dati prodotti dal software *Quantum ESPRESSO* [12]. *Quantum-SEF* è stato scritto in linguaggio Python, consta di circa 1500 righe divise in moduli estendibili ed è un software Open-Source liberamente disponibile su Github[11].

1.1 L'equazione di Dyson

In un sistema a molti corpi ogni particella interagisce con le altre e a causa di questa interazione l'evoluzione del suo stato è in generale diversa da quella che sarebbe stata nel caso della particella isolata. Un caso limite è quello di N fermioni non interagenti (un sistema a molti corpi in cui le particelle presentano un'interazione nulla). Questo sistema ha una soluzione molto semplice, infatti la soluzione a molti corpi può essere fattorizzata nel prodotto di soluzioni a singola particella ed è quindi sufficiente trovare una soluzione per il sistema a singola particella per ottenere direttamente la soluzione del sistema completo. I calcoli DFT permettono di ricondursi ad un problema di questo tipo in cui le particelle non interagiscono tra loro ma solo con un potenziale di interazione medio.

Al limite opposto vi è un sistema fortemente correlato come un materiale superconduttore, in cui la funzione d'onda elettronica è ben lontana dall'essere fattorizzabile in soluzioni a singola particella e non è possibile limitarsi a considerare solo le interazioni tra elettroni vicini, ma è necessario studiare la funzione d'onda elettronica a molti elettroni nel suo complesso.

Nei casi intermedi capita spesso che il comportamento del sistema sia diverso da quello di N particelle non interagenti ma che possa essere studiato in modo perturbativo a partire dalla soluzione del sistema

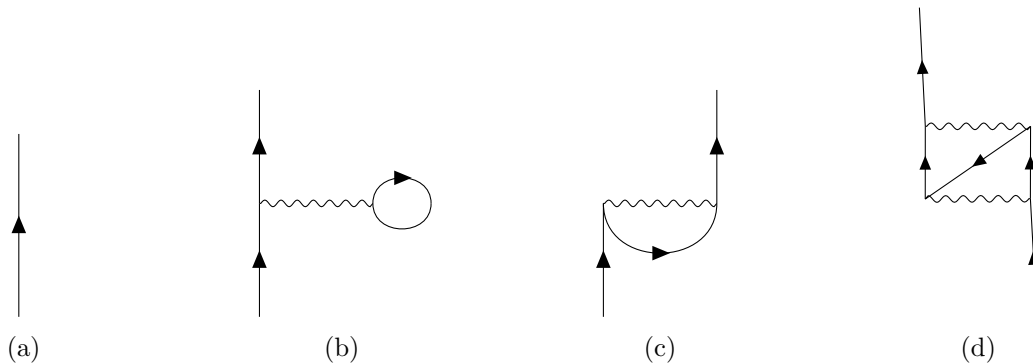


Figura 1.1: alcuni possibili processi di interazione

non interagente. In questi casi sono disponibili strumenti matematici e computazionali efficaci per studiare le proprietà del sistema.

L'oggetto matematico di principale interesse per descrivere il comportamento di un sistema è il propagatore G ad una particella, definito dalla seguente equazione:

$$iG(1, 2) = \langle N | T[\hat{\psi}(1)\hat{\psi}^\dagger(2)] | N \rangle \quad (1.1)$$

$$= \theta(t_1 - t_2) \langle N | \hat{\psi}(1)\hat{\psi}^\dagger(2) | N \rangle - \theta(t_2 - t_1) \langle N | \hat{\psi}^\dagger(2)\hat{\psi}(1) | N \rangle \quad (1.2)$$

in cui N è il numero di particelle, 1 e 2 rappresentano sinteticamente le coordinate spazio-temporali e di spin della rispettiva particella, T è l'operatore di *time-ordering* di Wick, che comprende un segno meno per uno scambio di operatori, con θ si indica la funzione di Heaviside e con ψ^\dagger e ψ gli operatori di creazione e distruzione in visuale di Heisemberg, i quali contengono quindi tutta la dipendenza dalle variabili temporali. Attraverso una trasformata di Fourier si ottiene la rappresentazione in frequenza del propagatore, anche detta di Lehmann:

$$G(\mathbf{r}_1, \mathbf{r}_2; \omega) = \sum_i \frac{A^i(\mathbf{r}_1)A^{i*}(\mathbf{r}_2)}{E^{N+1,i} - E^{N,0} - \omega - i\eta_i} - \sum_j \frac{B^j(\mathbf{r}_2)B^{j*}(\mathbf{r}_1)}{E^{N-1,j} - E^{N,0} + \omega - i\eta_j} \quad (1.3)$$

Le funzioni $A^i(\mathbf{r})$ e $B^j(\mathbf{r})$ sono chiamate ampiezze di quasiparticella mentre con $E^{N,i}$ o $E^{N,j}$ si indica l'energia dello stato i -esimo o j -esimo del sistema ad N elettroni. Questa rappresentazione consente di investigare le proprietà dei fenomeni di eccitazione carica $N \rightarrow N+1$ o $N \rightarrow N-1$, nelle quali il sistema acquisisce o perde un elettrone[10]. Infatti nell'equazione (1.3) compaiono le energie di quasiparticella $\epsilon_i = E^{N+1,i} - E^{N,0}$ ed $\epsilon_j = E^{N-1,j} - E^{N,0}$, che sono l'opposto delle energie di affinità elettronica e di ionizzazione, e i termini η_i che corrispondono agli inversi delle vite medie associate alle transizioni. Sfortunatamente un calcolo esatto di G è troppo difficile da svolgere su un computer classico, ma si può trovare una prima approssimazione della soluzione per mezzo della DFT, riconducendosi ad un sistema in cui le particelle non interagiscono tra loro ma solo con un campo di interazione medio. A questo sistema corrisponde un propagatore diverso da G , che quindi viene denotato G_0 , e anche da questo tramite l'equazione (1.3) si possono ottenere degli autovalori di energia ϵ_i^0 , ma questi autovalori non sono abbastanza precisi per lo studio delle proprietà di eccitazione carica.

Il propagatore G può essere ricavato a partire da G_0 attraverso un'espansione perturbativa che considera i contributi di tutte le espressioni descritte dai diagrammi di Feynman ammessi di cui alcuni esempi sono riportati in figura 1.1. Questa impostazione del problema è vantaggiosa perchè G_0 viene ottenuto dalla soluzione di un problema a singola particella ed è quindi relativamente facile da ricavare.

Al termine di questo procedimento si ottiene un risultato di grande importanza, l'equazione di Dyson:

$$G(1, 2) = G_0(1, 2) + \int d^3d^4 G_0(1, 3)\Sigma(3, 4)G(4, 2) \quad (1.4)$$

In questa equazione compare il termine Σ , la funzione self-energia. Siccome G_0 si può considerare noto, questa equazione permette di calcolare G se si conosce un'approssimazione della self-energia Σ .

1.2 GW

Il problema del calcolo di G è stato affrontato in modo rigoroso e sistematico da Lars Hedin [2], il quale ricavò un sistema di cinque equazioni integro-differenziali autoconsistenti che permettono in linea di principio di ricavare G . Queste equazioni sono estremamente difficili da risolvere in modo esatto e la self-energia ha un'espressione particolarmente complicata: $\Sigma = iGW\Gamma$, in cui il "vertice completo" $\Gamma(1, 2, 3)$ dipende da tre punti dello spazio-tempo. Per semplificare l'espressione si usa l'approssimazione $\Gamma(1, 2, 3) = \delta(1, 2)\delta(1, 3)$, con la quale [9] le equazioni di Hedin diventano

$$G(1, 2) = G_0(1, 2) + \int G_0(1, 3)\Sigma(3, 4)G(4, 2)d(3, 4) \quad (1.5)$$

$$\Gamma(1, 2, 3) = \delta(1, 2)\delta(1, 3) \quad (1.6)$$

$$\chi_0(1, 2) = -iG(1, 2)G(2, 1) \quad (1.7)$$

$$W(1, 2) = v(1, 2) + \int v(1, 3)\chi_0(3, 4)W(4, 2)d(3, 4) \quad (1.8)$$

$$\Sigma(1, 2) = iG(1, 2)W(1^+, 2) \quad (1.9)$$

L'equazione (1.5) è l'equazione di Dyson. Il termine χ_0 nell'equazione (1.7) indica la suscettibilità irriducibile. Nella (1.8) compaiono il potenziale coulombiano v e l'interazione schermata W , che tiene conto degli effetti di schermatura del potenziale coulombiano a causa degli elettroni circostanti. L'equazione (1.9) esprime la self-energia in funzione del propagatore G e dell'interazione schermata W ed è la ragione per cui questa approssimazione è nota come approssimazione GW .

Le equazioni GW sono ancora autoconsistenti e possono essere risolte con metodi iterativi ma più spesso si preferisce risolvere il problema nell'approssimazione G_0W_0 che corrisponde alla prima iterazione delle equazioni di Hedin in approssimazione GW; questo metodo è estremamente efficace e notevolmente più semplice della soluzione autoconsistente.

Si parte risolvendo un problema a singola particella definito da una hamiltoniana

$$\hat{h}^{\text{DFT}} = -\frac{1}{2}\nabla^2 + v_{\text{ext}} + v_H + v_{\sigma}^{\text{xc}} \quad (1.10)$$

in cui v_{ext} è il potenziale esterno, v_H il potenziale di Hartree e v_{σ}^{xc} è il potenziale di correlazione-scambio. Un esempio tipico è l'hamiltoniana di Kohn-Sham che è alla base della DFT. Dall'equazione di Dyson si ricava un'equazione agli autovalori nota come equazione di quasiparticella. Per la similarità tra questa equazione e l'equazione agli autovalori della DFT si può usare la teoria perturbativa per calcolare gli autovalori di quasiparticella a partire dalle autofunzioni $|\phi_i^0\rangle$ della DFT. Si ottiene così la seguente equazione autoconsistente unidimensionale, facile da risolvere con metodi iterativi, in cui per semplicità si scrive la self-energia $\Sigma_i(\epsilon_i)$ come somma della self-energia di correlazione $\Sigma_i^c(\epsilon_i)$, dipendente dall'energia, e della self-energia di scambio Σ_i^x , indipendente dall'energia.

$$\epsilon_i = \epsilon_i^0 + \langle \phi_i^0 | \Sigma_i^x - v^{\text{xc}} | \phi_i^0 \rangle + \langle \phi_i^0 | \Sigma_i^c(\epsilon_i) | \phi_i^0 \rangle \quad (1.11)$$

La difficoltà che rimane a questo punto è il calcolo della self-energia: infatti riscrivendo l'equazione (1.9) nel dominio delle frequenze si ottiene

$$\Sigma_{\sigma}(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int d\omega' e^{i\omega'\eta} G_0^{\sigma}(\mathbf{r}, \mathbf{r}', \omega + \omega') W_0(\mathbf{r}, \mathbf{r}', \omega') \quad (1.12)$$

in cui compare un integrale difficile da valutare a causa della presenza di molti poli lungo l'asse reale.

Capitolo 2

Calcolo numerico di Σ

2.1 Continuazione analitica

Per ottenere delle stime precise degli autovalori di energia con il metodo GW è necessario poter calcolare precisamente la self-energia attorno all'asse reale, ma come si è visto è molto difficile calcolare la funzione in questa regione con l'equazione (1.12). Invece sull'asse immaginario la self-energia è molto più regolare ed è facile da calcolare; una volta che la funzione è stata valutata sull'asse immaginario si può effettuare una continuazione analitica sull'asse reale. Tramite questa continuazione analitica è poi possibile valutare la funzione anche lungo l'asse reale. Il risultato di questo procedimento è rappresentato in figura 2.1.

Sono stati analizzati diversi metodi per la continuazione analitica. In ciascun caso, i dati di partenza sono i valori della self-energia sull'asse immaginario; sulla base di questi si ottiene un approssimante per la self-energia che viene impiegato per i calcoli GW. Si ottengono infine le energie di quasiparticella dell'equazione (1.11). I dati di partenza sono stati calcolati con il software *Quantum ESPRESSO*, un software open-source per calcoli di struttura elettronica e modellizzazione di materiali su scala nanometrica. *Quantum ESPRESSO* implementa anche la continuazione analitica della self-energia con il metodo dei multipoli e nelle successive sezioni i valori di energia ottenuti da *Quantum ESPRESSO* sono confrontati con quelli ottenuti indipendentemente da *Quantum-SEF* con il metodo dei multipoli o con altri metodi.

2.2 Metodo dei multipoli

Il metodo dei multipoli è il più semplice tra quelli analizzati in questo lavoro di tesi, ma nonostante la sua semplicità è molto efficace e permette inoltre di illustrare un approccio che è comune anche ad altri metodi. Così come il calcolo della correzione GW è stato ricondotto allo svolgimento di una continuazione analitica, quest'ultima viene ricondotta ad un problema di minimizzazione, che può essere affrontato con i numerosi metodi presenti in letteratura sviluppati per questo scopo. Nella procedura di minimizzazione si tiene conto di eventuali vincoli imposti da considerazioni sulla forma dell'approssimante o da considerazioni teoriche sulla soluzione che si vuole trovare. La soluzione di questo problema fornisce una funzione della forma cercata che approssima i dati della self-energia forniti.

Il problema della continuazione analitica è un problema inverso in quanto si tratta di ricostruire da un insieme di dati la funzione che li ha prodotti. Come spesso accade per problemi di questo tipo, il problema della continuazione analitica della self-energia è mal condizionato, ovvero è estremamente sensibile ad un errore nei dati iniziali ed ammette soluzioni molto diverse di cui solo alcune riproducono correttamente la self-energia sull'asse reale.

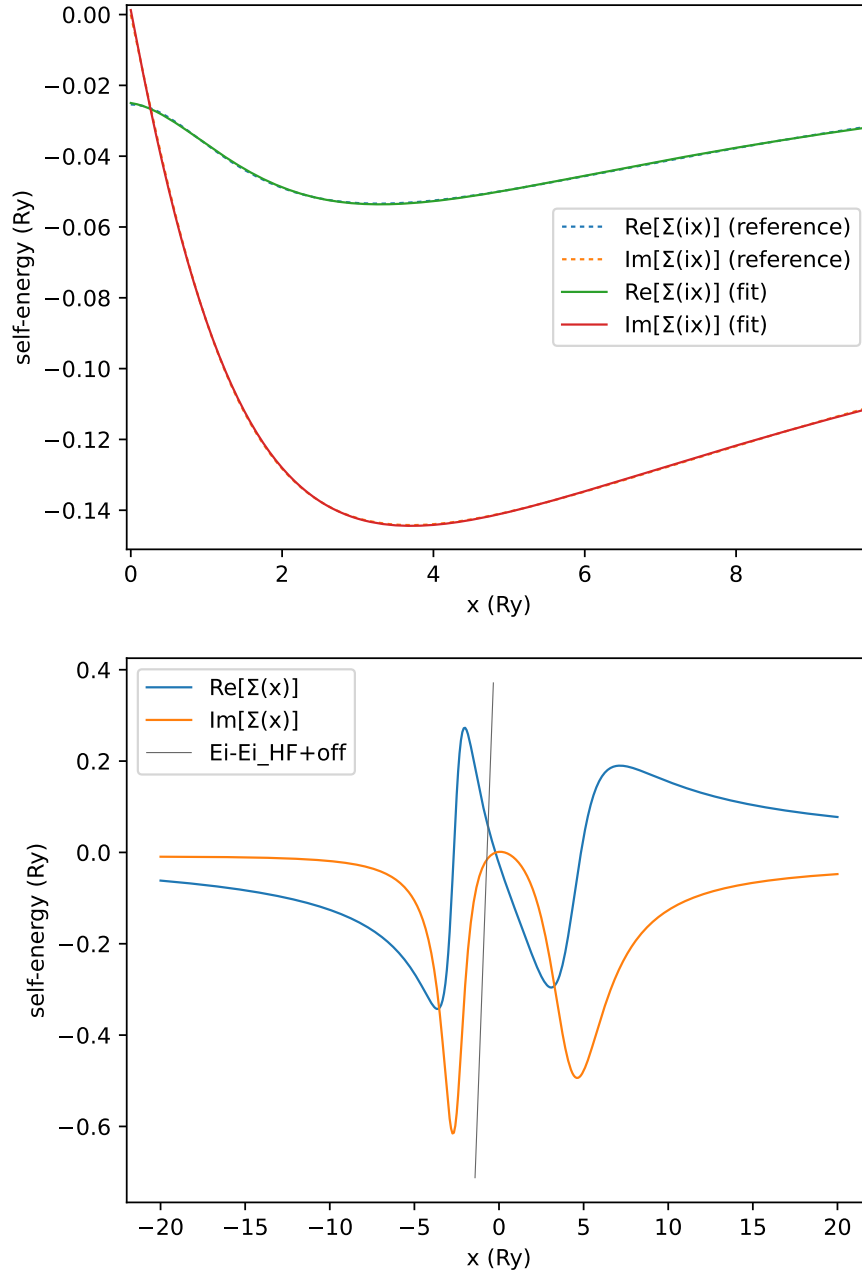


Figura 2.1: Continuazione analitica della self-energia con il metodo dei multipoli. Nel grafico in alto è rappresentata la self-energia lungo l'asse immaginario. I valori esatti calcolati da *Quantum ESPRESSO* sono descritti dalle curve tratteggiate, mentre le curve continue, sovrapposte alle precedenti, sono state calcolate con l'approssimante a multipoli. Qui la corrispondenza è perfetta. Nel grafico in basso è rappresentata la self-energia lungo l'asse reale. L'intersezione tra la retta diagonale e la self-energia individua la soluzione dell'equazione di quasiparticella. Una piccola imperfezione nel fit del grafico in alto è in grado di produrre comportamenti anomali nelle curve del grafico in basso.

2.2.1 Formulazione matematica

Il metodo dei multipoli per la continuazione analitica della self-energia prevede che la self-energia venga approssimata da una funzione del tipo

$$S(z) = a_0 + \sum_{i=1}^n \frac{a_i}{z - b_i} \quad (2.1)$$

La ragione di questa scelta per la funzione approssimante consegue dalla forma della self-energia, che è simile a quella di una funzione a multipoli.

Effettuare la continuazione analitica corrisponde a individuare i parametri a_i^*, b_i^* che meglio descrivono i valori noti sull'asse immaginario, ovvero minimizzano una funzione obiettivo che quantifica quanto l'approssimazione si distanzia dalla funzione ricercata. Sia θ il vettore dei parametri,

$$\theta = \begin{pmatrix} a_0 \\ a_1 \\ b_1 \\ a_2 \\ \vdots \\ b_n \end{pmatrix}$$

e sia

$$\mathbf{z} = \begin{pmatrix} z_0 \\ z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$$

il vettore costituito dai punti sull'asse immaginario in cui viene campionata la self-energia. Sia $S_\theta(z)$ la funzione a multipoli con i parametri dati da θ come indicato in (2.1). Siano \mathbf{s}_θ e σ i vettori definiti da $s_{\theta,i} = S_\theta(z_i)$, $\sigma_i = \Sigma(z_i)$. Si sceglie come funzione obiettivo la norma l^2

$$\lambda(\theta) = \|\mathbf{s}_\theta - \sigma\| \quad (2.2)$$

2.2.2 Minimizzazione della funzione obiettivo

Il problema della continuazione analitica col metodo dei multipoli è stato ricondotto ad un problema di minimizzazione. Lo spazio di ricerca per i parametri θ di un approssimante ad n poli ha dimensione $2n+1$ ed in generale può essere molto grande. Questo tipo di problema viene affrontato con algoritmi di minimizzazione al primo ordine che individuano un minimo locale della funzione obiettivo utilizzando il gradiente per determinare la direzione in cui è più efficiente effettuare un passo di minimizzazione. La ricerca del minimo globale è un problema in generale non risolvibile, ma con una scelta ragionevole delle condizioni iniziali l'algoritmo converge al minimo globale.

Il calcolo del gradiente della funzione $\lambda(\theta)$ è a sua volta un problema non banale; nel caso del problema a multipoli è possibile calcolare a mano il gradiente e specificarlo esplicitamente nella definizione dell'algoritmo, ma questo calcolo andrebbe ripetuto per ogni scelta di funzione approssimante e molto spesso non è immediato da calcolare (già per approssimanti di Padé inizia a diventare un calcolo eccessivamente lungo). Si può stimare il gradiente con metodi di differenziazione numerica alle differenze finite, ma questi risultano poco efficienti per spazi di ricerca di dimensione elevata. Per questo motivo si è scelto di affrontare il problema tramite tecniche di differenziazione automatica per mezzo della libreria JAX [6] [7].

Dopo aver definito funzioni per il calcolo di $S_\theta(z)$, $\Sigma(z)$ e $\lambda(\theta)$ si usa JAX per ottenere il gradiente di $\lambda(\theta)$ che si inserisce in un algoritmo di minimizzazione, come il metodo del gradiente coniugato o BFGS. Si ottiene così θ^* con il quale si ottiene finalmente l'approssimante a multipoli che può essere usato nel calcolo GW.

2.2.3 Risultati

Come prima cosa è stato necessario verificare che *Quantum-SEF* fosse in grado di processare i file prodotti da *Quantum ESPRESSO* e che implementasse correttamente i calcoli GW. Si è usato *Quantum ESPRESSO* per eseguire i calcoli GW per il metano; la cartella contenente i file corrispondenti è indicata con `$methane`. *Quantum ESPRESSO* esegue la continuazione analitica con un metodo a multipoli e salva risultati finali e alcuni risultati intermedi nel file `$methane/methane_gww.out`. Qui sono presenti anche i risultati del calcolo dell'approssimante ed è possibile estrarre i valori dei parametri trovati da *Quantum ESPRESSO*. Questi parametri sono stati usati per ricostruire lo stesso approssimante a multipoli con *Quantum-SEF* e sono stati eseguiti indipendentemente i calcoli GW successivi. I valori di energia calcolati *Quantum-SEF* risultano identici a quelli prodotti da *Quantum ESPRESSO*. Nelle righe seguenti sono riportati i risultati di questo calcolo, in cui i valori di energia sono misurati in elettronvolt.

Command: `qsef gwwparams $methane`

```
state 1; Quantum-SEF: -20.12104 DFT: -16.6124 Quantum ESPRESSO: -20.12104
state 2; Quantum-SEF: -13.58752 DFT: -9.11161 Quantum ESPRESSO: -13.58752
state 3; Quantum-SEF: -13.59307 DFT: -9.11161 Quantum ESPRESSO: -13.59307
state 4; Quantum-SEF: -13.58416 DFT: -9.11161 Quantum ESPRESSO: -13.58416
state 5; Quantum-SEF: 0.2109600 DFT: -0.56519 Quantum ESPRESSO: 0.2109600
```

Una volta verificato il corretto funzionamento del calcolo GW, si può esaminare il primo metodo di continuazione analitica, quello a multipoli. In questo caso i dati di partenza sono solamente i valori della self-energia salvati nei file `$methane/ch4-{re,im}_on_im0000{1..5}` e i parametri dell'approssimante vengono ricavati indipendentemente da *Quantum-SEF*. Si usa un approssimante con 2 poli.

Command: `qsef multipole 2 $methane`

```
state 1; Quantum-SEF: -21.39625 DFT: -16.6124 Quantum ESPRESSO: -20.12104
state 2; Quantum-SEF: -13.64765 DFT: -9.11161 Quantum ESPRESSO: -13.58752
state 3; Quantum-SEF: -13.64940 DFT: -9.11161 Quantum ESPRESSO: -13.59307
state 4; Quantum-SEF: -13.64649 DFT: -9.11161 Quantum ESPRESSO: -13.58416
state 5; Quantum-SEF: 0.2341317 DFT: -0.56519 Quantum ESPRESSO: 0.2109600
```

I risultati ottenuti sono in buon accordo con le stime di *Quantum-ESPRESSO* e le energie degli orbitali 2, 3, 4, 5 differiscono dai valori calcolati da *Quantum ESPRESSO* di meno di un decimo di elettronvolt.

2.3 Approssimanti di Padé

Una tecnica molto usata per la continuazione analitica è quella dell'approssimante di Padé, che consiste nell'interpolazione dei valori campionati dalla self-energia con un rapporto tra due polinomi. Il metodo di Padé si basa sull'ipotesi che $G(z)$ sia ben descritta da una frazione continua finita, della forma

$$G(z) = G_{r+1}(z) = \frac{\lambda_0^2}{z - e_0 - \frac{\lambda_1^2}{\ddots \frac{\lambda_r^2}{z - e_r}}} \quad (2.3)$$

Dal confronto tra (2.3) e (1.4) si osserva che anche $\Sigma(z)$ può essere scritta come

$$\Sigma(z) = \Sigma_r(z) = \frac{\lambda_1^2}{z - e_1 - \frac{\lambda_2^2}{\ddots \frac{\lambda_r^2}{z - e_r}}} \quad (2.4)$$

La funzione in (2.4) è esprimibile anche come funzione razionale, che motiva la scelta dell'approssimante di Padé:

$$\Sigma_r(z) = P_r(z) = \frac{\sum_{i=1}^{k+1} a_i z^{i-1}}{\sum_{i=1}^r b_i z^{i-1} + z^r} \quad (2.5)$$

La procedura più usata per determinare l'approssimante [3] di una funzione complessa $f(z)$ consiste nel risolvere $2r$ equazioni lineari, scritte in forma matriciale come

$$K\mathbf{v} = \mathbf{y} \quad (2.6)$$

dove

$$\mathbf{v} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_r \\ b_1 \\ \vdots \\ b_r \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} z_1^r f(z_1) \\ z_2^r f(z_2) \\ \vdots \\ z_M^r f(z_M) \end{bmatrix}$$

$$K = \begin{bmatrix} 1 & \cdots & z_1^{r-1} & -f(z_1) & \cdots & -f(z_1) z_1^{r-1} \\ 1 & \cdots & z_2^{r-1} & -f(z_2) & \cdots & -f(z_2) z_2^{r-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cdots & z_M^{r-1} & -f(z_M) & \cdots & -f(z_M) z_M^{r-1} \end{bmatrix}$$

e $M, r \in \mathbb{N}, N := 2r, M = N$.

La soluzione di questo sistema fornisce i coefficienti di un approssimante di Padé che passa per tutti i punti specificati: $P_r(z_i) = f(z_i)$; tuttavia a causa della condizione $M = N$ esso tende a non estrapolare correttamente il comportamento della funzione interpolata e risulta poco precisa in punti diversi dagli z_i , un problema noto come overfitting. Questo problema viene risolto scegliendo diversamente i parametri N e M in modo che $N \leq M$. In questa situazione però la matrice K di (2.6) non è invertibile e bisogna considerare l'equazione (2.6) come un problema di regressione lineare da risolvere col metodo dei minimi quadrati.

L'approssimante di Padé è estremamente efficace quando i dati a disposizione sulla self-energia sono molto precisi ma altrimenti diventa rapidamente meno affidabile e in molti casi produce soluzioni non fisicamente accettabili. Anche una piccola variazione nei dati di ingresso produce una grande variazione nei risultati e se si ripete il calcolo molte volte cambiando semplicemente i punti in cui viene campionata la self-energia si ottengono risultati che presentano variazioni casuali. L'idea alla base di un recente studio [4] è che un approssimante costruito facendo la media pesata di un insieme di diversi approssimanti di Padé sia molto più affidabile di ogni singolo approssimante di Padé perchè le oscillazioni casuali si compensano a vicenda. I pesi possono essere scelti in diversi modi, ad esempio ponendoli uguali a 0 o 1 in modo da selezionare solo un sottoinsieme di funzioni che sono in buon accordo tra loro.

2.3.1 Soluzione numerica

Si osserva che nella matrice compaiono termini della forma z_M^i con $i = 0, \dots, r-1$, che attraversano diversi ordini di grandezza. Eseguendo una decomposizione ai valori singolari si osserva che anche i valori singolari di K appartengono ad ordini di grandezza molto diversi e si osserva che il numero di condizionamento di K raggiunge spesso valori dell'ordine di 10^{10} .

Le difficoltà nella soluzione di questo problema sono principalmente due. Il numero di condizionamento evidenzia che il problema è mal condizionato, quindi la soluzione è estremamente sensibile ad errori su \mathbf{v} . Inoltre è necessario conservare in memoria numeri che differiscono di diversi ordini di grandezza. Per questi motivi è necessario rappresentare i numeri in *complex128* (coppie di *float64*). Questa scelta

permette di rappresentare correttamente tutti i valori e ridurre l'errore numerico dei calcoli intermedi ma rende il calcolo meno efficiente, specialmente su hardware specializzato come GPU e TPU, e corrisponde alla massima precisione numerica supportata da JAX. L'elevata precisione numerica usata nei calcoli intermedi non migliora invece la precisione dei dati iniziali; per questo ci si aspetta che i risultati finali presentino comunque una deviazione dalla soluzione dei minimi quadrati.

La soluzione del sistema 2.6 viene individuata usando l'algoritmo `scipy.linalg.lstsq` della libreria `scipy`. Per avvicinarsi ulteriormente al minimo della funzione di errore dei minimi quadrati si può raffinare la soluzione fornita da `scipy` con il metodo di minimizzazione descritto in 2.2.2 con JAX. Si definisce quindi la funzione $\lambda(\mathbf{x}) = \|\mathbf{b} - A\mathbf{x}\|$, si differenzia con JAX e si inserisce in un algoritmo di minimizzazione come BFGS. La procedura è analoga a quella utilizzata nel caso dei multipoli 2.2.2 ma in questo caso lo spazio di ricerca è molto più grande ($\dim \approx 100$) quindi è necessario sfruttare JAX per delegare il calcolo ad una GPU che è in grado di completare il calcolo più rapidamente. Si osserva che le soluzioni raffinate nel modo appena descritto sono leggermente migliori di quelle calcolate con `scipy` ma la differenza è piccola. Siccome la procedura di minimizzazione col gradiente è computazionalmente intensiva non è sempre stato possibile applicarla (e.g. i calcoli della seconda parte di 2.3.2)

La procedura per ottenere un approssimante di Padé può essere ripetuta variando i punti di campionamento della self-energia e i parametri N e M secondo quanto descritto in [4]. Dalla media pesata di questi si ottiene un approssimante migliore che viene denominato AveragePade.

2.3.2 Risultati

In modo analogo a quanto fatto per il metodo a multipoli, è stato applicato il metodo degli approssimanti di Padé per ottenere un'approssimazione AveragePade della self energia da usare per i calcoli GW. In questo caso il confronto con i risultati di *Quantum ESPRESSO* è meno utile perchè sono stati usati metodi diversi per stimare la self energia.

```
Command: qsef pade --type avgLS --precise_iter 500 $methane
```

```
state 1; Quantum-SEF: -21.18195 DFT: -16.6124 Quantum ESPRESSO: -20.12104
state 2; Quantum-SEF: -13.60915 DFT: -9.11161 Quantum ESPRESSO: -13.58752
state 3; Quantum-SEF: -13.61078 DFT: -9.11161 Quantum ESPRESSO: -13.59307
state 4; Quantum-SEF: -13.60795 DFT: -9.11161 Quantum ESPRESSO: -13.58416
state 5; Quantum-SEF: 0.2435491 DFT: -0.56519 Quantum ESPRESSO: 0.2109600
```

In 2.3 e 2.3.1 si è detto che un approssimante di Padé tende ad essere molto sensibile a variazioni nei dati di ingresso mentre ci si aspetta che AveragePade produca risultati più affidabili. Questo è stato verificato direttamente ripetendo i calcoli descritti in 2.3.1 per un insieme di dati leggermente diversi fino ad ottenere i risultati finali GW, dei quali è stata analizzata la variabilità.

I dati iniziali sono stati generati con una procedura inversa rispetto a quella di 2.3. A partire da un approssimante di Padé descritto dai parametri \mathbf{v} è stato costruito un insieme $\{\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^{500}\}$ in cui ogni elemento \mathbf{v}^i è stato ottenuto da \mathbf{v} introducendo un piccolo errore, in modo che l'elemento j -esimo v_j^i di \mathbf{v}^i sia dato da $v_j^i = v_j \cdot r_{i,j}$ con i $r_{i,j}$ tratti da una distribuzione normale con $\mu = 1$ e $\sigma = 0.01$. Per ognuno dei 500 \mathbf{v}_i è stato costruito il corrispondente approssimante di Padé che è stato usato per calcolare la self-energia lungo l'asse immaginario (analoghi a quelli forniti da *Quantum ESPRESSO*). Questi dati sono poi stati analizzati come in 2.3 ed è stato costruito un set di risultati rappresentato in figura 2.2. La figura rappresenta un istogramma dei valori di energia GW e la linea verticale indica il valore che corrisponde all'approssimante di partenza con parametri \mathbf{v} .

L'approssimante di partenza fornisce un risultato di $-13.628eV$. La distribuzione finale ha media $-13.637eV$ e deviazione standard $0.0171eV$. Dalla figura 2.2 si osserva che anche quando i dati di input sono soggetti ad un piccolo errore si ottengono risultati distribuiti nella vicinanza del risultato iniziale, quindi il metodo fornisce risultati coerenti nonostante la natura mal condizionata del problema.

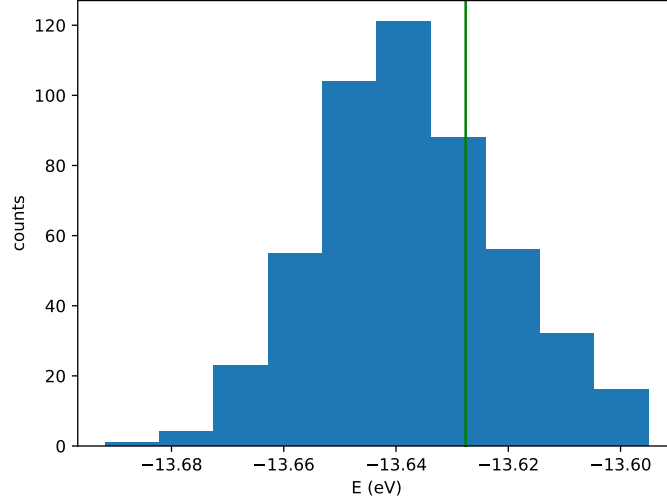


Figura 2.2: Istogramma delle energie GW

2.4 Reti neurali

I metodi descritti nelle sezioni precedenti sono stati sviluppati *ad hoc* per la continuazione analitica della self-energia sfruttando informazioni a priori (come la presenza di poli attorno all'asse reale) in modo da imporre vincoli per ridurre l'ambiguità sulla soluzione. Recentemente è stato descritto un metodo che non necessita di informazioni a priori basato sulle reti neurali[8], una tecnologia che è alla base dei più importanti successi nel campo dell'intelligenza artificiale negli ultimi anni.

2.4.1 Formulazione del problema

Si vuole risolvere un problema di regressione con una rete neurale in modo da stimare i parametri della rete neurale che permettono di inferire i valori di self-energia sull'asse reale (variabile dipendente) a partire da quelli sull'asse immaginario (variabile indipendente). Il processo attraverso il quale vengono stimati questi parametri è il processo di *training* e coinvolge il processamento di un elevato numero di coppie di dati che associano un valore della variabile indipendente con il corrispondente valore della variabile dipendente. L'insieme di tutte queste coppie costituisce il *training dataset*. Successivamente l'accuratezza del modello viene misurata su un altro *dataset* (*validation dataset*), confrontando le previsioni del modello con i valori corretti.

Il metodo delle reti neurali è stato usato sia per effettuare la continuazione analitica del propagatore sia per effettuare la continuazione analitica della self-energia. Il primo caso corrisponde a quanto descritto in letteratura[8] mentre il secondo permette di usare un'equazione come (1.11) per ottenere le energie di quasiparticella. Per ciascuno dei due casi viene costruito un dataset indipendente.

Dataset per G

Il *training dataset* è costituito da coppie che associano i valori del propagatore sull'asse immaginario con quelli calcolati poco sopra l'asse reale (funzione spettrale $\rho(\omega)$).

Mentre il passaggio dall'asse immaginario all'asse reale è un problema mal condizionato, il procedimento inverso è ben definito:

$$G(z) = \int_{-\infty}^{+\infty} d\omega \frac{1}{z - \omega} \rho(\omega) \quad (2.7)$$

Siccome si conosce qualitativamente la forma della funzione spettrale, è possibile generare il *dataset* costruendo diverse funzioni spettrali arbitrarie e generando i dati sull'asse immaginario con la formula (2.7).

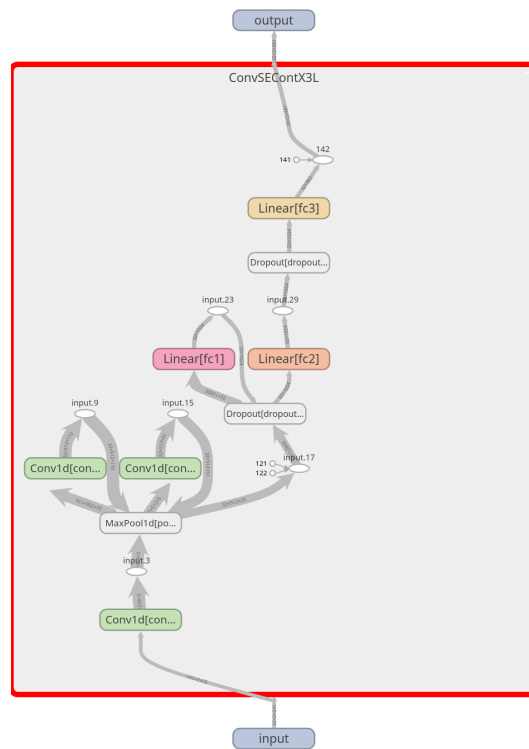


Figura 2.3: Architettura della rete neurale, costituita da tre layer convoluzionali e tre layer densi.

Per ogni elemento del dataset la funzione $\rho(\omega)$ viene generata sommando n_p gaussiane di altezza, larghezza e posizione casuale e viene normalizzata ad 1. Vengono generati 18000 elementi con $n_p = 1$, 18000 elementi con $n_p = 2$, 20000 elementi con $n_p = 3$, e 10000 elementi per ciascuno dei valori da 4 a 8. Siccome il dataset deve essere rappresentativo dei dati reali, che non sono privi di errore, si introduce sull'asse immaginario un rumore gaussiano con $\mu = 0$ e $\sigma = 0.01$. Per valutare l'accuratezza del modello si costruisce un *validation dataset* con le stesse caratteristiche del *training dataset* ma 100 volte più piccolo.

Dataset per Σ

Costruire un *dataset* di valori di funzioni self-energia è più difficile. Per ogni elemento del dataset è stata costruita una funzione a multipoli con due o quattro poli scegliendo parametri ragionevoli per una funzione self-energia, ad esempio parametri simili a quelli ottenuti in 2.2. Dalla funzione a multipoli sono stati estratti i valori relativi all'asse immaginario e a quello reale. Il procedimento è stato ripetuto fino ad ottenere 700000 elementi raccolti in un *dataset* da 4.1GiB.

Architettura della rete neurale

Per costruire la rete neurale è stata usata la libreria *PyTorch* usando *Tensorboard* per la visualizzazione dei dati di *training*. È stata usata una rete neurale convoluzionale come quella rappresentata in 2.3 con circa 17 milioni di parametri.

2.4.2 Risultati

Continuazione analitica di G

Al termine della procedura di *training* il modello risulta in grado di effettuare la continuazione analitica sull'asse reale; un esempio è rappresentato in figura 2.4.

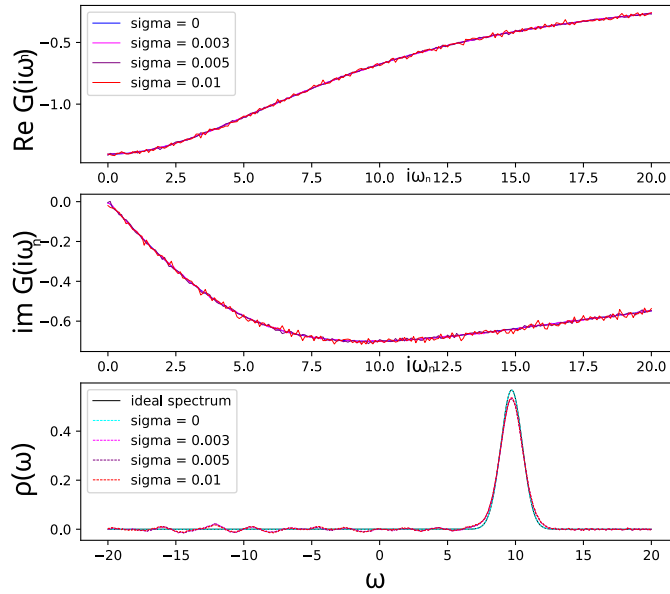


Figura 2.4: Continuazione analitica con reti neurali. Ad ogni livello di rumore (distribuzione gaussiana con $\sigma = 0, 0.003, 0.005, 0.010$) è associato un colore diverso. Nell'ultimo riquadro la curva esatta è rappresentata in nero, ed è sovrapposta con quella blu corrispondente all'assenza di rumore. Per livelli di rumore maggiori la previsione è meno precisa ma la posizione del picco viene comunque individuata correttamente.

Dalla figura si osserva che per dati molto precisi ($\sigma = 0$) il risultato è ottimo e la curva blu prevista dal modello si sovrappone esattamente a quella nera. In presenza di rumore ($\sigma = 0.003, 0.005, 0.010$) la qualità della previsione diminuisce, ma il modello è comunque in grado di individuare correttamente le posizioni dei picchi.

Continuazione analitica di Σ

La continuazione analitica prodotta dalla rete neurale per la self-energia, di cui un esempio è raffigurato in figura 2.5, è buona ma presenta alcuni difetti. La figura mostra che la rete neurale approssima bene i dati corretti perchè la curva blu è molto vicina a quella nera. I valori prodotti sono accettabili anche in presenza di piccole sorgenti di errore nei dati in ingresso, mentre non sono attendibili in presenza di sorgenti di errori più grandi (curva arancione).

Quando questa stima della self-energia viene usata per i calcoli GW si ottengono i risultati seguenti:

Command: `qsef neural $methane`

```
state 1; QSEF GW energy: -23.97507   DFT: -16.6124   Quantum ESPRESSO: -20.12104
state 2; QSEF GW energy: -9.938170  DFT: -9.11161   Quantum ESPRESSO: -13.58752
state 3; QSEF GW energy: -12.60911  DFT: -9.11161   Quantum ESPRESSO: -13.59307
state 4; QSEF GW energy: -10.33072  DFT: -9.11161   Quantum ESPRESSO: -13.58416
state 5; QSEF GW energy: 0.9978836   DFT: -0.56519   Quantum ESPRESSO: 0.2109600
```

Si ritiene che questi risultati siano poco attendibili perchè sono molto diversi da quelli ottenuti dai metodi precedenti, i quali invece tendono ad essere in accordo. Questa discrepanza è probabilmente dovuta al fatto che la popolazione di funzioni da cui è stato costruito il dataset non è rappresentativa dei veri dati di self-energia che si incontrano in calcoli come quelli per il metano. Di conseguenza ci si può aspettare che la rete neurale sia in grado di effettuare molto bene la continuazione analitica di funzioni a multipoli con due o quattro poli, come si vede in figura, mentre sia meno accurata in presenza di funzioni di tipo leggermente diverso, come per la self-energia. Questo problema non era presente nel caso della continuazione analitica di G perchè in quel caso il *dataset* era stato costruito

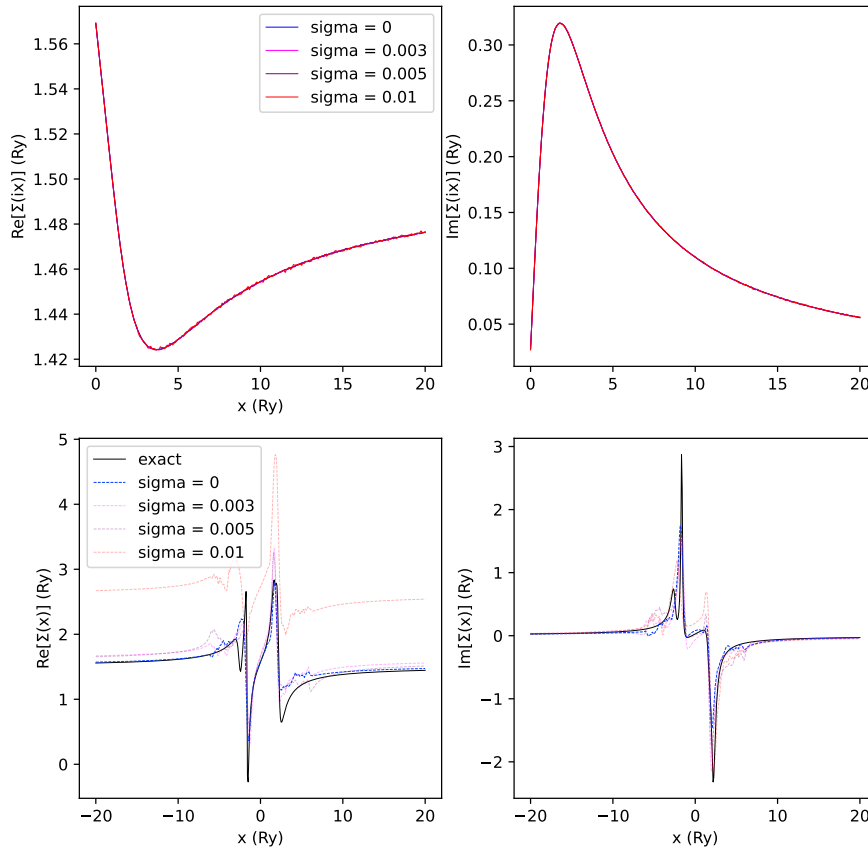


Figura 2.5: Continuazione analitica della self-energia con rete neurale. Nella prima riga sono raffigurati i valori di self-energia sull'asse immaginario: a sinistra parte reale, a destra parte immaginaria. Nella seconda riga i valori sono riferiti all'asse reale; la curva nera è la curva che la rete neurale deve approssimare, la curva blu è la curva prodotta dalla rete neurale a partire da dati esatti, mentre in altri colori sono raffigurate le curve prodotte dalla rete neurale in presenza di rumore nei dati di *input*.

utilizzando l'equazione esatta (2.7). Appare evidente che la difficoltà nell'utilizzo del metodo delle reti neurali possa risiedere nella costruzione di un dataset rappresentativo dei dati reali.

2.5 Conclusione

Il calcolo della self-energia sull'asse reale è uno dei passaggi più importanti che bisogna effettuare per applicare il metodo GW.

In questo lavoro di tesi si è affrontato il problema dal punto di vista della continuazione analitica, implementando alcuni metodi efficaci che sono stati descritti in letteratura e scrivendo il software *Quantum-SEF*, che è in grado di processare i dati prodotti da *Quantum ESPRESSO* sulla self-energia per svolgere i calcoli di continuazione analitica secondo i metodi descritti ed effettuare infine i calcoli GW.

È stato possibile ottenere buoni risultati con il metodo a multipoli e il metodo degli approssimanti di Padè, e nel caso del metodo a multipoli i risultati sono confrontabili con quelli ottenuti da *Quantum ESPRESSO* tramite lo stesso metodo. Gli stati 2, 3 e 4 formano un tripletto e le energie dovrebbero essere degeneri ma i calcoli GW non ne tengono conto a priori, quindi il fatto di ottenere valori simili suggerisce che la self-energia sia stata calcolata in modo accurato. In particolare le energie del tripletto calcolate da *Quantum-SEF*, sia con il metodo a multipoli che con il metodo AveragePade, variano di molto poco (circa $0.001eV$).

Il metodo delle reti neurali è il più sofisticato tra quelli descritti e ha il vantaggio di non necessitare di informazioni a priori sulla funzione. Esso è efficace per la continuazione analitica di G ed individua correttamente le posizioni dei picchi ma con livelli di rumore elevato produce caratteristiche non fisiche. La rete neurale è meno accurata quando viene usata per effettuare la continuazione analitica della self-energia, perchè non è stata sottoposta al *training* con un *dataset* rappresentativo dei dati sui quali viene usata nei calcoli GW. Nel caso della continuazione analitica della self-energia secondo la procedura descritta in sezione 2.4 il vantaggio di non necessitare di informazioni a priori sulla funzione viene annullato dal fatto che sia necessario fare assunzioni sulla forma della funzione per costruire il dataset.

Le reti neurali sono una tecnologia che ha avuto un successo recente: dal 2017, anno in cui si è diffusa la libreria Tensorflow, esse sono state alla base degli enormi progressi compiuti nel campo dell'intelligenza artificiale, mentre prima le risorse computazionali a disposizione erano insufficienti per la maggior parte delle applicazioni e la mancanza di librerie come Tensorflow rendevano più impegnativa la creazione e l'utilizzo di reti neurali. È ancora più recente l'impiego delle reti neurali nel campo della fisica: quanto è stato fatto in 2.4 si basa su un articolo pubblicato nel 2019. L'ultima parte di questo lavoro di tesi conferma che l'utilizzo delle reti neurali per la continuazione analitica della self-energia è molto promettente e possiede ancora molte potenzialità che non sono state esplorate. Anche se i risultati che sono stati ottenuti in sezione 2.4 non sono migliori di quelli ottenuti con i metodi precedenti, è possibile che uno studio più approfondito e l'utilizzo di architetture più sofisticate come i Transformer[5] sia in grado di effettuare la continuazione analitica della self-energia in modo più accurato di quanto già possibile con il metodo dei multipoli o quello degli approssimanti di Padè.

Bibliografia

- [1] Gordon Baym e N. David Mermin. «Determination of Thermodynamic Green's Functions». In: *Journal of Mathematical Physics* 2.2 (1961), pp. 232–234. DOI: 10.1063/1.1703704. eprint: <https://doi.org/10.1063/1.1703704>. URL: <https://doi.org/10.1063/1.1703704>.
- [2] Lars Hedin. «New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem». English. In: *Physical Review* 139.3A (1965), A796–A823. ISSN: 0031-899X.
- [3] K. S. D. Beach, R. J. Gooding e F. Marsiglio. «Reliable Padé analytical continuation method based on a high-accuracy symbolic computation algorithm». In: *Phys. Rev. B* 61 (8 2000), pp. 5147–5157. DOI: 10.1103/PhysRevB.61.5147. URL: <https://link.aps.org/doi/10.1103/PhysRevB.61.5147>.
- [4] Johan Schött et al. «Analytic continuation by averaging Padé approximants». In: *Phys. Rev. B* 93 (7 2016), p. 075104. DOI: 10.1103/PhysRevB.93.075104. URL: <https://link.aps.org/doi/10.1103/PhysRevB.93.075104>.
- [5] Ashish Vaswani et al. «Attention Is All You Need». In: (2017). DOI: 10.48550/ARXIV.1706.03762. URL: <https://arxiv.org/abs/1706.03762>.
- [6] James Bradbury et al. *JAX: composable transformations of Python+NumPy programs*. Ver. 0.3.13. 2018. URL: <http://github.com/google/jax>.
- [7] Matthew James Johnson Roy Frostig e Chris Leary. «Compiling machine learning programs via high-level tracing». In: *SysML* (2018).
- [8] Hongkee Yoon, Jae-Hoon Sim e Myung Joon Han. «Analytic continuation via domain knowledge free machine learning». In: *Phys. Rev. B* 98 (24 2018), p. 245101. DOI: 10.1103/PhysRevB.98.245101. URL: <https://link.aps.org/doi/10.1103/PhysRevB.98.245101>.
- [9] Dorothea Golze, Marc Dvorak e Patrick Rinke. «The GW Compendium: A Practical Guide to Theoretical Photoemission Spectroscopy». In: *Frontiers in Chemistry* 7 (lug. 2019). DOI: 10.3389/fchem.2019.00377.
- [10] Paolo Umari. «The G W Method for Excited States Calculations». In: *Heterogeneous Catalysts*. John Wiley & Sons, Ltd, 2021. Cap. 20, pp. 483–495. ISBN: 9783527813599. DOI: <https://doi.org/10.1002/9783527813599.ch27>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527813599.ch27>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527813599.ch27>.
- [11] Andrea Lizzit. *Quantum-SEF*. URL: https://github.com/andrea-lizzit/quantum_sef.
- [12] *Quantum-Espresso*. URL: <https://www.quantum-espresso.org/>.