

Fondamenti Teorici e Metodi Computazionali per la Regressione Ortogonale Rigorosa ed Error Analysis

Trattato di Analisi Statistica Sperimentale e Architettura del Codice `FitPipeline`

Laboratorio di Fisica Sperimentale e Computazionale

21 ottobre 2026

Sommario

Questo trattato fornisce una disamina teorica e computazionale completa della stima dei parametri in presenza di errori su entrambe le coordinate (modelli *Errors-in-Variables*). Nella prima parte si sintetizza l'architettura a tre stadi implementata in `FitPipeline`. Nella seconda parte viene sviluppata approfonditamente la teoria statistica: dal principio di massima verosimiglianza per variabili vettoriali alla formulazione della regressione ortogonale (ODR), dall'origine geometrica delle degenerazioni periodiche alle dimostrazioni asintotiche del bias da varianza mobile. Si analizza infine la convergenza asintotica della distribuzione chi-quadro, la geometria delle superfici di livello con HESSE e MINOS e la teoria quantitativa dei residui normalizzati (*pulls*).

Indice

1	Architettura e Metodi del Codice <code>chi_squared_fit.py</code>	2
1.1	Componenti Strutturali	2
1.1.1	Contenitori dei Risultati: <code>ParametroFit</code> e <code>RisultatiFit</code>	2
1.2	Strategie per la Funzione Obiettivo di Minuit	2
1.3	Pipeline a Tre Stadi e Ricerca Globale (<code>FitPipeline</code>)	3
2	Teoria Statistica Avanzata dei Minimi Quadrati e dell'ODR	4
2.1	Derivazione del Principio dei Minimi Quadrati dalla Verosimiglianza	4
2.2	Geometria della Proiezione Ortogonale e Degenerazioni Periodiche	5
2.2.1	Origine dell'Aliasing e dei Minimi Spuri	5
2.3	Dimostrazione Analitica del Bias da Varianza Effettiva	5
2.3.1	Perché il Metodo A Risolve il Teorema del Bias	7
2.4	Statistica Asintotica della Distribuzione χ^2 e Gradi di Libertà	7
2.5	Geometria delle Matrici di Errore: Curvatura (HESSE) vs Profilo (MINOS)	8
2.5.1	L'Approssimazione Parabolica Locale: HESSE	8
2.5.2	Il Principio del Profilo di Verosimiglianza: MINOS	8
2.6	Analisi dei Residui Normalizzati (<i>Pulls</i>) e Diagnostica di Consistenza	9
2.6.1	Definizione e Teoria dei Pulls	9
2.6.2	Proprietà Diagnostiche dei Momenti Statistici dei Pulls	9

1 Architettura e Metodi del Codice `chi_squared_fit.py`

L'obiettivo dell'infrastruttura software è automatizzare l'analisi di dati sperimentali bidimensionali garantendo tre proprietà fondamentali:

1. **Consistenza statistica:** assenza di bias da denominatori mobili nei minimi quadrati generalizzati;
2. **Invarianza globale:** capacità di evitare trappole di minimi locali spuri o sub-armoniche anche in presenza di parametri di primo tentativo (p_0) errati;
3. **Stima corretta delle incertezze:** estrazione di matrici di covarianza ben condizionate e calcolo di intervalli di confidenza asimmetrici tramite `iminuit`.

1.1 Componenti Strutturali

1.1.1 Contenitori dei Risultati: `ParametroFit` e `RisultatiFit`

L'incapsulamento dell'output separa rigorosamente le stime paraboliche simmetriche da quelle ottenute esplorando il profilo di verosimiglianza reale:

- **ParametroFit:** aggrega la stima centrale $\hat{\theta}_j$, l'incertezza simmetrica locale da matrice di curvatura $\sigma_{j,\text{hesse}} = \sqrt{2(\mathbf{H}^{-1})_{jj}}$, le deviazioni asimmetriche $\Delta_{\text{up}}, \Delta_{\text{down}}$ determinate dal solutore MINOS, e l'indice di asimmetria percentuale:

$$\alpha_j = \frac{\Delta_{\text{up}} - \Delta_{\text{down}}}{\frac{1}{2}(\Delta_{\text{up}} + \Delta_{\text{down}})} \times 100 \quad (1)$$

- **RisultatiFit:** aggrega i parametri stimati, il valore ottimo χ_{min}^2 , i gradi di libertà $\nu = N - p$, il chi-quadro ridotto $\chi_{\text{rid}}^2 = \chi^2/\nu$, il valore $p = \int_{\chi_{\text{min}}^2}^{\infty} f_{\chi^2}(z; \nu) dz$, il vettore dei residui r_i e la matrice di covarianza $p \times p$.

1.2 Strategie per la Funzione Obiettivo di Minuit

Minuit opera configurato con `errordef = Minuit.LEAST_SQUARES` ($\Delta\chi^2 = 1.0$ per la copertura a 1σ). Nel codice sono fornite due implementazioni alternative della funzione di costo:

1. `_CostoODR_linearizzato` (Varianza Effettiva Mobile):

$$\chi_{\text{eff}}^2(\boldsymbol{\theta}) = \sum_{i=1}^N \frac{(y_i - f(x_i; \boldsymbol{\theta}))^2}{\sigma_{y,i}^2 + \left(\frac{\partial f}{\partial x} \Big|_{x_i; \boldsymbol{\theta}} \sigma_{x,i} \right)^2} \quad (2)$$

Le derivate numeriche $\frac{\partial f}{\partial x}$ vengono ricalcolate a ogni step dell'ottimizzatore. Come dimostrato nella Sezione 2.3, questa dipendenza di σ_{eff} da $\boldsymbol{\theta}$ sposta sistematicamente il minimo verso pendenze maggiori.

2. `_CostoODR` (Metodo A a Varianza Congelata - Standard di Laboratorio):

$$\sigma_{\text{tot},i}^2 = \sigma_{y,i}^2 + \left(\frac{\partial f}{\partial x} \Big|_{x_i; \mathbf{p}_{\text{ref}}} \sigma_{x,i} \right)^2 = \text{costante} \quad \forall i \quad (3)$$

I pesi vengono pre-calcolati una sola volta sul vettore $\mathbf{p}_{\text{ref}}$ restituito dalla stima preliminare con ODRPACK nativo. Il funzionale:

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^N \frac{(y_i - f(x_i; \boldsymbol{\theta}))^2}{\sigma_{\text{tot},i}^2} \quad (4)$$

possiede denominatori statici: il gradiente è analiticamente non distorto e l'Hessiana è rigorosamente simmetrica e definita positiva vicino al minimo.

1.3 Pipeline a Tre Stadi e Ricerca Globale (FitPipeline)

La classe `FitPipeline` esegue un flusso a tre barriere numeriche:

1. **Pre-fit Locale:** esegue `scipy.odr.ODR` con i valori di p_0 forniti dall'utente.
2. **Controllo di Qualità e Fallback Stocastico:** valuta il chi-quadro ridotto grezzo del pre-fit. Se $\chi^2_{\text{rid}} > 8.0$, l'algoritmo rileva un arresto su un falso minimo o sub-armonica e scatena `_ricerca_globale_fallback`. Questo metodo applica l'algoritmo di *Dual Annealing* su uno spazio delimitato da bordi fisici adattivi, combinando l'esplorazione Monte Carlo a raffreddamento simulato con una discesa locale di tipo Nelder-Mead/BFGS.
3. **Inizializzazione e Analisi Minuit:** i parametri non distorti ottenuti dall'ODR finale vengono usati come ancoraggio statico per `_CostoODR`. Minuit esegue `migrad()`, calcola la curvatura esatta con `hesse()` ed estrae gli errori asimmetrici con `minos()`.

2 Teoria Statistica Avanzata dei Minimi Quadrati e dell'ODR

2.1 Derivazione del Principio dei Minimi Quadrati dalla Verosimiglianza

Consideriamo un processo di misura in cui sia la variabile indipendente sia la variabile dipendente sono grandezze fisiche stocastiche. Sia (x_i^*, y_i^*) lo stato vero della natura per l' i -esima osservazione, vincolato a giacere esattamente sulla varietà unidimensionale definita dal modello teorico:

$$y_i^* = f(x_i^*; \theta^*) \quad (5)$$

dove $\theta^* \in \mathbb{R}^p$ è il vettore dei parametri fisici veri. L'apparato sperimentale genera una coppia osservata (x_i, y_i) affetta da errori casuali additivi:

$$x_i = x_i^* + \delta_i, \quad y_i = y_i^* + \epsilon_i \quad (6)$$

Sotto l'ipotesi che gli errori sperimentali siano governati da processi stocastici gaussiani a media nulla, non correlati tra punti distinti e non correlati tra x e y sullo stesso punto, il vettore degli scarti $\boldsymbol{\eta}_i = (\delta_i, \epsilon_i)^T$ possiede matrice di covarianza diagonale:

$$\mathbf{V}_i = \begin{pmatrix} \sigma_{x,i}^2 & 0 \\ 0 & \sigma_{y,i}^2 \end{pmatrix} \quad (7)$$

La densità di probabilità congiunta di osservare la misura (x_i, y_i) dato il punto vero $(x_i^*, f(x_i^*; \theta))$ è data da:

$$P(x_i, y_i | x_i^*, \theta) = \frac{1}{2\pi\sigma_{x,i}\sigma_{y,i}} \exp \left(-\frac{1}{2} \left[\left(\frac{x_i - x_i^*}{\sigma_{x,i}} \right)^2 + \left(\frac{y_i - f(x_i^*; \theta)}{\sigma_{y,i}} \right)^2 \right] \right) \quad (8)$$

Data l'indipendenza statistica delle N misure, la funzione di verosimiglianza globale (Likelihood) per l'intero campione vale:

$$L(\theta, \mathbf{x}^*) = \prod_{i=1}^N P(x_i, y_i | x_i^*, \theta) = \left(\prod_{i=1}^N \frac{1}{2\pi\sigma_{x,i}\sigma_{y,i}} \right) \exp \left(-\frac{1}{2} \sum_{i=1}^N \left[\left(\frac{x_i - x_i^*}{\sigma_{x,i}} \right)^2 + \left(\frac{y_i - f(x_i^*; \theta)}{\sigma_{y,i}} \right)^2 \right] \right) \quad (9)$$

Il principio di massima verosimiglianza richiede di massimizzare L simultaneamente rispetto ai parametri fisici di interesse θ e agli N parametri di disturbo $\mathbf{x}^* = (x_1^*, \dots, x_N^*)$. Definendo la log-verosimiglianza negativa $\mathcal{L} = -\ln L$:

$$\mathcal{L}(\theta, \mathbf{x}^*) = \frac{1}{2} \sum_{i=1}^N \left[\left(\frac{x_i - x_i^*}{\sigma_{x,i}} \right)^2 + \left(\frac{y_i - f(x_i^*; \theta)}{\sigma_{y,i}} \right)^2 \right] + \sum_{i=1}^N \ln(2\pi\sigma_{x,i}\sigma_{y,i}) \quad (10)$$

Poiché il secondo termine è una costante dipendente unicamente dalle precisioni strumentali note a priori, la massimizzazione di L equivale rigorosamente alla minimizzazione del funzionale di costo quadratico bidimensionale:

$$S(\theta, \mathbf{x}^*) = \sum_{i=1}^N \left[\left(\frac{x_i - x_i^*}{\sigma_{x,i}} \right)^2 + \left(\frac{y_i - f(x_i^*; \theta)}{\sigma_{y,i}} \right)^2 \right] \quad (11)$$

Se poniamo $\delta_i = x_i^* - x_i$, l'equazione (11) assume la forma canonica dell'*Orthogonal Distance Regression* (ODR):

$$S(\theta, \boldsymbol{\delta}) = \sum_{i=1}^N \left[\left(\frac{\delta_i}{\sigma_{x,i}} \right)^2 + \left(\frac{y_i - f(x_i + \delta_i; \theta)}{\sigma_{y,i}} \right)^2 \right] \quad (12)$$

Nel caso limite in cui l'incertezza sulla coordinata indipendente sia trascurabile rispetto alla dinamica del problema ($\sigma_{x,i} \rightarrow 0$), la condizione di consistenza fisica impone che $x_i^* \rightarrow x_i$, ossia $\delta_i \rightarrow 0$. Il funzionale collassa naturalmente sui minimi quadrati verticali ordinari (OLS):

$$\lim_{\sigma_{x,i} \rightarrow 0} S(\theta, \boldsymbol{\delta}) = \chi_{\text{OLS}}^2(\theta) = \sum_{i=1}^N \left(\frac{y_i - f(x_i; \theta)}{\sigma_{y,i}} \right)^2 \quad (13)$$

2.2 Geometria della Proiezione Ortogonale e Degenerazioni Periodiche

Per un vettore di parametri θ assegnato, il minimo rispetto a ciascun parametro di disturbo δ_i si determina imponendo la stazionarietà della derivata parziale unidimensionale:

$$\frac{\partial}{\partial \delta_i} \left[\left(\frac{\delta_i}{\sigma_{x,i}} \right)^2 + \left(\frac{y_i - f(x_i + \delta_i; \theta)}{\sigma_{y,i}} \right)^2 \right] = 0 \quad (14)$$

Eseguendo la differenziazione:

$$\frac{2\delta_i}{\sigma_{x,i}^2} - \frac{2[y_i - f(x_i + \delta_i; \theta)]}{\sigma_{y,i}^2} \frac{\partial f}{\partial x} \Big|_{x_i + \delta_i} = 0 \quad (15)$$

Riorganizzando i termini, si ricava la **condizione geometrica fondamentale di ortogonalità pesata**:

$$\frac{y_i - f(x_i + \delta_i; \theta)}{\delta_i} = - \frac{\sigma_{y,i}^2}{\sigma_{x,i}^2} \frac{1}{\frac{\partial f}{\partial x} \Big|_{x_i + \delta_i}} \quad (16)$$

L'equazione (16) stabilisce che nel piano riscalato delle coordinate adimensionali:

$$u = \frac{x}{\sigma_x}, \quad v = \frac{y}{\sigma_y} \quad (17)$$

il vettore congiungente il dato sperimentale (u_i, v_i) con il punto ottimo sulla curva (u_i^*, v_i^*) è rigorosamente perpendicolare al vettore tangente alla curva stessa.

2.2.1 Origine dell'Aliasing e dei Minimi Spuri

Quando il modello $f(x; \theta)$ è periodico (es. $f(x) = A \sin(\omega x + \phi)$), la derivata prima $\partial f / \partial x$ oscilla indefinitamente tra $-A\omega$ e $+A\omega$. L'equazione (16) cessa di ammettere una soluzione unica e diventa multi-modale, presentando infiniti zeri locali lungo l'asse x . Se l'algoritmo di ottimizzazione sposta il parametro di frequenza $\omega \rightarrow \infty$, il periodo spaziale dell'onda $T = 2\pi/\omega$ collassa a valori inferiori alla risoluzione sperimentale σ_x . In tale limite asintotico:

$$\forall (x_i, y_i) \quad \exists \delta_i \text{ con } |\delta_i| \leq \frac{\pi}{\omega} \rightarrow 0 \quad \text{tale che} \quad f(x_i + \delta_i) = y_i \quad (18)$$

Di conseguenza, il residuo verticale si annulla identicamente ($y_i - f = 0$) con scostamenti δ_i arbitrariamente piccoli, portando il costo globale $S \rightarrow 0$ ($\chi_{\text{rid}}^2 \rightarrow 0$). Si genera così una buca di potenziale infinitamente profonda (*minimo degenerare ad alta frequenza*) che attrae i solutori privi di regolarizzazione. Per questa ragione, qualunque regressione ortogonale su oscillatori non lineari richiede la definizione di confini fisici (**limits**) su ω o una scansione globale preventiva.

2.3 Dimostrazione Analitica del Bias da Varianza Effettiva

Dimostriamo in modo formale perché il metodo euristico basato sulla varianza effettiva mobile genera stime asintoticamente inconsistenti e distorte dei parametri. Consideriamo la funzione di costo linearizzata:

$$\chi_{\text{lin}}^2(\theta) = \sum_{i=1}^N \frac{r_i^2(\theta)}{\sigma_{\text{eff},i}^2(\theta)} \quad (19)$$

dove definiamo il residuo verticale $r_i(\theta) = y_i - f(x_i; \theta)$ e la varianza efficace propagata al primo ordine:

$$\sigma_{\text{eff},i}^2(\theta) = \sigma_{y,i}^2 + (f'_i(\theta) \sigma_{x,i})^2, \quad \text{avendo posto } f'_i(\theta) \equiv \frac{\partial f}{\partial x} \Big|_{x_i; \theta} \quad (20)$$

Calcoliamo la derivata parziale esatta di χ_{lin}^2 rispetto a una componente θ_j del vettore dei parametri:

$$\begin{aligned}\frac{\partial \chi_{\text{lin}}^2}{\partial \theta_j} &= \sum_{i=1}^N \frac{\partial}{\partial \theta_j} \left[r_i^2(\boldsymbol{\theta}) \cdot (\sigma_{\text{eff},i}^2(\boldsymbol{\theta}))^{-1} \right] \\ &= \sum_{i=1}^N \left[2r_i \frac{\partial r_i}{\partial \theta_j} \frac{1}{\sigma_{\text{eff},i}^2} - \frac{r_i^2}{(\sigma_{\text{eff},i}^2)^2} \frac{\partial \sigma_{\text{eff},i}^2}{\partial \theta_j} \right]\end{aligned}\quad (21)$$

Poiché $\frac{\partial r_i}{\partial \theta_j} = -\frac{\partial f(x_i; \boldsymbol{\theta})}{\partial \theta_j}$ e differenziando la varianza:

$$\frac{\partial \sigma_{\text{eff},i}^2}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \left[\sigma_{y,i}^2 + (f'_i(\boldsymbol{\theta}))^2 \sigma_{x,i}^2 \right] = 2f'_i(\boldsymbol{\theta}) \frac{\partial f'_i(\boldsymbol{\theta})}{\partial \theta_j} \sigma_{x,i}^2 \quad (22)$$

otteniamo l'espressione rigorosa del gradiente:

$$\frac{\partial \chi_{\text{lin}}^2}{\partial \theta_j} = \underbrace{-2 \sum_{i=1}^N \frac{y_i - f(x_i; \boldsymbol{\theta})}{\sigma_{\text{eff},i}^2(\boldsymbol{\theta})} \frac{\partial f(x_i; \boldsymbol{\theta})}{\partial \theta_j}}_{\text{Termine Fisico di Discesa } G_j(\boldsymbol{\theta})} - \underbrace{2 \sum_{i=1}^N \frac{(y_i - f(x_i; \boldsymbol{\theta}))^2}{(\sigma_{\text{eff},i}^2(\boldsymbol{\theta}))^2} \left(f'_i(\boldsymbol{\theta}) \frac{\partial f'_i(\boldsymbol{\theta})}{\partial \theta_j} \sigma_{x,i}^2 \right)}_{\text{Termine di Bias Spuro } B_j(\boldsymbol{\theta})} \quad (23)$$

Nel punto di minimo stimato $\hat{\boldsymbol{\theta}}$, l'algoritmo impone l'annullamento del gradiente globale:

$$\left. \frac{\partial \chi_{\text{lin}}^2}{\partial \theta_j} \right|_{\hat{\boldsymbol{\theta}}} = 0 \implies G_j(\hat{\boldsymbol{\theta}}) = -B_j(\hat{\boldsymbol{\theta}}) \quad (24)$$

Calcoliamo ora il valore atteso statistico del termine B_j sul collettivo delle realizzazioni sperimentali, valutato attorno ai parametri veri $\boldsymbol{\theta}^*$. Dalla definizione di varianza totale, il valore atteso del quadrato del residuo è:

$$\mathbb{E} [(y_i - f(x_i; \boldsymbol{\theta}^*))^2] = \sigma_{\text{eff},i}^2(\boldsymbol{\theta}^*) \quad (25)$$

Sostituendo nel valore atteso del termine B_j :

$$\begin{aligned}\mathbb{E} [B_j(\boldsymbol{\theta}^*)] &= -2 \sum_{i=1}^N \frac{\mathbb{E} [(y_i - f(x_i; \boldsymbol{\theta}^*))^2]}{(\sigma_{\text{eff},i}^2(\boldsymbol{\theta}^*))^2} \left(f'_i(\boldsymbol{\theta}^*) \frac{\partial f'_i(\boldsymbol{\theta}^*)}{\partial \theta_j} \sigma_{x,i}^2 \right) \\ &= -2 \sum_{i=1}^N \frac{\sigma_{\text{eff},i}^2(\boldsymbol{\theta}^*)}{(\sigma_{\text{eff},i}^2(\boldsymbol{\theta}^*))^2} \left(f'_i(\boldsymbol{\theta}^*) \frac{\partial f'_i(\boldsymbol{\theta}^*)}{\partial \theta_j} \sigma_{x,i}^2 \right) \\ &= - \sum_{i=1}^N \frac{1}{\sigma_{\text{eff},i}^2(\boldsymbol{\theta}^*)} \frac{\partial \sigma_{\text{eff},i}^2(\boldsymbol{\theta}^*)}{\partial \theta_j} = - \frac{\partial}{\partial \theta_j} \left(\sum_{i=1}^N \ln \sigma_{\text{eff},i}^2(\boldsymbol{\theta}^*) \right) \neq 0\end{aligned}\quad (26)$$

Il valore atteso del termine spuro **non è nullo**. Al contrario, per il termine canonico G_j , essendo $\mathbb{E}[y_i - f(x_i; \boldsymbol{\theta}^*)] = 0$:

$$\mathbb{E} [G_j(\boldsymbol{\theta}^*)] = 0 \quad (27)$$

Di conseguenza, al valore vero dei parametri:

$$\mathbb{E} \left[\left. \frac{\partial \chi_{\text{lin}}^2}{\partial \theta_j} \right|_{\boldsymbol{\theta}^*} \right] = -\mathbb{E} [B_j(\boldsymbol{\theta}^*)] \neq 0 \quad (28)$$

La condizione di stazionarietà non è verificata sul parametro vero: lo stimatore $\hat{\boldsymbol{\theta}}$ ricavato minimizzando χ_{lin}^2 è **intrinsecamente distorto (biased)**. L'ottimizzatore viene forzato a deviare verso regioni dello spazio parametrico in cui $\sigma_{\text{eff},i}^2$ è più grande (per rendere B_j positivo), preferendo curve con pendenze artificialmente ripide pur di dilatare i denominatori.

2.3.1 Perché il Metodo A Risolve il Teorema del Bias

Nel Metodo A, la derivata spaziale viene calcolata numericamente sul vettore $\mathbf{p}_{\text{ref}}$ scaturito dalla stima preliminare e congelata come costante numerica:

$$\sigma_{\text{tot},i}^2 \equiv \sigma_{y,i}^2 + \left(\frac{\partial f}{\partial x} \bigg|_{x_i; \mathbf{p}_{\text{ref}}} \sigma_{x,i} \right)^2 \implies \frac{\partial \sigma_{\text{tot},i}^2}{\partial \theta_j} \equiv 0 \quad \forall j \quad (29)$$

Ne consegue identicamente:

$$B_j(\boldsymbol{\theta}) \equiv 0 \implies \mathbb{E} \left[\frac{\partial \chi_{\text{Metodo A}}^2}{\partial \theta_j} \bigg|_{\boldsymbol{\theta}^*} \right] = 0 \quad (30)$$

L'annullamento identico del gradiente di scala elimina la distorsione asintotica, garantendo stime non distorte ed efficienti al limite di grandi campioni.

2.4 Statistica Asintotica della Distribuzione χ^2 e Gradi di Libertà

Sia $\hat{\boldsymbol{\theta}}$ lo stimatore di massima verosimiglianza ricavato su un campione di N punti con p parametri liberi. Espandendo in serie di Taylor la funzione $f(x_i; \boldsymbol{\theta})$ attorno al vettore vero $\boldsymbol{\theta}^*$:

$$f(x_i; \boldsymbol{\theta}) \approx f(x_i; \boldsymbol{\theta}^*) + \sum_{j=1}^p J_{ij}(\theta_j - \theta_j^*), \quad J_{ij} = \frac{\partial f(x_i; \boldsymbol{\theta})}{\partial \theta_j} \bigg|_{\boldsymbol{\theta}^*} \quad (31)$$

In forma matriciale vettoriale, introducendo la matrice Jacobiana $\mathbf{J} \in \mathbb{R}^{N \times p}$ e la matrice diagonale dei pesi $\mathbf{W} \in \mathbb{R}^{N \times N}$ con $W_{ii} = 1/\sigma_{\text{tot},i}^2$, il vettore dei residui linearizzato vale:

$$\mathbf{r}(\boldsymbol{\theta}) = \mathbf{y} - \mathbf{f}(\boldsymbol{\theta}) \approx \Delta \mathbf{y} - \mathbf{J} \Delta \boldsymbol{\theta} \quad (32)$$

dove $\Delta \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{W}^{-1})$. La minimizzazione quadratica:

$$\chi^2(\boldsymbol{\theta}) = \mathbf{r}^T \mathbf{W} \mathbf{r} = (\Delta \mathbf{y} - \mathbf{J} \Delta \boldsymbol{\theta})^T \mathbf{W} (\Delta \mathbf{y} - \mathbf{J} \Delta \boldsymbol{\theta}) \quad (33)$$

fornisce la soluzione standard per i parametri ottimi:

$$\Delta \hat{\boldsymbol{\theta}} = (\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{W} \Delta \mathbf{y} \quad (34)$$

Sostituendo $\Delta \hat{\boldsymbol{\theta}}$ all'interno del residuo ottimo:

$$\hat{\mathbf{r}} = \Delta \mathbf{y} - \mathbf{J} \Delta \hat{\boldsymbol{\theta}} = [\mathbf{I}_N - \mathbf{J}(\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{W}] \Delta \mathbf{y} = (\mathbf{I}_N - \mathbf{P}) \Delta \mathbf{y} \quad (35)$$

dove $\mathbf{P} = \mathbf{J}(\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{W}$ è la matrice di proiezione ortogonale nello spazio pesato (*Hat Matrix*). Calcoliamo il valore ottimo della forma quadratica:

$$\chi_{\min}^2 = \hat{\mathbf{r}}^T \mathbf{W} \hat{\mathbf{r}} = \Delta \mathbf{y}^T (\mathbf{I}_N - \mathbf{P})^T \mathbf{W} (\mathbf{I}_N - \mathbf{P}) \Delta \mathbf{y} \quad (36)$$

Poiché $\mathbf{P} \mathbf{J} = \mathbf{J}$, si dimostra per sostituzione algebrica diretta che:

$$(\mathbf{I}_N - \mathbf{P})^T \mathbf{W} (\mathbf{I}_N - \mathbf{P}) = \mathbf{W} (\mathbf{I}_N - \mathbf{P}) \quad (37)$$

Definendo il vettore standardizzato $\mathbf{z} = \mathbf{W}^{1/2} \Delta \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$, la forma quadratica diventa:

$$\chi_{\min}^2 = \mathbf{z}^T \left[\mathbf{I}_N - \mathbf{W}^{1/2} \mathbf{J} (\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{W}^{1/2} \right] \mathbf{z} = \mathbf{z}^T \mathbf{M} \mathbf{z} \quad (38)$$

La matrice $\mathbf{M}$ è simmetrica e idempotente ($\mathbf{M}^2 = \mathbf{M}$). In virtù del **Teorema di Cochran**, una forma quadratica $\mathbf{z}^T \mathbf{M} \mathbf{z}$ di variabili normali standard segue esattamente una distribuzione chi-quadro il cui numero di gradi di libertà è pari alla traccia di $\mathbf{M}$:

$$\begin{aligned} \nu = \text{tr}(\mathbf{M}) &= \text{tr}(\mathbf{I}_N) - \text{tr} \left(\mathbf{W}^{1/2} \mathbf{J} (\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{W}^{1/2} \right) \\ &= N - \text{tr} \left((\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} (\mathbf{J}^T \mathbf{W} \mathbf{J}) \right) = N - \text{tr}(\mathbf{I}_p) = N - p \end{aligned} \quad (39)$$

Resta così dimostrato che la riduzione dei gradi di libertà da N a $N - p$ deriva direttamente dal rango della proiezione sui vincoli imposti dai parametri stimati.

2.5 Geometria delle Matrici di Errore: Curvatura (HESSE) vs Profilo (MINOS)

2.5.1 L'Approssimazione Parabolica Locale: HESSE

Attorno al punto di stazionarietà $\hat{\boldsymbol{\theta}}$, lo sviluppo di Taylor del costo è privo di termini lineari:

$$\chi^2(\boldsymbol{\theta}) = \chi^2(\hat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{H}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \mathcal{O}(\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^3) \quad (40)$$

dove l'Hessiana $\mathbf{H} \in \mathbb{R}^{p \times p}$ definisce l'operatore di curvatura locale:

$$H_{jk} = \left. \frac{\partial^2 \chi^2}{\partial \theta_j \partial \theta_k} \right|_{\hat{\boldsymbol{\theta}}} \quad (41)$$

In virtù del limite di Cramér-Rao, l'inversa della matrice di curvatura fornisce una stima asintoticamente efficiente della matrice di covarianza $\mathbf{V}$:

$$\mathbf{V} = \text{cov}(\hat{\boldsymbol{\theta}}) = 2 \mathbf{H}^{-1} \quad (42)$$

I coefficienti di correlazione lineare tra le stime dei parametri si ricavano dalla normalizzazione degli elementi fuori diagonale:

$$\rho_{jk} = \frac{V_{jk}}{\sqrt{V_{jj}V_{kk}}} \in [-1, +1] \quad (43)$$

L'ellissoide di confidenza a 1σ è definito dal luogo dei punti:

$$\Delta\chi^2(\boldsymbol{\theta}) = \chi^2(\boldsymbol{\theta}) - \chi^2(\hat{\boldsymbol{\theta}}) = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{V}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq 1.0 \quad (44)$$

Se proiettiamo questo ellissoide lungo l'asse del singolo parametro θ_j , la semi-ampiezza dell'intervallo coincide con l'errore standard HESSE:

$$\sigma_{j,\text{hesse}} = \sqrt{V_{jj}} \quad (45)$$

2.5.2 Il Principio del Profilo di Verosimiglianza: MINOS

Nel momento in cui il modello presenta forti non-linearità parametriche, i termini del terzo ordine $\mathcal{O}(\|\Delta\boldsymbol{\theta}\|^3)$ cessano di essere trascurabili: le superfici equipotenziali di χ^2 si deformano assumendo geometrie a pera, a banana o a sella, invalidando l'ipotesi di confidenza simmetrica. Il metodo MINOS supera l'approssimazione parabolica costruendo la **funzione di profilo di verosimiglianza**:

$$\chi_{\text{prof}}^2(\theta_j) \equiv \min_{\{\theta_k: k \neq j\}} \chi^2(\theta_j, \theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_p) \quad (46)$$

Per ogni valore fissato di θ_j , MINOS rilassa il sistema ri-ottimizzando tutti i restanti $p-1$ parametri dimensionali, individuando il minimo vincolato lungo il fondovalle della superficie di costo. Gli estremi dell'intervallo di confidenza al livello di confidenza β (dove $\beta = 1.0$ corrisponde al 68.27% per un singolo grado di libertà) sono definiti dalle radici esatte dell'equazione scalare implicita:

$$\chi_{\text{prof}}^2(\theta_j) - \chi_{\min}^2 = \Delta\chi_{\text{target}}^2 \quad (\text{con } \Delta\chi_{\text{target}}^2 = 1.0) \quad (47)$$

Tale equazione ammette in generale due radici disuguali:

$$\theta_j^+ = \hat{\theta}_j + \Delta\theta_j^+, \quad \theta_j^- = \hat{\theta}_j - \Delta\theta_j^- \quad (48)$$

con $\Delta\theta_j^+ \neq \Delta\theta_j^-$. Questa discrepanza rappresenta la vera asimmetria statistica della misura, correlando direttamente la curvatura non gaussiana con la dispersione dei residui sperimentali.

2.6 Analisi dei Residui Normalizzati (*Pulls*) e Diagnostica di Consistenza

La valutazione della bontà del fit non può esaurirsi nell'esame del valore globale χ_{rid}^2 , il quale rappresenta una media globale che può mascherare deviazioni sistematiche locali o compensazioni accidentali tra punti con varianze sovrastimate e sottostimate. Lo strumento diagnostico locale per eccellenza è lo studio dei ****residui normalizzati**** o *pulls*.

2.6.1 Definizione e Teoria dei Pulls

Dato il punto sperimentale (x_i, y_i) , il residuo verticale grezzo rispetto alla curva stimata è:

$$r_i = y_i - f(x_i; \hat{\theta}) \quad (49)$$

Nel formalismo dell'ODR Metodo A, la varianza combinata complessiva associata a tale osservazione vale:

$$\sigma_{\text{tot},i}^2 = \sigma_{y,i}^2 + \left(\left. \frac{\partial f}{\partial x} \right|_{x_i; \mathbf{p}_{\text{ref}}} \sigma_{x,i} \right)^2 \quad (50)$$

Il residuo normalizzato (pull) z_i è definito come il rapporto adimensionale:

$$z_i = \frac{y_i - f(x_i; \hat{\theta})}{\sigma_{\text{tot},i}} \quad (51)$$

Se il modello fisico $f(x; \theta)$ descrive fedelmente il fenomeno naturale e le incertezze sperimentali $\sigma_{x,i}$ e $\sigma_{y,i}$ sono state stimate correttamente e hanno natura gaussiana, ciascun pull z_i costituisce una variabile casuale che deve seguire asintoticamente la ****distribuzione normale standard****:

$$z_i \sim \mathcal{N}(\mu = 0, \sigma^2 = 1) \quad (52)$$

2.6.2 Proprietà Diagnostiche dei Momenti Statistici dei Pulls

Dato il collettivo dei pulls campionari $\{z_i\}_{i=1}^N$, si calcolano i momenti campionari:

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N z_i, \quad s_z^2 = \frac{1}{N-1} \sum_{i=1}^N (z_i - \bar{z})^2 \quad (53)$$

L'interpretazione delle deviazioni dai valori ideali fornisce indicazioni precise sullo stato dell'analisi:

- **Media non nulla** ($\bar{z} \neq 0$): indica la presenza di un errore sistematico costante (offset non modellizzato, taratura dello zero errata o fondo non sottratto).
- **Varianza significativamente minore di uno** ($s_z^2 \ll 1$): le barre di incertezza strumentali fornite in ingresso (σ_x o σ_y) sono state sovrastimate a priori, oppure il modello ha un numero eccessivo di parametri liberi rispetto alla complessità del segnale (*overfitting*).
- **Varianza significativamente maggiore di uno** ($s_z^2 \gg 1$): le incertezze strumentali sono state sottostimate, oppure sono presenti sorgenti di rumore non termico/gaussiano (fluttuazioni flicker, disturbi ambientali o errori di quantizzazione).
- **Correlazione seriale tra residui adiacenti**: se il grafico temporale dei pulls mostra andamenti a onde o raggruppamenti di segni uguali consecutivi, il modello teorico è matematicamente inadeguato (ad esempio, si sta fittando un fenomeno quadratico con una relazione lineare).

L'analisi congiunta di $\chi_{\text{rid}}^2 \approx 1$, $p\text{-value} \in [0.05, 0.95]$ e una distribuzione dei pulls con $\bar{z} \approx 0$ e $s_z \approx 1$ costituisce il criterio necessario e sufficiente per dichiarare un fit statisticamente valido e conforme ai criteri di riproducibilità scientifica internazionale.