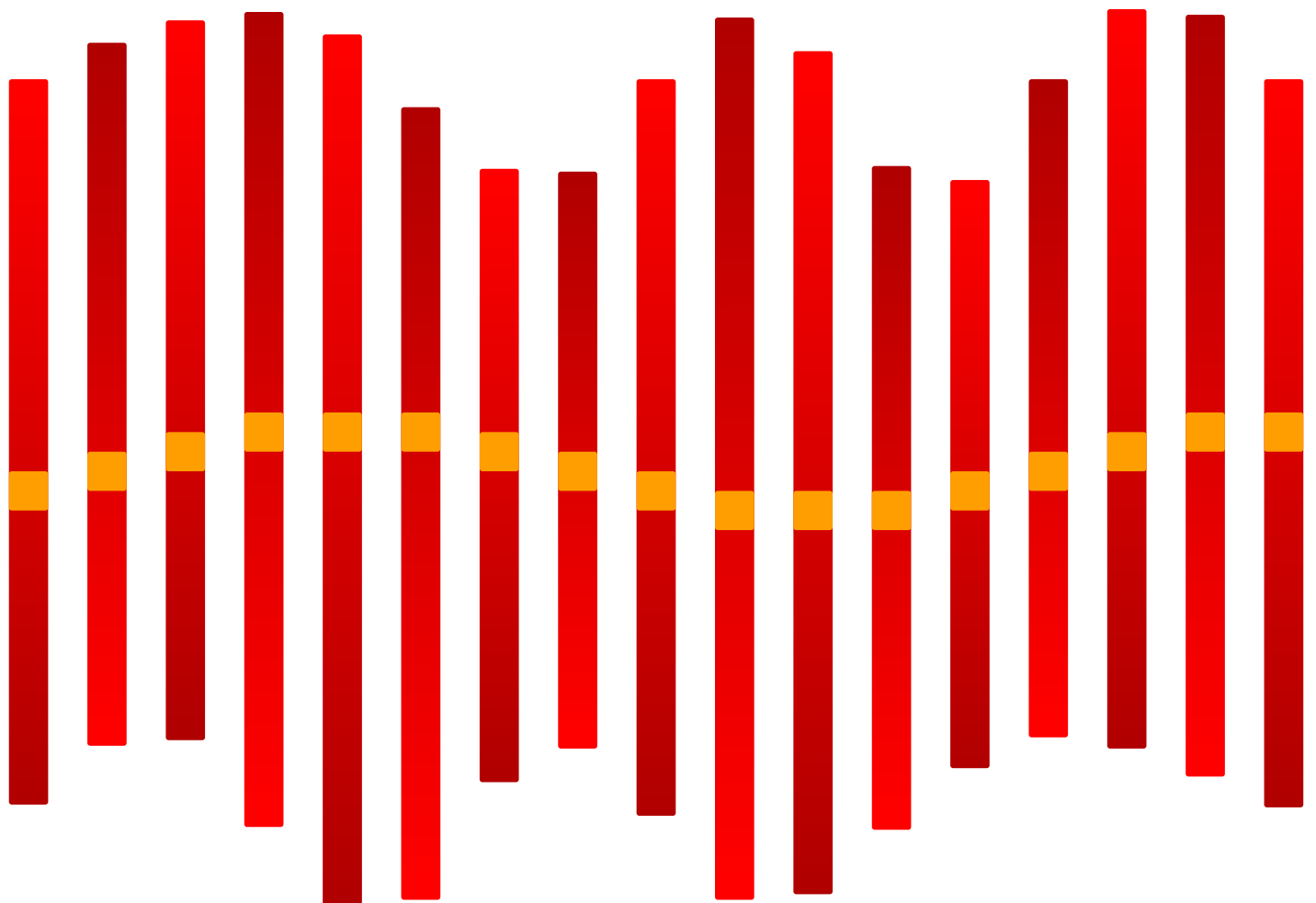


**Numero II**  
**Anno 2013**



# **I** Viaggio nella Scienza **thaca**



**Il calcolo delle variazioni**

---

## **Ithaca: Viaggio nella Scienza**

---

Una pubblicazione del Dipartimento di Matematica e Fisica “Ennio De Giorgi” dell’Università del Salento.

Registrazione presso il Tribunale di Lecce n. 6 del 30 Aprile 2013.  
e-ISSN: 2282-8079

Direttore Responsabile  
**Luigi Spedicato**

Ideatore  
**Giampaolo Co’**

Comitato di Redazione  
**Rocco Chirivì,**  
**Maria Luisa De Giorgi,**  
**Luigi Martina,**  
**Giuseppe Maruccio,**  
**Marco Mazzeo,**  
**Francesco Paparella,**  
**Carlo Sempi.**

Segreteria di Redazione  
**Daniela Dell’Anna.**

© 2013 Dipartimento di Matematica e Fisica “*Ennio de Giorgi*”.

© 2013 per i singoli articoli dei rispettivi autori.

Il materiale di questa pubblicazione può essere riprodotto nei limiti stabiliti dalla licenza  
“*Creative Commons Attribuzione – Condividi allo stesso modo 3.0 Italia*” (CC BY-SA 3.0 IT).

Per il testo della licenza: <http://creativecommons.org/licenses/by-sa/3.0/it/deed.it>

**Ithaca: Viaggio nella Scienza**

è disponibile sul sito:

<http://ithaca.unisalento.it/>

Scriveteci all’indirizzo:

[ithaca@unisalento.it](mailto:ithaca@unisalento.it)

---

**3** Ιθάκη  
Κωνσταντίνος Π. Καβάφης

**5** In questo numero

**7** Il problema di Newton dei profili aerodinamici ottimi  
Giuseppe Buttazzo

**17** Calcolo delle Variazioni e segmentazione di immagini  
Michele Carriero, Antonio Leaci, Franco Tomarelli

**29** Il problema isoperimetrico  
Alessio Figalli

**33** Integrale di Feynman e limite classico  
Matteo Beccaria

**41** Il principio variazionale nella fisica dei sistemi quantistici a molticorpi  
Giampaolo Co'

**49** Adam Smith, John Nash, il prezzo dell'anarchia e la decadenza della società moderna

Vittorio Bilò

**57** La fine del ghiaccio artico ed il futuro del clima

Francesco Paparella

La lezione mancata

**67** L'equazione di Eulero–Lagrange

Rocco Chirivì

---

# Ιθάκη

**Κωνσταντίνος Π. Καβάφης**

(*Costantino P. Kavafis*) Poeta greco, 1863 – 1933.

---

Quando ti metterai in viaggio per Itaca  
devi augurarti che la strada sia lunga  
fertile in avventure e in esperienze.  
I Lestrigoni e i Ciclopi  
o la furia di Nettuno non temere,  
non sarà questo il genere d'incontri  
se il pensiero resta alto e un sentimento  
fermo guida il tuo spirito e il tuo corpo.  
I Ciclopi e Lestrigoni, no certo  
nè nell'irato Nettuno incapperai  
se non li porti dentro  
se l'anima non te li mette contro.

Devi augurarti che la strada sia lunga.  
Che i mattini d'estate siano tanti  
quando nei porti - finalmente, e con che gioia -  
toccherai terra tu per la prima volta:  
negli empori fenici indugia e acquista  
madreperle, coralli, ebano e ambre  
tutta merce fina, anche profumi  
penetranti d'ogni sorta, più profumi  
inebrianti che puoi,  
va in molte città egizie  
impara una quantità di cose dai dotti.

Sempre devi avere in mente Itaca -  
raggiungerla sia il pensiero costante.  
Soprattutto, non affrettare il viaggio;  
fa che duri a lungo, per anni, e che da vecchio  
metta piede sull'isola, tu, ricco  
dei tesori accumulati per strada  
senza aspettarti ricchezze da Itaca.  
Itaca ti ha dato il bel viaggio,  
senza di lei mai ti saresti messo  
in viaggio: che cos'altro ti aspetti?  
E se la trovi povera, non per questo Itaca ti avrà deluso.  
Fatto ormai savio, con tutta la tua esperienza addosso  
già tu avrai capito ciò che Itaca vuole significare.

*Traduzione italiana di Nelo Risi e Margherita Dalmati*

Σὰ βγεῖς στὸν πηγαμὸ γιὰ τὴν Ἰθάκη,  
νὰ εὐχέσαι νὰ 'ναὶ μακρὺς ὁ δρόμος,  
γεμάτος περιπέτειες, γεμάτος γνώσεις.  
Τοὺς Λαιστρυγῶνας καὶ τοὺς Κύκλωπας,  
τὸν θυμωμένο Ποσειδῶνα μὴ φοβάσαι,  
τέτοια στὸν δρόμο σου ποτέ σου δὲν θὰ βρεῖς,  
ἂν μὲν' ἡ σκέψις σου ὑψηλή, ἂν ἐκλεκτὴ  
συγκίνησις τὸ πνεῦμα καὶ τὸ σῶμα σου ἀγγίζει.  
Τοὺς Λαιστρυγῶνας καὶ τοὺς Κύκλωπας,  
τὸν ἄγριο Ποσειδῶνα δὲν θὰ συναντήσεις,  
ἂν δὲν τοὺς κουβανεῖς μὲς στὴν ψυχὴ σου,  
ἂν ἡ ψυχὴ σου δὲν τοὺς στήνει ἐμπρὸς σου.

Νὰ εὐχέσαι νὰ 'ναὶ μακρὺς ὁ δρόμος.  
Πολλὰ τὰ καλοκαιρινὰ πρωινὰ νὰ εἶναι  
ποῦ μὲ τί εὐχαρίστηση, μὲ τί χαρὰ  
θὰ μπαίνεις σὲ λιμένας πρωτοειδωμένους.  
Νὰ σταματήσεις σ' ἐμπορεῖα Φοινικικά,  
καὶ τὲς καλὲς πραγμάτειες ν' ἀποκτήσεις,  
σεντέφια καὶ κοράλλια, κεχρμιπάρια κ' ἔβενους,  
καὶ ἡδονικὰ μυρωδικὰ κάθε λογῆς,  
ὅσο μπορεῖς πιὸ ἄφθονα ἡδονικὰ μυρωδικὰ.  
Σὲ πόλεις Αἰγυπτιακὲς πολλὲς νὰ πᾶς,  
νὰ μάθεις καὶ νὰ μάθεις ἀπ' τοὺς σπουδασμένους.

Πάντα στὸ νοῦ σου νὰ 'χεις τὴν Ἰθάκη.  
Τὸ φθάσιμον ἐκεῖ εἶν' ὁ προορισμός σου.  
Ἀλλὰ μὴ βιάζεις τὸ ταξίδι διόλου.  
Καλλίτερα χρόνια πολλὰ νὰ διαρκέσει.  
Καὶ γέρος πᾶ ν' ἀράξεις στὸ νησί,  
πλούσιος μὲ ὅσα κέρδισες στὸν δρόμο,  
μὴ προσδοκῶντας πλοῦτη νὰ σὲ δώσει ἡ Ἰθάκη.  
Ἡ Ἰθάκη σ' ἔδωσε τ' ὠραῖο ταξίδι.  
Χωρὶς αὐτὴν δὲν θὰ 'βγαίνεις στὸν δρόμο.  
Ἄλλα δὲν ἔχει νὰ σὲ δώσει πᾶ.  
Κι ἂν πτωχικὴ τὴν βρεῖς, ἡ Ἰθάκη δὲν σὲ γέλασε.  
Ἔτσι σοφὸς ποὺ ἔγινες, μὲ τόση πείρα,  
ἦδη θὰ τὸ κατάλαβες οἱ Ἰθάκες τι σημαίνουν.



---

# In questo numero

---

Il tema conduttore di questo numero di *Ithaca* è costituito dal Calcolo delle Variazioni. Si tratta di un tema classico tanto per la Matematica quanto per la Fisica. Benché abbia la sua origine nel Settecento con i lavori di Maupertuis, Euler e Lagrange (tra gli altri), ha acquistato nuova vita con il passare del tempo e ha continuato a affascinare matematici e fisici; ne sia testimone la lezione speciale di Richard Feynman nelle sue *Lectures* [1]. Il lettore troverà articoli che mostrano la straordinaria vitalità che questo tema ancora mostra, adattandosi perfettamente ai recenti sviluppi. Brevemente

- Buttazzo ricerca il profilo di un corpo solido che presenti la minima resistenza al moto, un problema che risale a Newton;
- Carriero, Leaci e Tomarelli applicano il Calcolo delle Variazioni alla segmentazione delle immagini;
- Figalli affronta da un punto di vista molto moderno il problema antichissimo, noto talvolta come *il problema di Didone*, di trovare per un perimetro di lunghezza fissata quello che racchiude l'area massima;
- Beccaria se ne serve per mostrare come, utilizzando il formalismo degli integrali di Feynmann, sia possibile trattare i problemi della meccanica quantistica mantenendo uno stretto contatto con il suo limite classico;
- Co' usa un principio variazionale per descrivere sistemi quantistici a molti corpi;
- Bilò applica la teoria dei giochi a problemi di scienze sociali: anche in questo caso si

tratta di trovare, fra tutte le possibili soluzioni, quelle che rendano massima o minima una predeterminata quantità.

Il numero è completato dall'articolo di Paparella sulle fluttuazioni delle dimensioni della banchisa dell'Artico, che non rientra nel tema principale, ma offre un'informazione aggiornata su un argomento di attualità. Chiude questo numero la *lezione mancata* di Chirivì che può servire da introduzione per il lettore che non abbia mai incontrato il calcolo delle variazioni.

Il numero si apre con una poesia di Kavafis su Itaca, che, letta con gli occhi di chi è interessato alla scienza, trasporta nel mito Omerico lo sforzo e la tensione della ricerca scientifica, ponendo l'accento sul momento della ricerca del risultato più che sul risultato stesso.



[1] R. P. FEYNMAN, R. B. LEIGHTON, M. SANDS: *The Feynman Lectures on Physics*, Cap. 19 - Vol.II, Addison Wesley, London (1969).







---

# Il problema di Newton dei profili aerodinamici ottimi

**Giuseppe Buttazzo**

Dipartimento di Matematica, Università di Pisa

---

**L**a ricerca del profilo di un corpo solido che presenta la minima resistenza al moto è uno dei primi problemi del calcolo delle variazioni e fu posto da Newton nel 1685. Newton non disponeva della teoria dei fluidi e delle equazioni a derivate parziali della fluidodinamica, tuttavia sviluppò un modello molto semplice per calcolare la resistenza al moto di un corpo solido in un fluido.

Nel suo *Principia Mathematica* Newton scrive:

Si globus & cylindrus æqualibus diametris descripti, in medio raro ex particulis æqualibus & ad æquales ab invicem distantias libere dispositis constante, secundum plagam axis cylindri, æquali cum velocitate moveantur: erit resistentia globi duplo minor quam resistentia cylindri. [...] Quam quidem propositio-nem in construendis navibus non inutilem futuram esse censeo.<sup>1</sup>

---

<sup>1</sup>Libro II, Proposizione XXXIV, Teorema XXVIII: *Una sfera ed un cilindro di uguale diametro si muovano parallelamente all'asse del cilindro, immersi in un mezzo rarefatto composto di particelle identiche, disposte senza vincoli a distanza costante l'una dall'altra: la resistenza della sfera sarà la metà della resistenza del cilindro.* Nel successivo scolio: *Senza dubbio ritengo che la proposizione non sarà inutile per l'ingegneria navale.*

Per capire quanto la resistenza incide sul moto di un corpo partiamo da un semplice problema di balistica: il lancio di un oggetto nel vuoto. Tale fenomeno è governato dalla legge di Newton

$$\text{massa} \cdot \text{accelerazione} = \text{forze in gioco}$$

dove le forze in gioco si riducono nel nostro caso alla sola forza di gravità. Supponendo la gravità costante, ipotesi ragionevole per lanci di breve gittata, ma non per lanci balistici di migliaia di km, con facili calcoli si ricava l'equazione della traiettoria:

$$y = x \tan \theta - \frac{g(1 + \tan^2 \theta)}{2v^2} x^2$$

La traiettoria è dunque un arco di parabola che non dipende dalla massa del corpo lanciato; nell'equazione precedente abbiamo indicato con

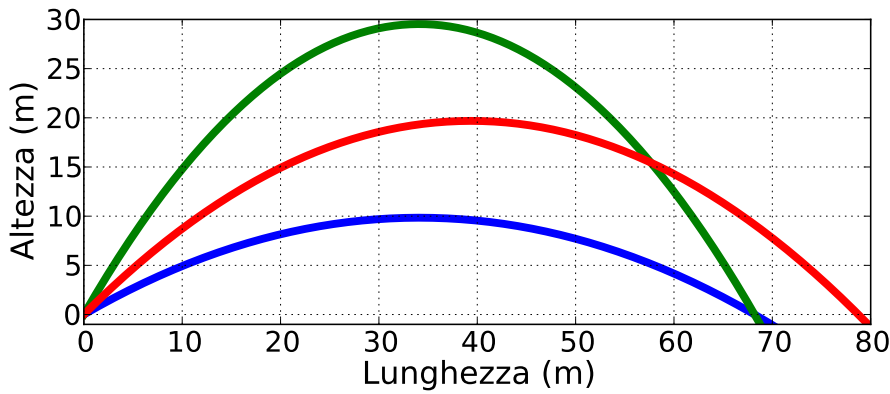
$\theta$  = angolo di tiro;  
 $v$  = velocità iniziale;  
 $g$  = accelerazione di gravità.

Ad esempio, usando i dati:

$$v \sim 100 \text{ km/h} \sim 27.78 \text{ m/sec};$$
$$g \sim 9.8 \text{ m/sec}^2 \text{ (sulla terra)}$$

si trovano le traiettorie riportate in Figura 1, con gittata

$$L = \frac{2v^2 \tan \theta}{g(1 + \tan^2 \theta)}$$



**Figura 1:**

$L \sim 68.2 \text{ m}$  con  $\theta = 30^\circ$ ,  
 $L \sim 68.2 \text{ m}$  con  $\theta = 60^\circ$ ,  
 $L \sim 78.7 \text{ m}$  con  $\theta = 45^\circ$ .

Si trova, come è ben noto, che la *gittata massima*  $L_{max}$  vale

$$L_{max} = \frac{v^2}{g} \quad \text{in corrispondenza di } \theta = 45^\circ$$

Un modello matematico più realistico (almeno per lanci balistici sulla Terra) deve però tener conto della *resistenza dell'aria* che è una forza  $R$  che dipende dalla velocità:

$$R = c \phi(v)$$

dove  $\phi$  è una funzione che si ricava sperimentalmente e  $c$  è una costante che dipende dalla *forma* dell'oggetto lanciato.

Utilizzando ad esempio  $\phi(v) = v$  l'equazione del moto è un'equazione differenziale ordinaria lineare del secondo ordine a coefficienti costanti e si trova con facili calcoli un'espressione esplicita della traiettoria, data da:

$$y = x \tan \theta + \frac{gm}{c} \left[ \frac{x}{v \cos \theta} + \frac{m}{c} \log \left( 1 - \frac{cx}{mv \cos \theta} \right) \right]$$

Le traiettorie non sono più archi di parabola; inoltre esse dipendono dalla massa dell'oggetto lanciato e dalla costante  $c$ , che a sua volta dipende dalla *forma* dell'oggetto stesso. Alcune traiettorie sono riportate in Figura 2, dove si è usato un coefficiente  $c$  piuttosto grande per evidenziare meglio l'effetto della resistenza dell'aria e la natura non parabolica delle curve.

Va notato che la gittata massima non si ottiene più per  $\theta = 45^\circ$  ma per angoli più piccoli, dipendenti dai dati del problema. Inoltre, la resistenza dell'aria fa diminuire notevolmente la gittata massima; la tabella in Figura 3 mostra l'effetto

della resistenza dell'aria sulla gittata massima di alcuni proiettili comuni.

La resistenza aerodinamica di un oggetto è data da

$$R = \frac{1}{2} \rho S C v^2$$

dove  $\rho$  è la densità del fluido in cui avviene il moto,  $v$  la velocità dell'oggetto,  $S$  l'area della sezione ortogonale al moto,  $C$  il coefficiente aerodinamico dipendente dalla forma dell'oggetto.

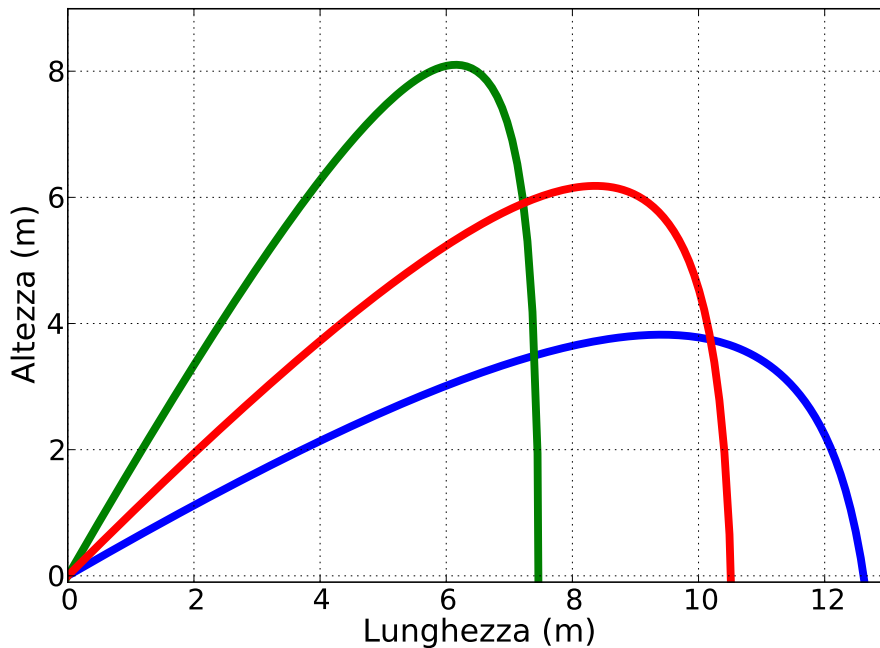
Nella pratica, si calcola la resistenza tramite *galleria del vento* e si deduce il coefficiente  $C$  dalla formula precedente. Qui di seguito, in Figura 4 alcuni coefficienti aerodinamici di alcune note automobili, ed in Figura 5 alcuni altri coefficienti aerodinamici.

Il modello di Newton per calcolare la resistenza aerodinamica di un corpo solido si ottiene facilmente supponendo le condizioni seguenti:

- il fluido è supposto costituito da particelle *indipendenti* che si muovono con velocità costante;
- la resistenza è dovuta soltanto agli urti delle particelle del fluido con la superficie del corpo, urti che sono supposti *perfettamente elastici*;
- ogni particella urta la superficie del corpo solido al più una sola volta;
- si trascurano l'attrito tangenziale ed altri effetti quali la *vorticità* e le *turbolenze*.

Notiamo che la condizione di singolo urto si ha ad esempio se il corpo solido in questione è *convesso*.

Con facili argomenti di trigonometria elementare si trova allora che la pressione in un punto



**Figura 2:**  
 $\theta = 30^\circ$ ,  $\theta = 60^\circ$ ,  $\theta = 45^\circ$ .

del profilo del corpo è proporzionale a  $\sin^2 \theta$ , dove  $\theta$  è l'inclinazione del profilo del corpo rispetto alla direzione del moto, come illustrato in Figura 6.

Descrivendo il profilo del corpo mediante il grafico di una funzione  $u(x)$  definita sulla sezione  $\Omega$  ortogonale al moto, si trova

$$\sin^2 \theta = \frac{1}{1 + \tan^2(\pi/2 - \theta)} = \frac{1}{1 + |\nabla u|^2}$$

e la resistenza totale sarà quindi proporzionale a

$$F(u) = \int_{\Omega} \frac{1}{1 + |\nabla u|^2} dx$$

tramite il coefficiente di proporzionalità  $\rho v^2$ , dove  $\rho$  è la densità del fluido e  $v$  la sua velocità.

Dividendo la resistenza totale per l'area della sezione  $\Omega$  otteniamo la resistenza relativa del profilo  $u$  (spesso chiamata *coefficiente*  $C_x$ ):

$$C(u) = \frac{F(u)}{|\Omega|}$$

Le seguenti osservazioni sono immediate:

- si ha sempre  $0 \leq C(u) \leq 1$ ;
- se il corpo ha un profilo *piatto*, cioè se  $u$  è costante, si ha  $C(u) = 1$ ;
- se il corpo è una *semisfera* di raggio  $R$  si ha

$u(x) = \sqrt{R^2 - |x|^2}$  e si trova

$$C(u) = \frac{F(u)}{\pi R^2} = \frac{2}{R^2} \int_0^R \frac{R^2 - r^2}{R^2} r dr = \frac{1}{2}$$

in accordo con quanto previsto da Newton nel 1685.

Alcuni corpi con  $C(u) = 1/2$  (o vicino a  $1/2$ ) sono illustrati nelle Figure 7 e 8.

Studiamo ora il problema di trovare il profilo ottimo in una classe di profili *ammissibili*:

$$\min \left\{ F(u) : u \text{ profilo ammissibile} \right\}$$

Calibro	Velocità m/s	Gittata in m.	nel vuoto m.
4,5 mm aria compressa	120	100	1469
4,5 mm aria compressa	200	200	4082
6/9 mm Flobert	225	700	5166
.22 corto	260	1000	6898
.22 Long Rifle	350	1370	12500
.22 Long Rifle HS	370	1500	13969
.22 Winch. Magnum	610	1800	37969
243 Winch.	1070	3200	116827
6,35 mm	220	800	4939
7,65 mm	285	1300	8288
9 mm corto	285	1300	8288
9 mm Para	350	1700	12500
.45 ACP	300	1620	9184
30 M1Carb.	600	2000	36735
7x70 mm	830	3500	70296
8x57 mm JS	830	3500	70296
6,5x57 mm	1020	4000	106163
7x57 mm	850	4500	73724
6,5x68 mm	1150	5000	134949

**Figura 3:** Effetto della resistenza dell'aria sulla gittata di alcuni proiettili.

Alfa Romeo Giulia	1964	0,43
Alfa Romeo Alfetta	1972	0,42
Alfa Romeo 33	1983	0,36
Alfa Romeo 90	1984	0,37
Alfa Romeo 75	1985	0,32
Alfa Romeo 75 Turbo Evoluzione	1987	0,30
Alfa Romeo 164	1988	0,31
Alfa Romeo RZ/SZ	1989	0,30
Alfa Romeo 155	1992	0,29
Alfa Romeo 156 Berlina	1997	0,31
Alfa Romeo 156 Sportwagon	1999	0,30
Alfa Romeo 159	2006	0,32
Alfa Romeo Mi.To. 1.6 JTDm 120 cv	2008	0,29
Alfa Romeo Mi.To. 78 cv	2008	0,35

**Figura 4:** I coefficienti aerodinamici di alcune Alfa Romeo.

Il problema considerato da Newton limitava la classe dei profili ammissibili a quelli con *simmetria radiale*, verificanti inoltre l'ipotesi di convessità in modo da verificare la condizione di urto singolo delle particelle di fluido con la superficie del corpo solido. La stessa limitazione veniva imposta in tutti i trattati di calcolo delle variazioni (Bliss, Bolza, Carathéodory, Cesari, Hestenes, Miele, Tonelli, Young, ...).

La prima formulazione generale del problema si trova in [1] ed in [2]. I problemi che si pongono sono:

- individuazione di una classe "naturale" di profili ammissibili;
- dimostrazione dell'esistenza di un profilo ottimo;
- studio delle "condizioni necessarie di ottimalità" che i profili ottimi verificano;
- questione della *radialità* dei profili ottimi nel caso in cui la sezione  $\Omega$  sia un cerchio;
- eventuale "rottura di simmetria" nel caso in cui ciò non avvenga e conseguente *non unicità* della soluzione ottima.

Il funzionale  $F$  non è convesso nè coercivo, per cui i *metodi diretti* del calcolo delle variazioni non sono applicabili. Inoltre, senza ulteriori restrizioni alla classe dei profili ammissibili, l'estremo inferiore del funzionale  $F$  è zero, come si vede subito considerando le funzioni

$$u_n(x) = n \operatorname{dist}(x, \partial\Omega)$$

Rumpler Tropfenwagen	1921	0,29
Volkswagen Maggiolino	1946	0,38
Volkswagen New Beetle	1999	0,38
Fiat 500	2009	0,32
Maserati Gran Sport	2009	0,33
Toyota Prius	2009	0,25
Mercedes Classe E Coupé	2009	0,24
Ferrari 360 Modena Novitec	2004	0,35
Aprilia RSV 1000 R		0,30
Go Kart		0,80
Formula Uno		0,90
Aereo leggero		0,12
Ciclista da turismo		1,00
Goccia d'acqua		0,05

**Figura 5:** Alcuni altri coefficienti aerodinamici.

(coni di base  $\Omega$  ed altezza che tende a  $+\infty$ ), per cui si ha

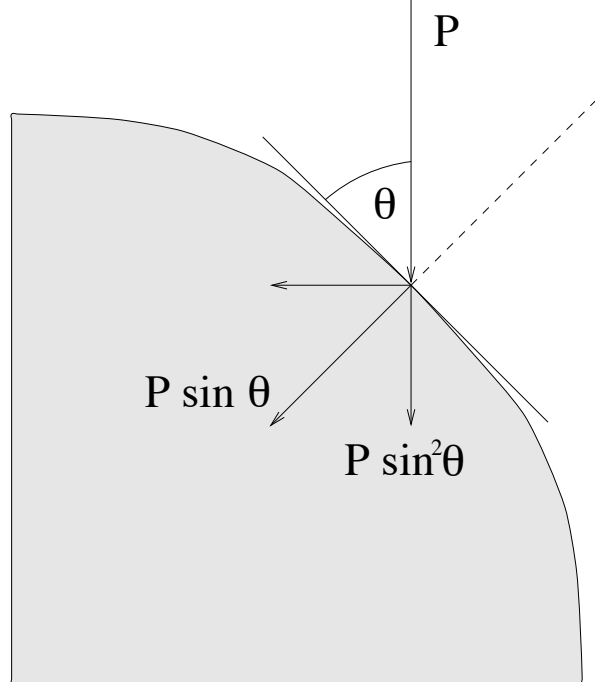
$$\lim_{n \rightarrow +\infty} F(u_n) = 0$$

ed evidentemente nessun corpo solido ha resistenza nulla. Dunque senza ulteriori restrizioni sulle funzioni ammissibili il profilo ottimo *non esiste*.

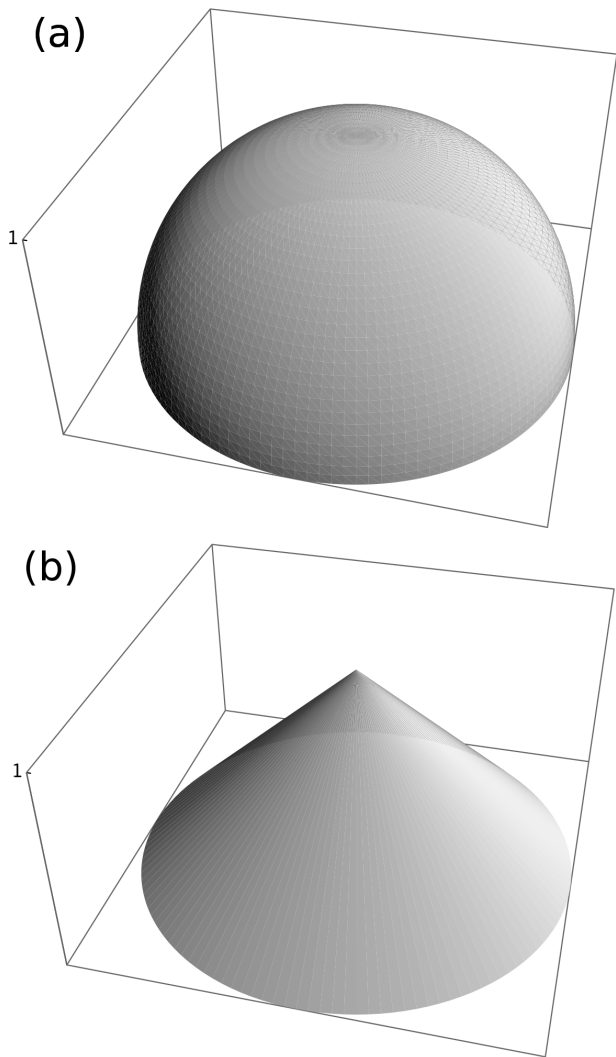
Si potrebbe erroneamente pensare che la non esistenza del minimo di  $F$  sia dovuta al fatto che le funzioni  $u_n$  non verificano limitazioni uniformi. Tuttavia, anche imponendo ai profili ammissibili un vincolo del tipo

$$0 \leq u(x) \leq M \quad \forall x \in \Omega$$

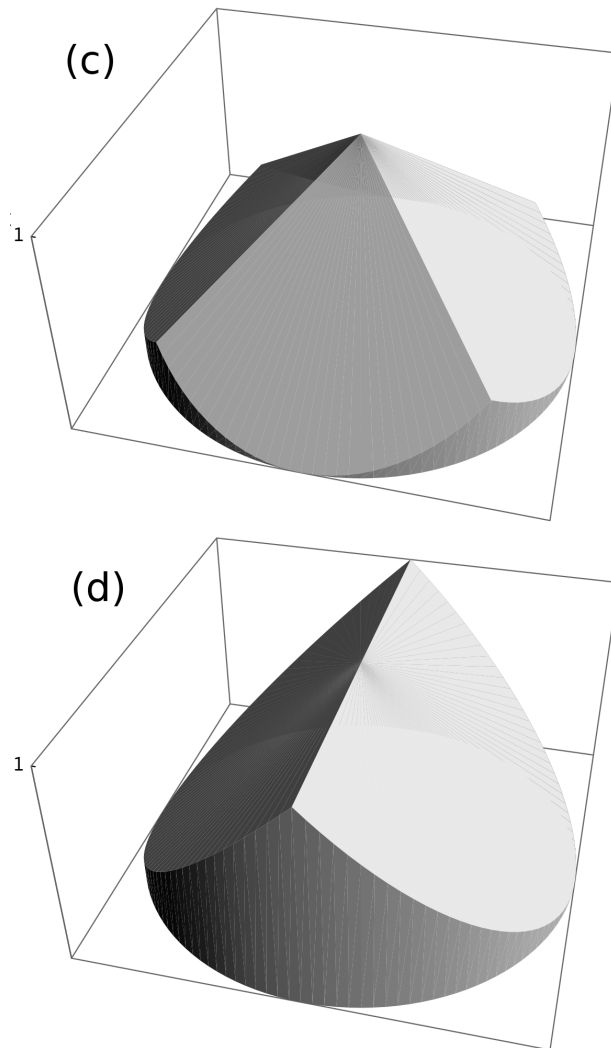
si ottiene ancora che l'estremo inferiore di  $F$  è



**Figura 6:** La legge di pressione di Newton.



**Figura 7:** (a) semisfera; (b) cono.



**Figura 8:** (c) una forma piramidale; (d) un'altra forma piramidale.

zero, come mostrano ad esempio le funzioni

$$u_n(x) = M \sin^2(n|x|)$$

per le quali si ha ancora

$$\lim_{n \rightarrow +\infty} F(u_n) = 0$$

Tuttavia, per profili del tipo del grafico di  $M \sin^2(n|x|)$  le particelle del fluido hanno *urti multipli* con il corpo, ed il modello di Newton, che considera solo il primo urto, non è più valido. Consideriamo allora soltanto corpi convessi per i quali gli urti sono sempre singoli; in altri termini, restringiamo l'analisi alle funzioni  $u(x)$  che sono concave, e quindi il problema di Newton diventa:

$$\min \left\{ F(u) : 0 \leq u \leq M, u \text{ concava in } \Omega \right\}$$

Vedremo che questo problema risulta ben posto. Condizioni più generali della convessità, che ancora assicurano un solo urto delle particelle di fluido con il corpo solido, sono state considerate ad esempio in [3], [4].

Nel caso unidimensionale (o quando il profilo dipende solo da una dimensione)  $\Omega$  è il segmento  $[0, L]$  ed il problema diventa:

$$\min \left\{ \int_0^L \frac{1}{1+|u'|^2} dx : 0 \leq u \leq M, u \text{ concava in } [0, L] \right\}$$

In tal caso le soluzioni ottime sono le seguenti, illustrate in Figura 9:

- se  $M \geq L/2$  il profilo ottimo è dato dal triangolo isoscele di altezza  $M$ ;

- se  $M < L/2$  il profilo ottimo è dato dal trapezio isoscele di altezza  $M$  e pendenza laterale uguale ad 1.

Nel caso bidimensionale di  $\Omega$  un cerchio di raggio  $R$ , e restringendosi alle funzioni  $u(r)$  con simmetria radiale, scrivendo la resistenza in coordinate polari si trova

$$F(u) = 2\pi \int_0^R \frac{r}{1 + |u'(r)|^2} dr$$

e di conseguenza il problema di Newton per corpi a simmetria cilindrica diventa allora:

$$\min \left\{ \int_0^R \frac{r}{1 + |u'(r)|^2} dr : u(0) = M, \right. \\ \left. u(R) = 0, u \text{ concava} \right\}$$

Tale problema di minimo ammette l'equazione di Eulero-Lagrange

$$ru' = C(1 + u'^2)^2 \quad \text{su } \{u' \neq 0\}$$

con  $C < 0$  costante. La soluzione  $u$  si può calcolare esplicitamente in forma parametrica utilizzando la funzione

$$f(t) = \frac{t}{(1+t^2)^2} \left( -\frac{7}{4} + \frac{3}{4}t^4 + t^2 - \log t \right) \quad \forall t \geq 1$$

che risulta strettamente crescente, e le quantità

$$T = f^{-1}(M/R), \quad r_0 = \frac{4RT}{(1+T^2)^2}$$

Si trova allora che il profilo ottimo  $u$  è dato da  $u(r) = M$  per  $r \in [0, r_0]$  e  $\forall t \in [1, T]$

$$\begin{cases} r(t) = \frac{r_0}{4t} (1+t^2)^2 \\ u(t) = M - \frac{r_0}{4} \left( -\frac{7}{4} + \frac{3}{4}t^4 + t^2 - \log t \right) \end{cases}$$

Osserviamo che:

- il profilo radiale ottimo ha sempre una zona piatta;
- la soluzione ottima radiale è lipschitziana ed è unica;
- la soluzione ottima radiale si annulla su  $\partial\Omega$ ;
- $|u'(r)| > 1$  per ogni  $r > r_0$  ed  $|u'(r_0^+)| = 1$ ; in particolare si ha  $|u'(r)| \notin ]0, 1[$ .

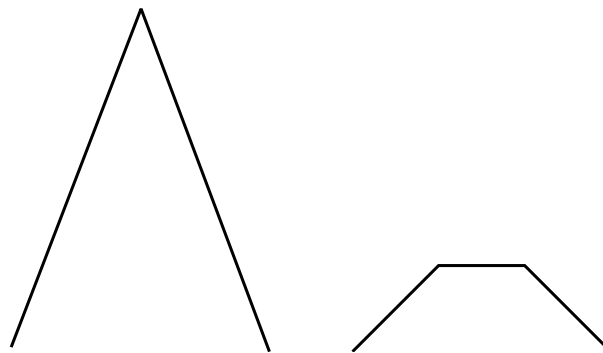


Figura 9: Profili ottimi in dimensione uno:  $M \geq L/2$  (sinistra),  $M \leq L/2$  (destra).

Alcuni profili ottimi sono illustrati nelle Figure 10, 11, 12.

La resistenza relativa  $C(u)$  dei corpi radiali ottimi, ed il raggio  $r_0$  della zona piatta superiore, dipendono solo dal rapporto  $M/R$ . Ad esempio si ha:

	$M/R = 1$	$M/R = 2$	$M/R = 3$	$M/R = 4$
$r_0/R$	0.35	0.12	0.048	0.023
$C(u)$	0.37	0.16	0.082	0.049

e per  $M/R \rightarrow +\infty$  si hanno le stime asintotiche:

$$r_0/R \approx \frac{27}{16} (M/R)^{-3} \quad \text{per } M/R \rightarrow +\infty$$

$$C(u) \approx \frac{27}{32} (M/R)^{-2} \quad \text{per } M/R \rightarrow +\infty$$

Nel problema di Newton, di minimo per il funzionale

$$\int_{\Omega} \frac{1}{1 + |Du|^2} dx$$

altri tipi di vincoli possono essere imposti (sempre mantenendo la condizione di concavità):

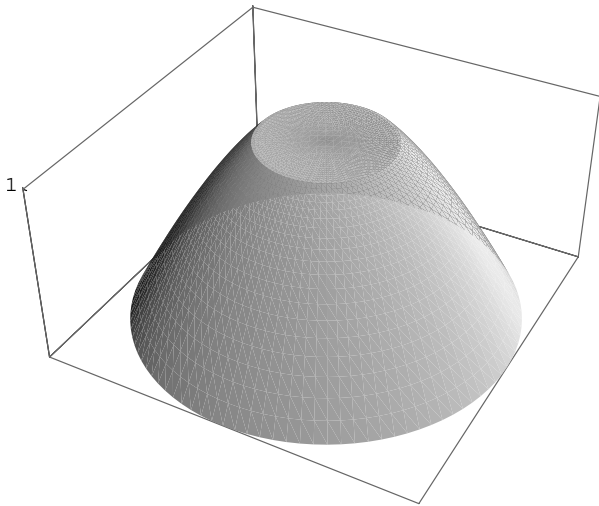
- vincoli di volume  $\int_{\Omega} u dx \leq m$ ;
- vincoli di superficie  $\int_{\Omega} \sqrt{1 + |\nabla u|^2} dx \leq m$ .

Per un panorama sulle possibili applicazioni del problema di Newton in aerodinamica si può consultare ad esempio il libro di A. Miele [6].

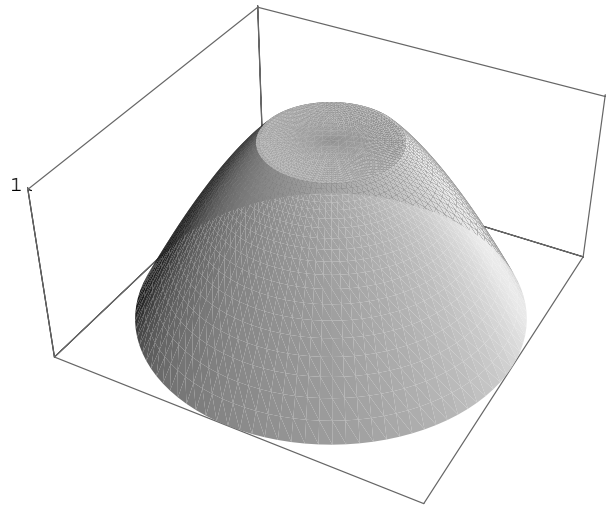
Il risultati di esistenza di un profilo aerodinamico ottimo è il seguente

**Teorema 1.** Se la sezione  $\Omega$  è convessa il problema di Newton ha soluzione ottima nella classe dei profili  $u(x, y)$  tali che

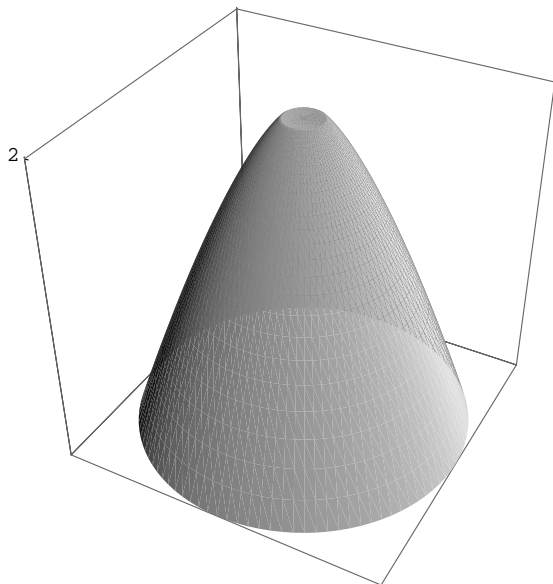
$$0 \leq u \leq M, \quad u \text{ è concava}$$



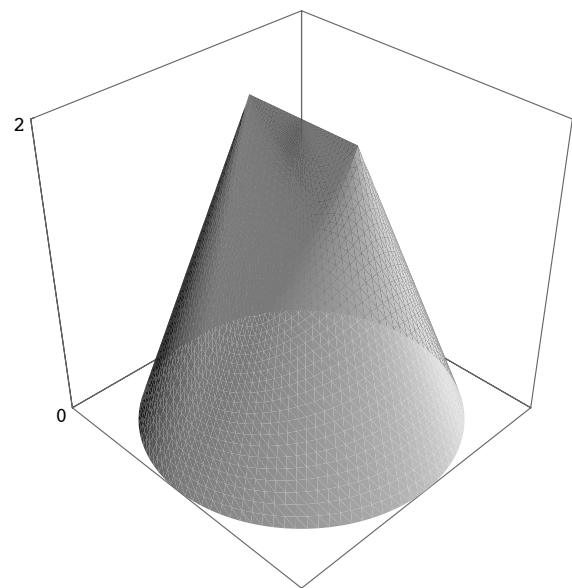
**Figura 10:** La forma ottima radiale nel caso  $M = R$ .



**Figura 12:** La forma ottima radiale nel caso  $M = R/2$ .



**Figura 11:** La forma ottima radiale nel caso  $M = 2R$ .



**Figura 13:** Un profilo "a cacciavite".

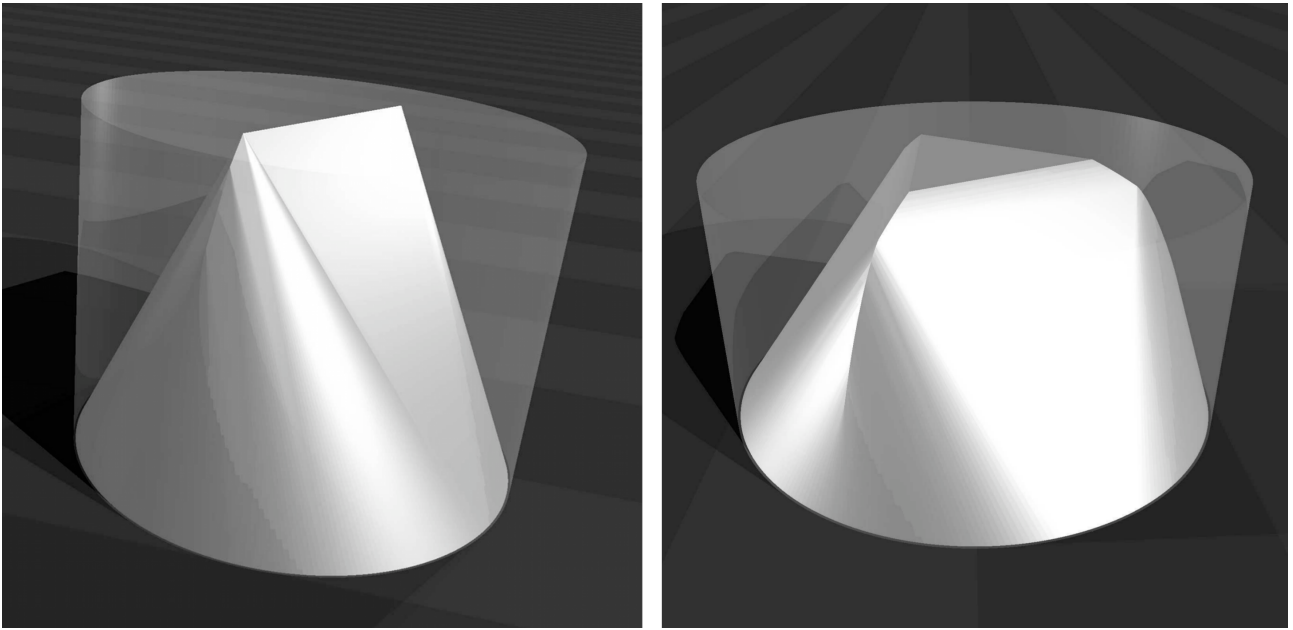
Faremo ora vedere che nel caso in cui  $\Omega$  è un disco, non si ha una soluzione con simmetria radiale. Mostreremo tale *rottura di simmetria* in vari modi. Un primo modo, sviluppato in [5], consiste nell'esibire un profilo con resistenza inferiore al miglior profilo radiale. Considerando un profilo del tipo illustrato in Figura 13 e scegliendo opportunamente la lunghezza del segmento in alto, cioè l'insieme  $\{u = M\}$ , si può calcolare la resistenza del profilo così ottenuto e si trova per  $M/R \rightarrow +\infty$ :

$$C(u) \approx 0.77(M/R)^{-2} < \frac{27}{32}(M/R)^{-2} \approx C(u_{rad})$$

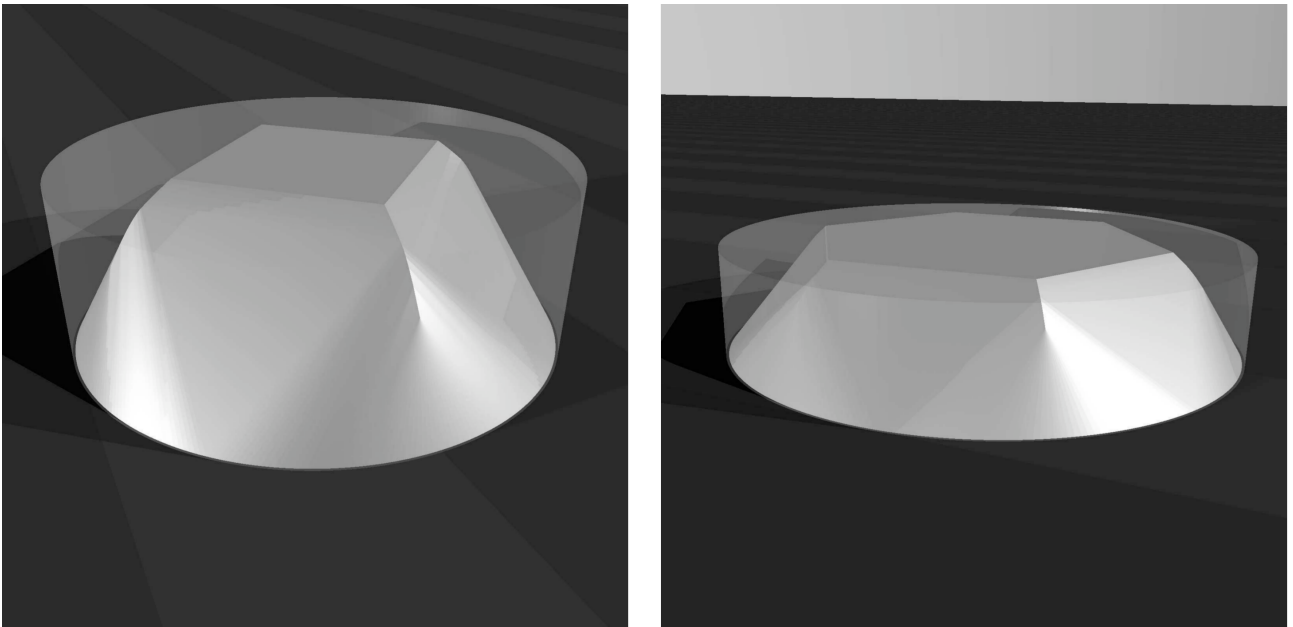
Dunque, per  $M/R$  abbastanza grande (maggiore di 2 nel calcolo fatto in [5]) la soluzione non può essere radiale. Resta il caso  $M/R$  piccolo, che si studia attraverso alcune condizioni necessarie di ottimalità, cioè condizioni che una soluzione, per il fatto che è ottima, deve verificare.

**Teorema 2.** Sia  $\Omega$  un insieme convesso di  $\mathbb{R}^N$  e sia  $u$  una soluzione del problema di Newton. Supponiamo che in un aperto  $\omega$  la funzione  $u$  sia di classe  $C^2$  e che in  $\omega$  sia  $u < M$ . Allora si ha

$$\det \nabla^2 u \equiv 0 \quad \text{in } \omega$$



**Figura 14:** *Un profilo ottimo abbastanza alto (sinistra), un profilo ottimo più basso (destra).*



**Figura 15:** *Un profilo ottimo ancora più basso (sinistra), un profilo ottimo molto basso (destra).*



**Corollario 3.** *Il problema di Newton non ha soluzioni radiali, per alcun valore di  $M$ . Infatti, applicando il teorema precedente alla soluzione radiale, in una regione dove essa è regolare, si trova una contraddizione.*

Caratterizzare le soluzioni del problema di Newton è ancora una *questione aperta*. In particolare non è noto come deve essere fatta la zona superiore  $\{u = M\}$  e non si sa se  $u$  è regolare nella zona  $\{u < M\}$ . Si hanno solo dei risultati numerici sulla forma dei profili ottimi; quelli che mostriamo qui nelle Figure 14 e 15 sono stati ottenuti da E. Oudet.



- [1] G. BUTTAZZO, B. KAWOHL: *On Newton's problem of minimal resistance*. Math. Intelligencer, **15** (1993), 7–12.
- [2] G. BUTTAZZO, V. FERONE, B. KAWOHL: *Minimum problems over sets of concave functions and related questions*. Math. Nachr., **173** (1995), 71–89.
- [3] M. COMTE, T. LACHAND-ROBERT: *Existence of minimizers for Newton's problem of the body of minimal resistance under a single impact assumption*. J. Anal. Math., **83** (2001), 313–335.

- [4] M. COMTE, T. LACHAND-ROBERT: *Newton's problem of the body of minimal resistance under a single-impact assumption*. Calc. Var. Partial Differential Equations, **12** (2) (2001), 173–211.
- [5] P. GUASONI: *Problemi di ottimizzazione di forma su classi di insiemi convessi*. Tesi di Laurea, Università di Pisa, 1995-1996.
- [6] A. MIELE: *Theory of Optimum Aerodynamic Shapes*. Academic Press, New York (1965).



**Giuseppe Buttazzo:** Ha conseguito la Laurea in Matematica nel 1976 presso l'Università di Pisa ed il Diploma nel 1976 presso la Scuola Normale Superiore di Pisa. Dal 1987 al 1990 è stato professore ordinario di Analisi Matematica all'Università di Ferrara e dal 1990 all'Università di Pisa. Si occupa di Calcolo delle Variazioni, Equazioni alle Derivate Parziali, Ottimizzazione.



---

# Calcolo delle Variazioni e segmentazione di immagini

*In sintesi, credo che il Calcolo delle Variazioni sia un settore molto vasto e variegato [...]; esso costituisce un'ottima "palestra" per capire quali sono le radici etiche e culturali del metodo scientifico.*

**Ennio De Giorgi**

**Michele Carriero**

*Dipartimento di Matematica e Fisica "E. De Giorgi" - Università del Salento*

**Antonio Leaci**

*Dipartimento di Matematica e Fisica "E. De Giorgi" - Università del Salento*

**Franco Tomarelli**

*Dipartimento di Matematica - Politecnico di Milano*

---

**I**n questa presentazione esponiamo alcuni problemi con discontinuità libere relativi alla segmentazione d'immagini, introducendo in particolare lo studio, dal punto di vista analitico e numerico, dei funzionali di Mumford & Shah e di Blake & Zisserman.

Il Calcolo delle Variazioni è l'ambito in cui problemi di minimo o di massimo e nozioni di equilibrio energetico trovano un linguaggio preciso e formalizzazioni per mezzo di principi variazionali.

La segmentazione di immagini è un problema rilevante sia nella elaborazione di immagini digitali che nella comprensione della visione biologica.

Qui presentiamo una breve rassegna di risulta-

ti recenti che sono collegati sia alla formulazione e alla sperimentazione di algoritmi per la segmentazione automatica di immagini digitali, sia alla comprensione della percezione visiva. Per un'ampia descrizione rinviamo al volume [1].

Una immagine digitale (che potremmo chiamare "discreta") proviene da un contesto continuo, attraverso un processo di campionamento spaziale e di quantizzazione della luminosità. In pratica l'immagine reale (analogica) è suddivisa in piccoli rettangoli (i *pixels*), a ciascuno dei quali è assegnato un numero (o tre numeri, per le immagini a colori) ottenuto utilizzando la media della luminosità nel rettangolo.

A partire dalla metà degli anni '80, nella elaborazione d'immagini sono stati utilizzati il Calcolo delle Variazioni e le Equazioni alle Derivate Parziali, settori della matematica già ben consolidati

## La misura di Hausdorff unidimensionale

La definizione formale è la seguente: per ogni sottoinsieme  $K$  di  $\mathbb{R}^2$ ,

$$\mathcal{H}^1(K) = \lim_{\varepsilon \downarrow 0} \mathcal{H}_\varepsilon^1(K) = \sup_{\varepsilon > 0} \mathcal{H}_\varepsilon^1(K)$$

(possibilmente anche  $+\infty$ ) dove

$$\mathcal{H}_\varepsilon^1(K) = \inf \left\{ \sum_{i=1}^{\infty} \text{diam} B_i; K \subset \bigcup_{i=1}^{\infty} B_i, \text{diam} B_i < \varepsilon \right\}$$

È possibile definire le misure di Hausdorff  $\mathcal{H}^s$   $s$ -dimensionali per ogni numero reale  $s$  con  $0 \leq s \leq 2$ , inserendo nella serie il termine  $c_s(\text{diam} B_i)^s$ . Queste misure intervengono nella teoria degli insiemi "frattali". Per  $s = 0$  si ottiene la misura che conta i punti, per  $s = 2$  si ottiene l'usuale misura di Lebesgue nel piano:  $\mathcal{H}^2(K) = \text{meas}(K)$ .

per lo studio del mondo fisico. In questo contesto sia l'immagine sia la sua elaborazione sono definite in ogni punto del dominio. Una volta formulato e risolto un problema di elaborazione d'immagini nel modello continuo, la soluzione è solitamente approssimata con metodi numerici, che ci forniscono un'immagine "digitale".

La descrizione si rivolge anche ai non specialisti nel Calcolo delle Variazioni; per questo tralasciamo l'analisi in un contesto più generale e ci limitiamo a casi modello: talvolta, al fine di semplificare l'esposizione, i risultati sono espressi con ipotesi che non sono minimali.

Esistono molti modi diversi per definire gli obiettivi della segmentazione e non vi è alcuna nozione universalmente accettata: questa esposizione è limitata ad alcuni modelli di decomposizione di una immagine, in cui è assegnata una funzione che descrive l'intensità del segnale associata a ciascun punto (tipicamente l'intensità della luce su uno schermo). In parole semplici, la segmentazione di una immagine consiste nella scomposizione dell'immagine per evidenziarne le linee di maggiore discontinuità dell'intensità luminosa.

La segmentazione può essere ottenuta mediante diverse tecniche (cfr. [1], [2]). Noi descriviamo alcuni approcci basati su un principio variazionale: la minimizzazione di una opportuna energia. Evidenziamo che le funzioni che descrivono l'intensità della luce in segmentazioni ammissibili possono presentare discontinuità.

Questi modelli di segmentazione sono utili nella descrizione della visione, e sono in grado di fornire anche utili indicazioni su questioni rilevanti per la fisiologia della visione. Per esempio, questi modelli possono aiutare nella comprensione di come l'enorme quantità di dati contenuti in una singola immagine possa essere ridotta e rapidamente trasformata, preservandone la geometria essenziale che è fondamentale per l'interpretazione dell'immagine stessa.

Le formalizzazioni variazionali dei modelli per la segmentazione forniscono una più profonda comprensione dell'analisi di una immagine, producono questioni matematiche interessanti (alcune ancora aperte) la cui soluzione, con le stime globali che ne derivano, contribuisce ad una migliore interpretazione della percezione visiva.

Questi modelli si inquadrano nell'ampia classe dei problemi con discontinuità libere, introdotta da Ennio De Giorgi ([3]). In questo ambito, strumenti moderni di teoria geometrica della misura, recenti sviluppi sulle superficie minime e sulla regolarità delle estremali in Calcolo delle Variazioni, consentono lo studio di problemi di minimo per funzionali in cui sono presenti sia termini di massa sia termini di superficie (di linea, nel caso bidimensionale): in tale contesto sono ammissibili funzioni discontinue (in senso matematico), e talvolta le loro discontinuità sono proprio le caratteristiche principali della soluzione.

Qui delineiamo le principali proprietà di due

modelli variazionali noti in letteratura come modello di Mumford & Shah (cfr. [4]) e modello di Blake & Zisserman (cfr. [5]). Questi approcci bilanciano simultaneamente la regolarizzazione del segnale e la lunghezza della segmentazione; spesso questi modelli sono più accurati nel rilevare le discontinuità, rispetto ad altre tecniche di filtraggio.

Limitiamo la nostra discussione a immagini bidimensionali monocromatiche. Per le immagini a colori rimandiamo all'analisi di segnali a valori vettoriali ([6]).

## Il funzionale di Mumford & Shah

Dato un aperto limitato  $\Omega \subset \mathbb{R}^2$ , consideriamo il seguente funzionale, introdotto in [4] da Mumford & Shah,

$$\begin{aligned} \mathcal{MS}(K, u) := & \iint_{\Omega \setminus K} (\|Du\|^2 + \mu|u - g|^2) dx_1 dx_2 \\ & + \alpha \text{length}(K \cap \Omega) \end{aligned} \quad (1)$$

dove  $K \subset \mathbb{R}^2$  è l'unione di una famiglia di curve (*a priori* incognite),  $x = (x_1, x_2)$  indica un punto in  $\Omega$ ,  $\text{length}(K \cap \Omega)$  indica la somma delle lunghezze di queste curve e  $u$  è una funzione scalare differenziabile in  $\Omega \setminus K$ ;  $\|Du\|$  denota la norma euclidea del gradiente  $Du = (\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2})$  di  $u$ . Il dato  $g$  è una funzione limitata definita su  $\Omega$ . Chiariamo nel successivo riquadro il termine lunghezza (totale), richiamando la definizione della **misura di Hausdorff unidimensionale**  $\mathcal{H}^1$ : in questo modo il funzionale  $\mathcal{MS}$  ha senso per ogni insieme chiuso  $K \subset \bar{\Omega}$ .

Vogliamo minimizzare il funzionale  $\mathcal{MS}$  sulla classe di tutte le coppie ammissibili  $(K, u)$ , con  $K$  chiuso e  $u$  regolare su  $\Omega \setminus K$ . Nel seguito chiameremo **segmentazione ottimale** ogni  $K$  che corrisponde ad un minimo di  $\mathcal{MS}$  e **immagine segmentata** la funzione correlata  $u$ .

Il funzionale  $\mathcal{MS}$  è stato introdotto nell'ambito della computer vision in [4]. L'insieme  $\Omega$  rappresenta lo schermo, il dato  $g$  è un'immagine digitale. Più precisamente, in tale contesto,  $g$  è l'immagine in ingresso e  $g(x_1, x_2)$  descrive il livello di intensità della luce nel punto  $(x_1, x_2) \in \Omega$ . Nella minimizzazione del funzionale  $\mathcal{MS}$ , l'immagine in ingresso (input)  $g$  si trasforma in una

immagine in uscita (output)  $u$  regolare fuori dell'insieme incognito  $K$ : il funzionale  $\mathcal{MS}$  penalizza grandi insiemi  $K$  e, al di fuori della loro unione, si richiede che la funzione  $u$  sia regolare (*denoised*) e "vicina" al dato  $g$ . Il primo termine forza  $u$  ad essere il più regolare possibile in un aperto di  $\Omega$ , il secondo termine impone una penalizzazione (in norma  $L^2(\Omega)$ ) per la deviazione di  $u$  da  $g$ , il terzo termine impedisce che l'insieme singolare  $K$  sia troppo lungo.

La scelta dei due parametri  $\mu$  e  $\alpha$  corrisponde a fissare soglie di scala e di contrasto. Più precisamente la sensibilità al contrasto è di  $(4\alpha^2\mu)^{1/4}$ , la scala è di  $\mu^{-1/2}$ , il limite di pendenza è di  $(\alpha^2\mu/4)^{1/4}$  (valori di pendenza sopra questa soglia vengono recuperati in modo non corretto introducendo una discontinuità artificiale) la resistenza al rumore è  $\alpha\mu$  ([5]).

La difficoltà principale per minimizzare il funzionale  $\mathcal{MS}$  è dovuta alla presenza dell'insieme incognito  $K$ . Se l'insieme chiuso  $K$  fosse assegnato, allora la funzione  $u$  per la quale  $\mathcal{MS}$  raggiunge il suo minimo, sarebbe l'unica soluzione  $u(K)$  del seguente problema di Neumann

$$\begin{cases} -\Delta u + \mu(u - g) = 0 & \text{in } \Omega \setminus K \\ \frac{\partial u}{\partial n} = 0 & \text{in } \partial\Omega \cup K \end{cases} \quad (2)$$

dove  $\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2}$  è l'operatore di Laplace di  $u$ ,  $\partial\Omega$  è la frontiera di  $\Omega$  e  $\frac{\partial u}{\partial n}$  è la derivata normale di  $u$ . Pertanto potremmo porre  $\mathcal{MS}(K) = \mathcal{MS}(K, u(K))$ , sicché il funzionale dipenderebbe dal solo insieme chiuso  $K$ .

D'altra parte, possiamo supporre  $\text{meas}(K) = 0$ , altrimenti  $\mathcal{MS}(K, u) = +\infty$ . In questo caso la funzione  $u$  è definita quasi ovunque in  $\Omega$  e  $K$  contiene l'insieme di discontinuità  $K(u)$  di  $u$ , quindi  $u$  è l'unica variabile significativa e  $\mathcal{MS}(u) = \mathcal{MS}(K(u), u)$ : in questo approccio l'insieme  $K(u)$  non è necessariamente chiuso.

Questa seconda strategia, proposta da De Giorgi ([3]) è molto generale, e ha portato a molti risultati ([7], [8]), le cui dimostrazioni sono alquanto tecniche e la loro descrizione va oltre gli scopi di questa esposizione. Tuttavia, per una migliore comprensione, alcune definizioni e nozioni basilari sono richiamate nell'ultima sezione nell'ambito dei problemi con discontinuità

## La $\Gamma$ -convergenza

Per chiarire la convergenza di tipo variazionale di  $\mathcal{MS}_\varepsilon$  ricordiamo la nozione di  $\Gamma$ -convergenza introdotta da Ennio De Giorgi (cfr. [11]): dato uno spazio metrico  $V$  e i funzionali

$$F_\varepsilon : V \rightarrow \mathbb{R} \cup \{+\infty\} \quad \varepsilon > 0$$

diciamo che  $F_\varepsilon$   $\Gamma$ -converge a un funzionale  $F$  definito su  $V$  se, per ogni  $v \in V$ :

$$\forall(v_\varepsilon) : v_\varepsilon \rightarrow v \implies F(v) \leq \liminf_{\varepsilon \rightarrow 0} F_\varepsilon(v_\varepsilon)$$

$$\exists(v_\varepsilon) : v_\varepsilon \rightarrow v \text{ tale che } F(v) \geq \limsup_{\varepsilon \rightarrow 0} F_\varepsilon(v_\varepsilon)$$

La principale proprietà della  $\Gamma$ -convergenza è la seguente:

se  $F_\varepsilon$   $\Gamma$ -converge a  $F$ ,  $v_\varepsilon$  sono minimizzanti di  $F_\varepsilon$  e  $v_\varepsilon \rightarrow v$  allora, sotto ipotesi generali sui funzionali  $F_\varepsilon$ ,  $v$  è un punto di minimo di  $F$  e  $\lim_{\varepsilon \rightarrow 0} F(v_\varepsilon) = F(v)$ .

libere.

Va osservato che non ci si può aspettare l'unicità, considerata la dipendenza della segmentazione ottimale dai parametri  $\mu$  e  $\alpha$ . Quando il parametro  $\alpha$  è molto grande, allora la segmentazione ottimale  $K$  risulta essere vuota. D'altra parte, se  $\alpha$  è molto piccola c'è un ampio insieme singolare  $K$ , a meno che non siamo in un caso banale ( $g$  costante). Un argomento di continuità mostra che possiamo trovare un valore limite di  $\alpha$  che implica l'esistenza di una segmentazione ottimale  $K$  non vuota e allo stesso tempo mantiene anche l'insieme vuoto come segmentazione ottimale.

### Proprietà dei minimi di $\mathcal{MS}$

Il seguente teorema di esistenza è stato dimostrato in [8], [9].

**Teorema (Esistenza di soluzioni).** Sia  $\Omega \subset \mathbb{R}^2$  un aperto. Poniamo

$$\begin{cases} \mu > 0, \alpha > 0, \Omega \text{ limitato,} \\ g \text{ limitata e misurabile in } \Omega. \end{cases} \quad (3)$$

Allora esiste almeno una coppia tra i sottoinsiemi chiusi  $K$  di  $\mathbb{R}^2$  e  $u \in C^1(\Omega \setminus K)$ , che minimizza il funzionale  $\mathcal{MS}$  con energia finita. Inoltre  $\sup |u| \leq \sup |g|$  e l'insieme  $K \cap \Omega$  è l'unione (al più numerabile) di archi regolari.

La dimostrazione del teorema, come già detto, è piuttosto lunga e tecnica ed è stata ottenuta con

i metodi diretti del Calcolo delle Variazioni. Una versione debole del problema di minimo posto, precisamente la minimizzazione del funzionale  $\mathcal{F}$  che dipende solo da  $u$  (per la definizione di  $\mathcal{F}$  rinviamo all'ultima sezione), è risolta dimostrando semicontinuità inferiore e compattezza nello spazio funzionale delle immagini con energia finita ([10]). Successivamente, provando la maggiore regolarità dei minimi deboli (concetto opportunamente affinato al fine di essere "essenziale") otteniamo una soluzione del problema iniziale ([8], [9]).

Per descrivere le proprietà di una segmentazione ottimale  $K$  usiamo la seguente notazione: per ogni  $x \in \mathbb{R}^2$  e per ogni numero reale  $r > 0$  indichiamo con  $B_r(x)$  il disco di centro  $x$  e raggio  $r$ , ovvero

$$B_r(x) = \{y \in \mathbb{R}^2 : \|y - x\| < r\}$$

Se  $(K, u)$  è una coppia minimizzante per  $\mathcal{MS}$ , allora possiamo aggiungere a  $K$  ogni insieme chiuso con lunghezza nulla ottenendo una nuova coppia ottimale  $(K', u)$ . Per questo motivo è utile introdurre la nozione di **coppia minimizzate essenziale** che ha la proprietà seguente:  $\mathcal{H}^1(K \cap B_r(x)) > 0$  per ogni  $x \in K$  e  $B_r(x) \subset \Omega$ .

Le quattro proposizioni che seguono, relative a una coppia essenziale minimizzante  $(K, u)$ , possono essere dimostrate con l'ipotesi (3) ([2]). Per semplicità consideriamo  $|g(x_1, x_2)| \leq 1$ . Con il termine *densità media* intendiamo il rapporto

tra la misura della parte di insieme singolare all'interno di un disco e il raggio del disco.

**Limitazione dall'alto per la densità media**

Per ogni  $x \in \Omega$ , per ogni  $0 < r \leq 1$ , con  $B_r(x) \subset \Omega$ , risulta

$$\mathcal{H}^1(K \cap B_r(x)) \leq \left(2 + \frac{\mu}{\alpha}\right) \pi r$$

**Limitazione dal basso per la densità media.**

Esiste  $\delta > 0$  tale che per ogni  $x \in K$ , per ogni  $0 < r \leq 1$  con  $B_r(x) \subset \Omega$  risulta

$$\mathcal{H}^1(K \cap B_r(x)) \geq \delta r$$

**Proprietà di concentrazione uniforme.** Per ogni  $\varepsilon > 0$  esiste  $\delta(\varepsilon) > 0$  tale che per ogni  $x \in K$ , per ogni  $R \leq 1$  con  $B_R(x) \subset \Omega$ , esiste  $B_r(y) \subset B_R(x)$  con  $r \geq \delta(\varepsilon)R$  e

$$\mathcal{H}^1(K \cap B_r(y)) \geq (2 - \varepsilon)r$$

Le precedenti proprietà esprimono il fatto che un algoritmo idoneo può eliminare le parti di  $K$  che sono essenzialmente isolate, dal momento che la teoria garantisce che con questa eliminazione la segmentazione migliora, nel senso (variazionale) che il valore del funzionale  $\mathcal{MS}$  diminuisce.

**Contenuto di Minkowski della segmentazione.** Per ogni aperto  $\Omega' \subset\subset \Omega$  vale la seguente uguaglianza

$$\lim_{r \rightarrow 0} \frac{|\{x \in \Omega'; d(x, K \cap \Omega') < r\}|}{2r} = \mathcal{H}^1(K \cap \Omega')$$

La precedente uguaglianza esprime la relazione tra la misura di Hausdorff e il **contenuto di Minkowski** di  $K$ . Questa è una informazione utile per le applicazioni; permette, ad esempio, di utilizzare lo schema di approssimazione introdotto in [12] che produce gli algoritmi sviluppati successivamente.

**Approssimazione variazionale di  $\mathcal{MS}$**

Minimizzare  $\mathcal{MS}$  è abbastanza difficile a causa della difficoltà dovuta alla presenza dell'insieme di discontinuità libera  $K$ . Dal momento che  $K$  è incognito, è stato necessario sviluppare nuovi metodi per trovare soluzioni numeriche.

Il primo approccio alla minimizzazione numerica di  $\mathcal{MS}$  fu proposto e attuato da Blake & Zisserman nel libro [5]: "the Graduated

Non-Convexity algorithm", che corrisponde ad un rilassamento della parte non-convessa del funzionale.

Qui descriviamo un altro metodo, proposto in [12] e implementato numericamente in [13] e in [14].

L'idea principale di questo metodo consiste nell'introdurre una funzione ausiliaria  $z$ , che assume valori prossimi a 1 lontano da  $K$  e valori prossimi a 0 vicino a  $K$ . Consideriamo, allora, i seguenti funzionali dipendenti da una coppia di funzioni (funzionali più gestibili, dal punto di vista numerico, di quanto lo sia  $\mathcal{MS}$ ): con  $\varepsilon > 0$ :

$$\begin{aligned} \mathcal{MS}_\varepsilon(z, u) = & \iint_{\Omega} (z^2 \|Du\|^2 + \mu|u - g|^2) dx_1 dx_2 \\ & + \alpha \iint_{\Omega} \left( \varepsilon \|Dz\|^2 + \frac{(z - 1)^2}{4\varepsilon} \right) dx_1 dx_2 \end{aligned} \tag{4}$$

L'ultimo addendo forza  $z$  a essere quasi ovunque uguale a 1, mentre il secondo integrale approssima la lunghezza di  $K$  nella minimizzazione di  $\mathcal{MS}_\varepsilon$ , per  $\varepsilon$  che tende a 0.

Presi  $V = L^2(\Omega) \times L^2(\Omega)$ ,  $F_\varepsilon(z, u) = \mathcal{MS}_\varepsilon(z, u)$  e

$$F(z, u) = \begin{cases} \mathcal{F}(u) & \text{se } z = 1 \text{ q.o. in } \Omega \\ +\infty & \text{altrove} \end{cases}$$

dove  $\mathcal{F}$  è una versione debole del funzionale  $\mathcal{MS}$  (vedi ultima sezione), sussiste il seguente risultato [12].

**Teorema.** Poniamo (3). Allora i funzionali  $\mathcal{MS}_\varepsilon$   $\Gamma$ -convergono a  $F$  per  $\varepsilon \rightarrow 0^+$ .

Eseguendo variazioni regolari e con supporto compatto in un intorno di una coppia minimizzante  $(z, u)$  del funzionale  $\mathcal{MS}_\varepsilon$  troviamo le equazioni di Eulero in  $\Omega$

$$\begin{cases} z^2 \Delta u + 2z Dz \cdot Du = \mu(u - g) \\ \Delta z = \frac{z - 1}{4\varepsilon^2} + \frac{z}{\varepsilon\alpha} \|Du\|^2 \\ \frac{\partial u}{\partial n} = \frac{\partial z}{\partial n} = 0, \quad \text{in } \partial\Omega \end{cases}$$

Questo sistema ora può essere risolto numericamente con schemi alle differenze finite (vedi [13], [14]).

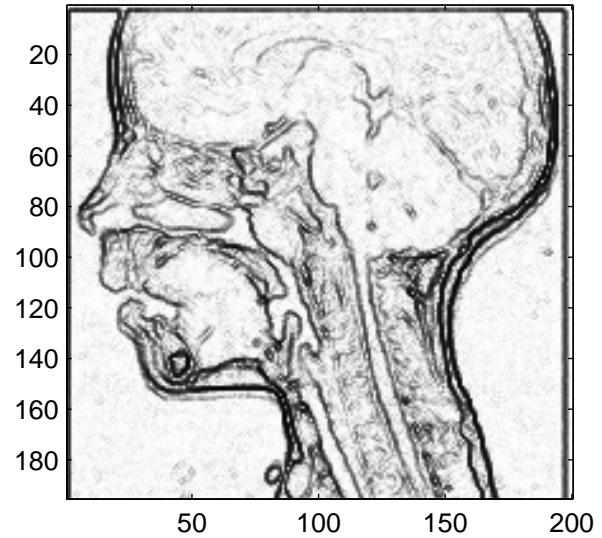
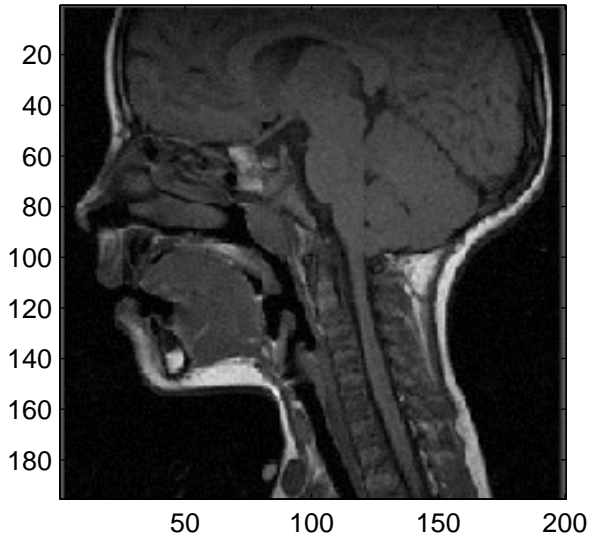


Figura 1: Segmentazione di una immagine ottenuta utilizzando il funzionale  $MS$

### Equazioni di Eulero per $MS$ , regolarità dell'insieme singolare ottimale e una congettura di E. De Giorgi

Oltre a (2), ulteriori condizioni di Eulero sono state ottenute per i minimi di  $MS$  ([4], [15]).

**Curvatura di  $K$  e salto del gradiente al quadrato.** Sia  $(K, u)$  un minimo di  $MS$  in  $\Omega$ ,  $g \in C^1(\Omega)$  e  $B \subset \Omega$  un disco aperto tale che  $K \cap B$  è il grafico di una funzione  $C^2$  e sia  $B_+$  (risp.  $B_-$ ) l'epigrafico (risp. sottografico) aperto connesso di tale funzione in  $B$ . Posto  $u \in C^1(\overline{B_+}) \cap C^1(\overline{B_-})$ , risulta

$$\llbracket \|Du\|^2 + \mu|u-g|^2 \rrbracket = \alpha \text{curv}(K) \quad \text{su } K \cap B$$

dove  $\llbracket \cdot \rrbracket$  denota la differenza delle tracce su  $B_+$  e  $B_-$ , e  $\text{curv}$  è la curvatura valutata orientando la curva con la normale che punta verso  $B_+$ .

Il seguente enunciato relativo a una **Congettura di Mumford & Shah ([4])** sulla regolarità dell'insieme singolare ottimale per  $MS$  non è stato ancora completamente dimostrato:

Sia  $(K, u)$  una coppia essenziale minimizzante  $MS$ , allora  $K$  è localmente in  $\Omega$  l'unione di un numero finito di archi  $C^{1,1}$ .

Ulteriori proprietà geometriche dei punti terminali di una segmentazione ottimale sono state studiate ([4], [15]): le condizioni ottenute avevano suggerito la seguente congettura (per ora parzialmente dimostrata, vedi [16])

#### Congettura (De Giorgi [3])

L'unica minimizzante locale non costante della parte principale del funzionale  $MS$  è la coppia  $(K, u)$ , con

$K = \{(x_1, 0); x_1 \leq 0\}$  e la funzione  $u$ , espressa in coordinate polari, data da

$$u(\rho, \theta) = \sqrt{\frac{2\alpha\rho}{\pi}} \sin \frac{\theta}{2} \quad \rho \geq 0, -\pi < \theta < \pi \quad (5)$$

(a meno del segno, di moti rigidi in  $\mathbb{R}^2$  e di costanti additive).

### Il funzionale di Blake & Zisserman

Lo studio di funzionali da minimizzare dipendenti da una energia di massa del secondo ordine e una discontinuità (incognita) di superficie (di linea, nel caso bidimensionale) ha attirato l'interesse in relazione a problemi di teoria della frattura, partizioni ottimali, piastre elasto-plastiche ([17]) e nella segmentazione di immagini.

Consideriamo ora il funzionale dipendente dalle derivate del secondo ordine ([5], [18])

$$\begin{aligned} \mathcal{BZ}(K_0, K_1, u) := & \iint_{\Omega \setminus (K_0 \cup K_1)} (\|D^2u\|^2 + \mu|u-g|^2) dx_1 dx_2 \\ & + \alpha \text{length}(K_0 \cap \Omega) \\ & + \beta \text{length}((K_1 \setminus K_0) \cap \Omega) \end{aligned} \quad (6)$$

dove  $\Omega \subset \mathbb{R}^2$  è un aperto limitato e  $\|D^2u\|$  denota la norma euclidea della matrice Hessiana  $D^2u = \left( \frac{\partial^2 u}{\partial x_i \partial x_j} \right)_{i,j=1,2}$  di  $u$ . I numeri reali positivi  $\alpha, \beta, \mu$  e la funzione limitata  $g$  sono assegnati, mentre  $K_0, K_1 \subset \mathbb{R}^2$  sono insiemi chiusi e  $u$  è



una funzione tale che  $u \in C^2(\Omega \setminus (K_0 \cup K_1)) \cap C^0(\Omega \setminus K_0)$ .

Gli insiemi chiusi  $K_0, K_1$  e la funzione  $u$  sono le incognite da trovare come minimi di (6) tra tutte le possibili terne ammissibili.

Il funzionale  $\mathcal{BZ}$  stima il grado di corrispondenza tra l'immagine  $g$  e una sua immagine segmentata  $u$ ;  $K_0 \cup K_1$  è la relativa segmentazione. Tale funzionale è stato introdotto da Blake & Zisserman (cfr. [5]) come una forma di energia da minimizzare per realizzare una segmentazione ottimale di un'immagine monocromatica.

La dipendenza dalle derivate seconde  $D^2u$  (invece che dalle derivate prime  $Du$ , come nel funzionale di Mumford-Shah) consente di rilevare anche l'insieme delle pieghe. Inoltre questo funzionale evita l'inconveniente dell'effetto rampa causato dalla sovra-segmentazione dei salti (che è la comparsa di una o più discontinuità spurie nell'immagine di output  $u$ , manifestate invece dal funzionale  $\mathcal{MS}$ , quando il dato  $g$  ha una pendenza abbastanza ripida).

Le discontinuità di  $u$  e di  $Du$  si evidenziano rispettivamente con gli insiemi  $K_0, K_1$  che sono "a priori" incogniti; per questo il problema di minimo associato risulta essere essenzialmente non-convesso, e per alcuni dati  $g$  può dar luogo a non unicità dei minimi. Un'altra difficoltà nell'analisi matematica del funzionale  $\mathcal{BZ}$  è il fatto che (6) non controlla le derivate prime, e inoltre il troncamento delle funzioni in competizione non riduce l'energia, mentre nel caso del funzionale  $\mathcal{MS}$  il troncamento riduce l'energia.

### Proprietà dei minimi di $\mathcal{BZ}$

In [18] è stato dimostrato il seguente risultato.

**Teorema (Esistenza di soluzioni).** *Sia  $\Omega \subset \mathbb{R}^2$  un aperto. Poniamo*

$$\begin{cases} \mu > 0, 0 < \beta \leq \alpha \leq 2\beta, \Omega \text{ limitato,} \\ g \text{ limitata e misurabile in } \Omega. \end{cases} \quad (7)$$

Allora esiste almeno una terna tra i sottoinsiemi di Borel  $K_0, K_1 \subset \mathbb{R}^2$  con  $K_0 \cup K_1$  chiuso e  $u \in C^2(\Omega \setminus (K_0 \cup K_1))$ ,  $u$  continua in senso approssimato su  $\Omega \setminus K_0$ , che minimizza il funzionale  $\mathcal{BZ}$  con energia finita. Inoltre gli insiemi  $K_0 \cap \Omega$  e  $K_1 \cap \Omega$  sono unione (al più numerabile) di archi regolari. La dimostrazione è stata ottenuta con i metodi di-

retti del Calcolo delle Variazioni. Una versione debole del problema posto, cioè la minimizzazione del funzionale  $\mathcal{G}$  che dipende solo da  $u$  (la definizione di  $\mathcal{G}$  è rinviata all'ultima sezione) è risolta, provando semicontinuità inferiore e compattezza nello spazio funzionale di immagini con energia finita ([19]). Una soluzione al problema di partenza è ottenuta successivamente ([18]), provando la maggiore regolarità dei minimi deboli (concetto opportunamente affinato al fine di essere "essenziale").

Le quattro proposizioni che seguono, relative a una terna essenziale minimizzante, sono state dimostrate (cfr. [20]) con l'ipotesi (7). Per semplicità assumiamo  $|g(x_1, x_2)| \leq 1$ .

**Limitazione dall'alto per la densità media.** *Per ogni  $x \in \Omega$ , per ogni  $0 < r \leq 1$  con  $B_r(x) \subset \Omega$ , risulta*

$$\mathcal{H}^1((K_0 \cup K_1) \cap B_r(x)) \leq \left(4 + \frac{\mu}{\beta}\right) \pi r$$

**Limitazione dal basso per la densità media.** *Esistono  $\varepsilon_1 > 0, \varrho_1 > 0$  tali che per ogni  $x \in K_0 \cup K_1$ , per ogni  $0 < r \leq \varrho_1$  con  $B_r(x) \subset \Omega$ , risulta*

$$\mathcal{H}^1((K_0 \cup K_1) \cap B_r(x)) \geq \varepsilon_1 r$$

### Proprietà dell'eliminazione

*Siano  $\varepsilon_1 > 0, \varrho_1 > 0$  come nel precedente enunciato e  $0 < r \leq \varrho_1$ . Se  $x \in \Omega$  e*

$$\mathcal{H}^1((K_0 \cup K_1) \cap B_r(x)) < \frac{\varepsilon_1}{2} r$$

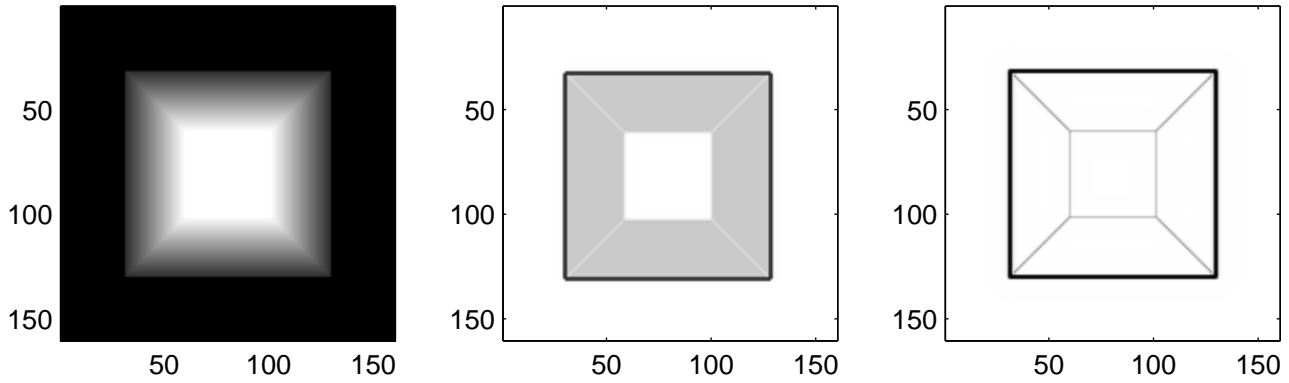
*allora*

$$(K_0 \cup K_1) \cap B_{r/2}(x) = \emptyset$$

Questa proprietà è un'informazione utile per l'analisi numerica del problema, nel senso che un idoneo algoritmo può eliminare tali parti isolate di  $K_0 \cup K_1$  ai fini di una segmentazione ottimale.

In considerazione di queste caratteristiche, possiamo affermare che la segmentazione ottimale  $(K_0, K_1, u)$  ridisegna uno schema omogeneo di linee essenziali dell'immagine iniziale  $g$ . In questo modo la segmentazione ottimale porta più chiaramente in luce il significato semantico insito in  $g$ .

**Contenuto di Minkowski della segmentazione.** *Per ogni insieme aperto  $\Omega' \subset\subset \Omega$  vale la*



**Figura 2:** Segmentazione di una immagine artificiale ottenuta utilizzando i funzionali  $\mathcal{MS}$  e  $\mathcal{BZ}$ . Nell'immagine centrale è evidenziato solo l'insieme di discontinuità  $K_0$ , con un effetto di "sovrasegmentazione". In quella a destra sono evidenziate le discontinuità  $K_0$  e anche le pieghe  $K_1$ .

seguinte uguaglianza

$$\lim_{r \rightarrow 0} \frac{|\{x \in \Omega; d(x, (K_0 \cup K_1) \cap \Omega') < r\}|}{2r} = \mathcal{H}^1((K_0 \cup K_1) \cap \Omega')$$

Anche questa informazione è importante per il calcolo numerico dei minimi, al fine di approssimare, nel senso della  $\Gamma$ -convergenza, il funzionale  $\mathcal{BZ}$  con funzionali ellittici per i quali si possono trovare algoritmi numerici efficienti.

### Approssimazione variazionale di $\mathcal{BZ}$

In analogia a quanto sviluppato per l'approssimazione di  $\mathcal{MS}$ , consideriamo due funzioni ausiliarie  $s, \sigma$  e definiamo i funzionali ([21])

$$\begin{aligned} \mathcal{BZ}_\varepsilon(s, \sigma, u) = & \iint_{\Omega} ((\sigma^2 + \kappa_\varepsilon) \|D^2 u\|^2 \\ & + \xi_\varepsilon (s^2 + \zeta_\varepsilon) \|Du\|^2 + \mu |u - g|^2) dx_1 dx_2 \\ & + (\alpha - \beta) \iint_{\Omega} \left( \varepsilon \|Ds\|^2 + \frac{(s-1)^2}{4\varepsilon} \right) dx_1 dx_2 \\ & + \beta \iint_{\Omega} \left( \varepsilon \|D\sigma\|^2 + \frac{(\sigma-1)^2}{4\varepsilon} \right) dx_1 dx_2 \end{aligned} \quad (8)$$

che approssimano  $\mathcal{BZ}$  nel senso della  $\Gamma$ -convergenza, per  $\varepsilon \rightarrow 0^+$ , sotto opportune scelte degli infinitesimi  $\kappa_\varepsilon, \xi_\varepsilon, \zeta_\varepsilon$ .

La minimizzazione dei funzionali  $\mathcal{BZ}_\varepsilon$  è stata implementata numericamente in alcuni casi ([21]).

Le pieghe sono evidenziate meglio da  $\mathcal{BZ}$  rispetto a  $\mathcal{MS}$ , come mostrato in fig. 2.

### Equazioni di Eulero per $\mathcal{BZ}$ e una congettura sul candidato ottimale

**Condizioni necessarie su  $K_0$ .** Assumiamo (7),  $(K_0, K_1, u)$  sia un minimo locale di  $\mathcal{BZ}$ ,  $B \subset \Omega$  un disco aperto e, per semplicità,  $K_0 \cap B$  sia il diametro di  $B$  parallelo all'asse  $x_1$ ,  $(K_1 \setminus K_0) \cap B = \emptyset$  e  $u \in C^3(\overline{B_+}) \cap C^3(\overline{B_-})$ . Allora le condizioni sui due lati di  $K_0$  sono:

$$\left( \frac{\partial^2 u}{\partial x_2^2} \right)^\pm = 0, \quad \left( \frac{\partial^3 u}{\partial x_2^3} + 2 \frac{\partial^3 u}{\partial x_1^2 \partial x_2} \right)^\pm = 0$$

su  $K_0 \cap B$ .

**Curvatura di  $K_0$  e salto dell'Hessiano al quadrato.** Sia  $(K_0, K_1, u)$  un minimo locale di  $\mathcal{BZ}$  in  $\Omega$ ,  $g \in C^1(\Omega)$  e  $B \subset \Omega$  un disco aperto, tale che  $K_0 \cap B$  è il grafico di una funzione  $C^3$  e  $B_+$  (risp.  $B_-$ ) l'epigrafico (risp. sottografico) aperto connesso di tale funzione in  $B$ . Assumiamo  $K_1 \cap B = \emptyset$  e  $u \in C^3(\overline{B_+}) \cap C^3(\overline{B_-})$ . Allora

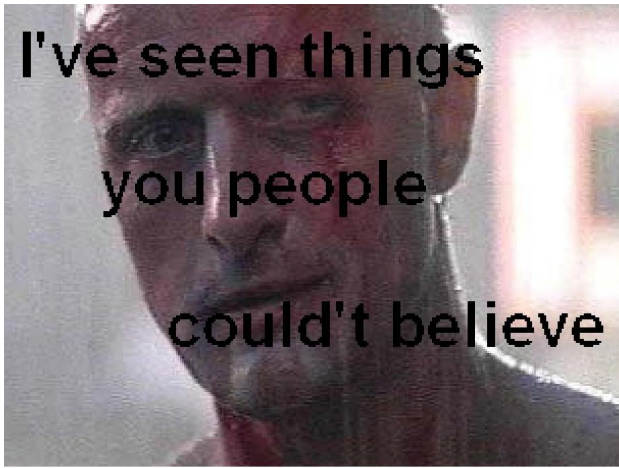
$$\left[ \|D^2 u\|^2 + \mu |u - g|^2 \right] = \alpha \text{curv}(K_0) \quad \text{su } K_0 \cap B$$

dove  $\text{curv}$  è la curvatura valutata orientando la curva con la normale che punta verso  $B_+$ .

Condizioni necessarie circa la curvatura di  $K_1$  sono anche provate in ([22]).

Una analisi fine del comportamento della segmentazione ottimale nei punti terminali dell'insieme singolare è stata fatta in [23], ottenendo condizioni integrali e geometriche nei **crack-tip** (cioè nei punti terminali interni di  $K_0$ ) e nei **crease-tip** (punti terminali interni di  $K_1$ ).

Dall'analisi della parte principale del funzionale



**Figura 3:** *Inpainting di una immagine ottenuta utilizzando il funzionale  $BZ$*

di Blake & Zissermann in  $\mathbb{R}^2$  è emerso ([22]) un candidato minimizzante locale in  $\mathbb{R}^2$ , che ha un insieme di salto non vuoto.

Tale candidato è  $(K_0, K_1, W)$  con

$$K_0 = \{(x_1, 0); x_1 \leq 0\}, \quad K_1 = \emptyset \quad (9)$$

e la funzione  $W$ , espressa in coordinate polari in  $\mathbb{R}^2$  con  $\theta \in (-\pi, \pi)$ , è:

$$W = \sqrt{\frac{\alpha\rho^3}{193\pi}} \left( \sqrt{21} \left( \sin \frac{\theta}{2} - \frac{5}{3} \sin \left( \frac{3}{2}\theta \right) \right) + \left( \cos \frac{\theta}{2} - \frac{7}{3} \cos \left( \frac{3}{2}\theta \right) \right) \right) \quad (10)$$

Il seguente elenco di proprietà mostra esplicitamente che  $W$  soddisfa tutte le condizioni necessarie provate e un principio variazionale di equi-partizione dell'energia di massa e di linea:

$$\left\{ \begin{array}{l} \Delta^2 W = 0 \quad \text{su } \mathbb{R}^2 \setminus K_0 \\ \left( \frac{\partial^2 W}{\partial x_2^2} \right)^\pm = 0, \quad \left( \frac{\partial^3 W}{\partial x_2^3} + 2 \frac{\partial^3 W}{\partial x_1^2 \partial x_2} \right)^\pm = 0 \\ \text{e } W^\pm = \pm \frac{8}{3} \sqrt{\frac{21\alpha\rho^3}{193\pi}} \quad \text{su } K_0 \\ \int_{B_\rho(\mathbf{0})} \|D^2 W\|^2 dx_1 dx_2 = \alpha\rho \\ \qquad \qquad \qquad = \alpha \mathcal{H}^1(K_0 \cap B_\rho(\mathbf{0})) \end{array} \right.$$

**Congettura** ([22]) *La terna  $(K_0, K_1, W)$ , definita in (9) (10) è un minimo locale della parte principale del funzionale di Blake-Zissermann in  $\mathbb{R}^2$ , e non esistono altri minimi locali non banali (a meno del segno, di moti rigidi in  $\mathbb{R}^2$  e dell'aggiunta di funzioni*

*affini).*

## Cenno al problema dell'inpainting

Il problema dell'inpainting consiste nel cercare di ricostruire una parte deteriorata oppure mancante di un'immagine, in maniera visivamente accettabile. Anche per questo problema è stata utilizzata una opportuna versione del funzionale  $MS$  ([24]) e del funzionale  $BZ$  ([25]). Nell'esempio in Figura 3 è stato rimosso il testo sovrascritto (l'immagine è a colori ed è stata trattata usando la versione di  $BZ$  per le funzioni vettoriali). Per brevità non ci soffermiamo ulteriormente su questo problema.

## Problemi con discontinuità libere

I problemi di minimo per i funzionali  $MS$  e  $BZ$  si inquadrano in una più ampia classe di problemi del Calcolo delle Variazioni, quella dei **Problemi con discontinuità libere**, in cui tra le incognite figurano, oltre che funzioni, anche insiemi che non sono necessariamente frontiere.

L'approccio generale introdotto da E. De Giorgi in [3] per risolvere i problemi con discontinuità libere è una nuova applicazione dei Metodi Diretti nel Calcolo delle Variazioni. L'idea iniziale è la **formulazione debole del problema di minimo** in una classe opportuna di funzioni, denominata  $SBV(\Omega)$ , i cui elementi ammettono discontinuità solo lungo insiemi di dimensione uno. I passi successivi, più delicati, consistono nel dimostrarlo

re la regolarità degli insiemi ottimali con discontinuità libere, per recuperare una soluzione del problema di minimo originario.

Riportiamo qui la definizione dello spazio  $BV$  di **funzioni con variazione limitata**, le cui derivate distribuzionali sono misure con variazione limitata, e alcune proprietà :

$v \in BV(\Omega)$  se e solo se  $v \in L^1(\Omega)$  (i.e.  $v$  è sommabile secondo Lebesgue) e la sua variazione totale

$$\|Dv\|_{\mathcal{M}(\Omega)} := \sup \left\{ \iint_{\Omega} v \operatorname{div} \phi \, dx_1 dx_2, \phi \in C_0^1(\Omega, \mathbb{R}^2), \|\phi\| \leq 1 \right\}$$

è finita. Per ogni funzione con variazione limitata  $v : \Omega \rightarrow \mathbb{R}$  e  $x \in \Omega$ ,  $z \in \mathbb{R}$ , poniamo  $z = \operatorname{ap} \lim_{y \rightarrow x} v(y)$  ( $z$  è il **limite approssimato** di  $v$  in  $x$ ) se

$$\lim_{\rho \rightarrow 0} \frac{1}{\pi \rho^2} \iint_{B_\rho(0)} v(x+y) \, dy_1 dy_2 = z$$

Per  $v \in S^1$ , denotiamo con  $v^+ = \operatorname{tr}^+(x, v, \nu)$  (e  $v^- = \operatorname{tr}^+(x, v, -\nu)$ ) se

$$\lim_{\rho \rightarrow 0} \frac{1}{\pi \rho^2} \iint_{B_\rho(0) \cap \{y \cdot \nu > 0\}} v(x+y) \, dy_1 dy_2 = v^+$$

L'insieme  $S_v = \{x \in \Omega : \nexists \operatorname{ap} \lim_{y \rightarrow x} v(y)\}$  è detto l'insieme singolare di  $v$ ;  $Dv$  denota il gradiente distribuzionale di  $v$  e  $\nabla v$  il **gradiente approssimato** di  $v$  (cioè la parte assolutamente continua di  $Dv$ ). Quando ha significato, poniamo  $\nabla^2 v = \nabla(\nabla v)$ . Ricordiamo inoltre le definizioni di alcune classi di funzioni aventi derivate che sono misure speciali nel senso di De Giorgi, e alcune proprietà :

$$SBV(\Omega) := \left\{ v \in BV(\Omega) : \|Dv\|_{\mathcal{M}(\Omega)} = \iint_{\Omega} \|\nabla v\| \, dy_1 dy_2 + \int_{S_v} |v^+ - v^-| \, d\mathcal{H}^1 \right\}$$

Per una generica funzione  $BV$  la variazione totale avrebbe un terzo addendo, detto parte cantoriana, che è escluso nel caso delle funzioni  $SBV$ .

$$GSBV(\Omega) := \left\{ v : \Omega \rightarrow \mathbb{R} ; -k \vee v \wedge k \in SBV_{loc}(\Omega) \forall k \in \mathbb{N} \right\}$$

$$GSBV^2(\Omega) := \{v \in GSBV(\Omega), \nabla v \in (GSBV(\Omega))^2\}$$

Le classi di funzioni  $GSBV(\Omega)$ ,  $GSBV^2(\Omega)$  non sono spazi vettoriali, e non sono sottoinsiemi di distribuzioni in  $\Omega$ , tuttavia variazioni regolari di una funzione di  $GSBV^2(\Omega)$  appartengono alla stessa classe. Se  $v \in SBV(\Omega)$  o  $GSBV(\Omega)$ , allora  $S_v$  è l'unione (al più numerabile) di archi regolari e  $\nabla v$  esiste quasi ovunque in  $\Omega$ .

### Formulazione debole del funzionale di Mumford & Shah ([10]).

Sia  $\Omega \subset \mathbb{R}^2$  un aperto; con l'ipotesi (3), definiamo il funzionale  $\mathcal{F} : SBV(\Omega) \rightarrow [0, +\infty]$  ponendo

$$\mathcal{F}(u) := \iint_{\Omega} (\|\nabla u\|^2 + \mu|u-g|^2) \, dx_1 dx_2 + \alpha \mathcal{H}^1(S_u)$$

La segmentazione ottimale per il modello di Mumford e Shah viene recuperata con la chiusura dell'insieme dei punti di discontinuità (salti) di una funzione che minimizza  $\mathcal{F}$  in  $SBV(\Omega)$ . Infatti, una funzione  $u$  minimizzante  $\mathcal{F}$  fornisce la segmentazione ottimale  $K = \overline{S_u}$  e  $\mathcal{F}(u) = \mathcal{MS}(K, u) = \min \mathcal{MS}$ . L'esistenza di una segmentazione ottimale è stata dimostrata in [8].

### Formulazione debole del funzionale di Blake & Zisserman ([19]).

Sia  $\Omega \subset \mathbb{R}^2$  un aperto; sotto l'ipotesi (7) definiamo il funzionale  $\mathcal{G} : GSBV^2(\Omega) \rightarrow [0, +\infty]$  ponendo

$$\mathcal{G}(u) := \iint_{\Omega} (\|\nabla^2 u\|^2 + \mu|u-g|^2) \, dx_1 dx_2 + \alpha \mathcal{H}^1(S_u) + \beta \mathcal{H}^1(S_{\nabla u} \setminus S_u)$$

La segmentazione ottimale per il modello di Blake e Zisserman viene recuperata con la chiusura dell'insieme dei punti di discontinuità (salti) e dell'insieme dei punti di discontinuità del gradiente (pieghe) di una funzione che minimizza  $\mathcal{G}$  in  $GSBV^2(\Omega)$ . L'esistenza di una segmentazione ottimale è stata dimostrata in [18].



- [1] G. AUBERT, P. KORNPBOST: *Mathematical problems in image processing. Partial differential equations and the calculus of variations.*, II ed., Springer, New York (2006).
- [2] J. M. MOREL, S. SOLIMINI: *Variational Models in Image Segmentation*, Birkhäuser, Basel (1994).
- [3] E. DE GIORGI: *Free discontinuity problems in Calculus of Variations*, in: J. L. Lions, R. Dautray (Eds.), "Frontiers in Pure and Applied Mathematics", North-Holland, Amsterdam (1991).
- [4] D. MUMFORD, J. SHAH: "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems", *Comm. Pure Appl. Math.* **42** (1989) 577-685.
- [5] A. BLAKE, A. ZISSERMAN: *Visual Reconstruction*, The MIT Press, Cambridge (1987).
- [6] M. CARRIERO, A. LEACI, F. TOMARELLI: *Local minimizers for a free gradient discontinuity problem in image segmentation.*, in G. Dal Maso & F. Tomarelli (Eds.), "Variational Methods for Discontinuous Structures", Birkhäuser, Basel (2002).
- [7] E. DE GIORGI, L. AMBROSIO: "Un nuovo tipo di funzionale del Calcolo delle Variazioni", *Atti Accad. Naz. Lincei, s.8* **82** (1988) 199-210.
- [8] E. DE GIORGI, M. CARRIERO, A. LEACI: "Existence theorem for a minimum problem with free discontinuity set", *Arch. Rational Mech. Anal.* **108** (1989) 195-218.
- [9] G. DAL MASO, J. M. MOREL, S. SOLIMINI: "A variational method in image segmentation: existence and approximation results", *Acta Mat.* **168** (1992) 89-151.
- [10] L. AMBROSIO: "A compactness theorem for a special class of functions of bounded variation", *Boll. Un. Mat. It.* **3-B** (1989) 857-881.
- [11] G. DAL MASO: *An introduction to  $\Gamma$ -convergence*, Birkhäuser, Basel (1993).
- [12] L. AMBROSIO, V. M. TORTORELLI: "Approximation of functionals depending on jumps by elliptic functionals via  $\Gamma$ -convergence", *Comm. Pure Appl. Math.* **43** (1991) 999-1036.
- [13] R. MARCH: "Visual reconstruction with discontinuities using variational methods", *Image and vision computing* **10** (1992) 30-38.
- [14] M. CARRIERO, A. FARINA AND I. SGURA: *Image segmentation in the framework of free discontinuity problems*, in: D. Pallara (Ed.), "Calculus of Variations: Topics from the Mathematical Heritage of E. De Giorgi", Quad. Mat., Dept. Math., Seconda Univ. Napoli, Caserta (2004).
- [15] M. CARRIERO, A. LEACI, D. PALLARA, E. PASCALI: "Euler Conditions for a Minimum Problem with Free Discontinuity Surfaces", *Preprint Dip. di Matematica di Lecce* **8** (1988) 1-20.
- [16] G. DAVID: *Singular sets of minimizers for the Mumford-Shah functional*, Birkhäuser, Basel (2005).
- [17] D. PERCIVALE, F. TOMARELLI: "From SBD to SBH: the elastic plastic plate", *Interfaces and Free Boundaries* **4** (2002) 1-29.
- [18] M. CARRIERO, A. LEACI, F. TOMARELLI: "Strong minimizers of Blake & Zisserman functional", *Ann. Scuola Normale Sup. Pisa, s.IV* **25** (1997) 257-285.
- [19] M. CARRIERO, A. LEACI, F. TOMARELLI: *A second order model in image segmentation: Blake & Zisserman functional*, in: R. Serapioni, F. Tomarelli (Eds.), "Variational Methods for Discontinuous Structures", Birkhäuser, Basel (1996).
- [20] M. CARRIERO, A. LEACI, F. TOMARELLI: *Density estimates and further properties of Blake & Zisserman functional*, in: R. Gilbert & Pardalos (Eds.), "From Convexity to Nonconvexity", Kluwer, Amsterdam (2001).
- [21] L. AMBROSIO, L. FAINA, R. MARCH: "Variational approximation of a second order free discontinuity problem in computer vision", *SIAM J. Math. Anal.* **32** (2001) 1171-1197.
- [22] M. CARRIERO, A. LEACI, F. TOMARELLI: "Necessary conditions for extremals of Blake & Zisserman functional", *C. R. Acad. Sci. Paris s.I* **334** (2002) 343-348.
- [23] M. CARRIERO, A. LEACI, F. TOMARELLI: "Euler equations for Blake & Zisserman functional", *Calc.Var. Partial Differential Equations* **32** (2008) 81-110.
- [24] T. F. CHAN AND J. SHEN: "Variational image inpainting", *Comm. Pure Appl. Math.* **LVIII** (2005) 579-619.
- [25] M. CARRIERO, A. LEACI, F. TOMARELLI: "Image inpainting via variational approximation of a Dirichlet problem with free discontinuity", *Advances in Calculus of Variations* (2013) to appear.



**Michele Carriero:** Laurea in matematica presso l'Università di Lecce. È Professore Ordinario di Analisi Matematica presso l'Università del Salento. Si occupa di Calcolo delle Variazioni ed Equazioni alle Derivate Parziali con applicazioni alla Teoria della Visione e alla Teoria dell'Elasticità.

**Antonio Leaci:** Laurea in matematica presso l'Università di Pisa e Diploma presso la Scuola Normale Superiore. È Professore Ordinario di Analisi Matematica presso l'Università del Salento. Si occupa di Calcolo delle Variazioni ed EDP con applicazioni alla Teoria della Visione e alla Teoria dell'Elasticità.

**Franco Tomarelli:** Laurea in matematica presso l'Università di Pisa e Diploma presso la Scuola Normale Superiore. È Professore Ordinario di Analisi Matematica presso il Politecnico di Milano. Si occupa di Calcolo delle Variazioni ed EDP con applicazioni alla Teoria della Visione e alla Meccanica dei Continui.



---

# Il problema isoperimetrico

**Alessio Figalli**

*Department of Mathematics and Institute for Computational Engineering and Sciences  
The University of Texas at Austin*

---

**U**n riferimento al “problema isoperimetrico”, anche detto delle “disuguaglianze isoperimetriche”, si trova già nella mitologia. Didone, regina di Tiro, costretta all’esilio dal fratello Pigmalione, si rifugiò in Nord Africa, nelle terre di Iarba, re dei Getuli, cui chiese non solo asilo ma anche della terra per costruire una nuova città: la futura Cartagine. Iarba, re ospitale ma piuttosto geloso dei suoi possedimenti, concesse alla regina tutta la terra che ella fosse riuscita a ricoprire con una pelle di bue. Didone non si perse affatto d’animo: presa una pelle di bue, iniziò a tagliarla a striscioline sottili, le legò tra di loro e costruì una lunga corda. Ora la questione è: quale forma dare a questa corda per racchiudere la maggior superficie possibile?

Questo è un bel problema matematico, un problema di “isoperimetria”, e consiste nel trovare la figura geometrica che, a parità di perimetro, ha la massima area. È dunque un tipico problema di “calcolo delle variazioni”, ovvero di ricerca di minimo o di massimo.

La soluzione di questi problemi dipende dalle condizioni al contorno: ad esempio, se ci si trova

nell’entroterra allora la forma migliore è un cerchio, mentre, se come nel caso di Didone la terra si affaccia sul mare, è molto meglio scegliere un semicerchio.

Il problema di Didone può essere letto in due modi. Infatti, si può scegliere se fissare la lunghezza della corda e chiedersi quale sia l’area massima che si vuole racchiudere dentro la figura, oppure se fissare l’area e cercare la lunghezza minima della corda che la racchiude. In questa seconda formulazione esso è un classico problema di “ricerca del minimo”.

Lo stesso problema ha senso in dimensione più alta: per esempio, nello spazio tridimensionale, fissato un volume qual è la forma ottimale per contenerlo usando una superficie la cui area sia la più piccola possibile? Le bolle di sapone forniscono la risposta, esse disegnano delle sfere.

Da un punto di vista matematico, formulare questi problemi è non banale. Infatti ci sono molti punti delicati che bisogna affrontare. Innanzitutto, come si misurano il volume di un insieme e il suo perimetro (ossia l’area del suo bordo)? Inoltre, quando si parla di ricerca di un minimo, bisogna definire una classe di elementi all’interno della quale si ricerca tale minimo.

Nel nostro caso, per insiemi regolari è facile definire volume e perimetro, ma il problema è il seguente: se cerchiamo un minimo in una classe troppo ristretta, allora è molto difficile dimostrare che tale minimo esista. Fa quindi parte della

teoria matematica trovare una classe abbastanza grande di insiemi in modo che:

1. tale classe contenga tutti gli insiemi “ragionevoli” che vorremmo ammettere come competitori, per esempio gli insiemi regolari, o quelli il cui bordo è almeno regolare a tratti;
2. tutti gli insiemi della classe abbiano proprietà sufficienti a definire una nozione di perimetro;
3. sia possibile dimostrare che un minimo esiste in questa classe;
4. sia possibile mostrare che, all’interno di tale classe, la palla è il minimo per il problema isoperimetrico.

I concetti adatti a rispondere alle domande qui sopra poste sono stati introdotti da De Giorgi negli anni '50, il quale ha dato una definizione matematicamente molto efficiente e maneggevole di “perimetro” di un insieme.

L’idea è la seguente: se il bordo di un insieme  $E$  è regolare in modo che sia possibile associare una forma di volume  $d\sigma$  sul bordo di  $E$ , allora, indicando con  $P(E)$  il perimetro di  $E$  e con  $\nu_E$  la normale esterna al bordo di  $E$ , per il Teorema di Stokes si ha

$$\begin{aligned} P(E) &= \int_{\partial E} d\sigma(y) \geq \int_{\partial E} X(y) \cdot \nu_E(y) d\sigma(y) \\ &= \int_E \operatorname{div}(X(x)) dx \end{aligned}$$

per ogni campo vettoriale regolare  $X : \mathbb{R}^n \rightarrow \mathbb{R}^n$  tale che  $\|X\|_\infty \leq 1$ .

Inoltre, se si sceglie  $X$  tale che  $X(y) = \nu_E(y)$  su  $\partial E$  (come è sempre possibile fare se  $\partial E$  è abbastanza regolare), allora nella formula sopra abbiamo un’uguaglianza. Quindi

$$P(E) = \sup \int_E \operatorname{div}(X(x)) dx$$

dove il sup è calcolato sui campi vettoriali  $X$  regolari con  $\|X\|_\infty \leq 1$ .

Notiamo ora che per definire il membro di destra non serve nessuna regolarità su  $\partial E$ , dato che stiamo solo integrando una funzione regolare su  $E$ . Quindi l’idea di De Giorgi è stata usare il membro di destra per *definire* il perimetro di un insieme  $E$ .

A questo punto il problema isoperimetrico si può riformulare nella seguente maniera: fissata una quantità di volume  $V > 0$ , l’obiettivo diventa minimizzare il perimetro  $P(E)$  tra tutti gli insiemi  $E$  il cui volume  $|E|$  (più precisamente, la cui misura di Lebesgue) è uguale a  $V$ . In altre parole, si considera

$$\min_{|E|=V} P(E)$$

Con la definizione di perimetro di De Giorgi è possibile usare tecniche classiche del Calcolo della Variazioni per dimostrare l’esistenza di un minimo, che denotiamo con  $\hat{E}_V$ . Il problema isoperimetrico diventa quindi mostrare che  $\hat{E}_V$  è una palla.

Un’osservazione interessante, anche se non strettamente necessaria nel nostro seguito, è la seguente: dato un insieme  $E$  di volume  $V$ , se definiamo  $\lambda_V := V^{-1/n}$  allora l’insieme  $\lambda_V E$  ottenuto dilatando  $E$  del fattore  $\lambda_V$  ha volume 1. Inoltre il suo perimetro è  $P(\lambda_V E) = \lambda_V^{n-1} P(E)$ . In particolare  $P(\hat{E}_V) \leq P(E)$  è equivalente a  $P(\lambda_V \hat{E}_V) \leq P(\lambda_V E)$ , da cui  $\lambda_V \hat{E}_V$  è un minimo del problema isoperimetrico a volume 1.

Da ciò si deduce che è sufficiente studiare il problema isoperimetrico solo per un fissato valore di  $V$ , dato che i minimi per gli altri volumi sono ottenuti semplicemente dilatando un minimo di volume  $V$ . In particolare, per comodità, sia  $\hat{E} \doteq \hat{E}_V$  con  $V = 1$ .

Il nostro obiettivo è dimostrare che  $\hat{E}$  è una palla. Per mostrare ciò, uno strumento fondamentale è la “simmetrizzazione di Steiner”. L’idea è la seguente: dato un insieme  $E$ , gli si applica una trasformazione che lo rende “più simmetrico” e che ha come proprietà fondamentali di preservarne il volume e farne decrescere il perimetro (vedi la figura nella pagina successiva).

Questo fatto permette di dedurre che se si applica tale trasformazione a  $\hat{E}$ , per minimalità il perimetro deve rimanere costante e questa informazione permetterà di dedurre che  $\hat{E}$  è una palla.

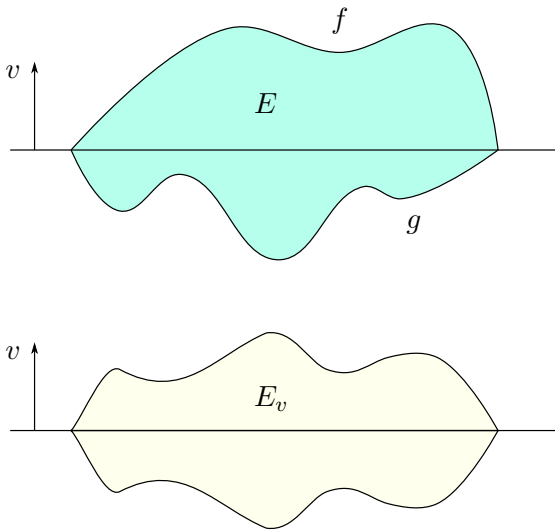
**Definizione:** Dato un insieme  $E \subset \mathbb{R}^n$  e un vettore unitario  $v \in \mathbb{R}^n$ , si definisce il “simmetrizzato di Steiner di  $E$  nella direzione  $v$ ”



come

$$E_v := \bigcup_{\substack{y \in \pi_v, \\ E \cap \ell_y^v \neq \emptyset}} \left\{ y + tv : |t| \leq \frac{1}{2} |E \cap \ell_y^v| \right\}$$

dove  $\pi_v := \{y \in \mathbb{R}^n : y \cdot v = 0\}$  è il piano perpendicolare a  $v$  passante per l'origine,  $\ell_y^v := \{y + tv : t \in \mathbb{R}\}$  è la retta parallela a  $v$  passante per  $y$ , e  $|E \cap \ell_y^v|$  denota la misura 1-dimensionale dell'insieme  $E \cap \ell_y^v$ . (Vedi la Figura 1, dove si utilizzano le notazioni della dimostrazione seguente.)



**Figura 1:** Costruzione dell'insieme simmetrizzato di Steiner.

Questa trasformazione soddisfa le seguenti proprietà:

- (a)  $|E_v| = |E|$  per ogni  $v$ ;
- (b)  $P(E_v) \leq P(E)$  e se vale l'uguaglianza allora  $E \cap \ell_y^v$  è un segmento per ogni  $y \in \pi_v$ ;
- (c) se  $E$  è convesso e  $P(E_v) = P(E)$  allora, per ogni  $v$ ,  $E_v = E + t_v v$  per un certo  $t_v \in \mathbb{R}$ ; ossia  $E_v$  è semplicemente un traslato di  $E$ .

La dimostrazione di queste proprietà non è troppo complicata. Per semplicità le verifichiamo in un caso particolare: denotando con  $\{e_1, \dots, e_n\}$  la base canonica di  $\mathbb{R}^n$ , supponiamo che  $v = e_n$  e che  $E$  sia della forma

$$E = \{(y, t) \in \mathbb{R}^n : -g(y) \leq t \leq f(y)\}$$

con  $f, g : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$  funzioni regolari. Allora in questo caso  $E_v$  è l'insieme dei vettori  $(y, t) \in \mathbb{R}^n$

per cui

$$-\frac{f(y) + g(y)}{2} \leq t \leq \frac{f(y) + g(y)}{2}$$

Notiamo quindi che (a) è immediata. Infatti

$$\begin{aligned} |E| &= \int_{\mathbb{R}^{n-1}} [f(y) + g(y)] dy \\ &= 2 \int_{\mathbb{R}^{n-1}} \frac{f(y) + g(y)}{2} dy \\ &= |E_v| \end{aligned}$$

Per quanto riguarda (b), per il perimetro abbiamo

$$\begin{aligned} P(E) &= \int_{\mathbb{R}^{n-1}} \left( \sqrt{1 + |\nabla f(y)|^2} \right. \\ &\quad \left. + \sqrt{1 + |\nabla g(y)|^2} \right) dy \\ P(E_v) &= 2 \int_{\mathbb{R}^{n-1}} \sqrt{1 + \left| \frac{\nabla f(y) + \nabla g(y)}{2} \right|^2} dy \end{aligned}$$

Allora, dato che la funzione  $\Phi(s) := \sqrt{1 + s^2}$  è convessa e crescente otteniamo

$$\begin{aligned} \Phi\left(\left|\frac{\nabla f(y) + \nabla g(y)}{2}\right|\right) &\leq \Phi\left(\frac{|\nabla f(y)| + |\nabla g(y)|}{2}\right) \\ &\leq \frac{\Phi(|\nabla f(y)|) + \Phi(|\nabla g(y)|)}{2} \end{aligned} \quad (1)$$

e (b) segue integrando la disuguaglianza sopra rispetto a  $y$ .

Infine, se vale  $P(E_v) = P(E)$ , dato che  $\Phi$  è strettamente convessa usando (1) deduciamo che  $|\nabla f(y)| = |\nabla g(y)|$  per ogni  $y$ , da cui  $f = g + \alpha$  per una certa costante  $\alpha \in \mathbb{R}$  e quindi  $E = E_v + (\alpha/2)e_n$ , il che dimostra (c).

Il caso per insiemi generici si ottiene localizzando opportunamente l'argomento appena descritto e ragionando per approssimazione (vedere per esempio [1, Sezione 3] per più dettagli).

Usando (a) e (b) otteniamo la proprietà seguente: se  $\hat{E}$  è un minimo allora  $\hat{E} \cap \ell_y^v$  è un segmento per ogni  $y \in \pi_v$  e  $v \in \mathbb{R}^n$ . Vogliamo mostrare che questo implica che  $\hat{E}$  è convesso. Infatti, dati due punti  $x, x' \in \hat{E}$ , prendiamo  $v$  parallelo a  $x' - x$  e consideriamo la retta  $\ell_y^v$  con  $y \in \pi_v$  tale che  $x, x' \in \ell_y^v$ . Allora, dato che  $\hat{E} \cap \ell_y^v$  è un segmento deduciamo in particolare che  $\tau x + (1 - \tau)x' \in \hat{E}$  per ogni  $\tau \in [0, 1]$ , dunque  $\hat{E}$  è convesso.

A questo punto applichiamo (c) per dedurre che, per ogni  $v \in \mathbb{R}^n$  con  $|v| = 1$ ,  $\hat{E}_v = \hat{E} + t_v v$  per un certo  $t_v \in \mathbb{R}$ . Se potessimo dire che  $t_v = 0$  per ogni  $v$  sarebbe facile concludere (come mostreremo in un attimo). Purtroppo in generale questo non è vero ma possiamo dimostrare che vale a meno di traslare  $\hat{E}$ . Più precisamente, definiamo

$$\hat{E}' := \hat{E} - \sum_{i=1}^n t_{e_i} e_i$$

Allora, usando la proprietà che i vettori  $e_i$  sono ortogonali, non è difficile verificare (ricordando la definizione del simmetrizzato di Steiner) che  $(\hat{E}')_{e_i} = \hat{E}'$  per ogni  $i = 1, \dots, n$ , ossia  $\hat{E}'$  è esattamente uguale al suo simmetrizzato di Steiner nelle direzioni  $e_i$  senza bisogno di applicare alcuna traslazione aggiuntiva. Dato che, per costruzione,  $(\hat{E}')_{e_i}$  è simmetrico rispetto al piano  $\{x_i = 0\}$ , deduciamo che

- $\hat{E}'$  è simmetrico rispetto a tutti i piani coordinati  $\{x_i = 0\}$ ,  $i = 1, \dots, n$

In particolare segue che  $\hat{E}' = -\hat{E}'$ , cioè  $\hat{E}'$  è simmetrico rispetto all'origine.

Ora, grazie a (c) applicato a  $\hat{E}'$  abbiamo  $(\hat{E}')_v = \hat{E}' + t_v v$  da cui, osservando che  $\hat{E}'$  è simmetrico rispetto all'origine e che  $(\hat{E}')_v$  è simmetrico rispetto al piano  $\pi_v$  (per costruzione), deduciamo che  $t_v = 0$ .

Abbiamo dunque dimostrato che

$$(\hat{E}')_v = \hat{E}' \quad \forall v \in \mathbb{R}^n, |v| = 1$$

ossia  $\hat{E}'$  è simmetrico rispetto al piano  $\pi_v$  per ogni  $v$ .

Possiamo ora far vedere che  $\hat{E}'$  è una palla centrata nell'origine. Infatti, prendiamo  $x \in \hat{E}'$  e consideriamo il punto  $x_v \in \hat{E}'$  ottenuto riflettendo  $x$  rispetto al piano  $\pi_v$ . Al variare di  $v$  sulla sfera unitaria i punti  $x_v$  descrivono una sfera di raggio  $|x|$ . Dunque abbiamo dimostrato che se  $x \in \hat{E}'$  allora la sfera di raggio  $|x|$  centrata nell'origine è contenuta in  $\hat{E}'$ . Dato che  $\hat{E}'$  è convesso,  $\hat{E}'$  (e quindi anche  $\hat{E}$ ) è una palla, come desiderato.



[1] F. MAGGI: "Some methods for studying stability in isoperimetric type problems", *Bull. Amer. Math. Soc* **45** (2008) 367–408.



**Alessio Figalli:** Laureato in matematica all'Università di Pisa e alla Scuola Normale Superiore nel 2006, consegue il dottorato di ricerca presso la Scuola Normale Superiore e l'École Normale Supérieure di Lione nel 2007. Nell'ottobre 2007 ottiene il posto di ricercatore presso il CNRS (Centre National de la Recherche Scientifique) a Nizza e nel settembre 2008 diventa professore all'École Polytechnique di Parigi. Nel settembre 2009 si trasferisce come professore associato all'Università del Texas ad Austin, dove nel settembre 2011 viene promosso a professore ordinario.

---

# Integrale di Feynman e limite classico

*Alice rise: «È inutile che ci provi», disse; «non si può credere a una cosa impossibile.» «Oserei dire che non ti sei allenata molto», ribatté la Regina. «Quando ero giovane, mi esercitavo sempre mezz'ora al giorno. A volte riuscivo a credere anche a sei cose impossibili prima di colazione.»*

— Alice nel paese delle meraviglie, Lewis Carroll, 1865

**Matteo Beccaria**

Dipartimento di Matematica & Fisica "Ennio De Giorgi" - Università del Salento

---

**Il formalismo basato sull'integrale di Feynman consente di trattare i problemi della meccanica quantistica mantenendo uno stretto contatto con il suo limite classico. In particolare, le soluzioni delle equazioni del moto che seguono dal principio variazionale continuano a giocare un ruolo centrale favorendo l'intuizione fisica. Alcuni esempi semplici illustrano questa caratteristica importante dell'approccio di Feynman alla quantizzazione.**

Il *principio di corrispondenza* è stato introdotto da Niels Bohr nel 1920 ed afferma che la meccanica quantistica deve riprodurre i risultati della meccanica classica per sistemi caratterizzati da un'azione grande rispetto alla costante di Planck (ridotta)  $\hbar$ .

Formulato in questi termini è un principio di buon senso, ma solo qualitativo. È quindi importante capire come tradurlo in termini quantitativi. Il problema non è banale poiché la meccanica

quantistica è basata su un formalismo completamente diverso da quello della meccanica classica. Infatti, la meccanica classica è basata su traiettorie  $(q(t), p(t))$  nello spazio delle fasi che soddisfano le equazioni del moto **non-lineari** che seguono dal principio variazionale

$$\delta S = 0 \quad (1)$$

dove  $S$  è l'azione, un funzionale di  $q(t)$  e  $p(t)$ . In maniera del tutto differente, in meccanica quantistica gli stati fisici sono descritti da uno spazio di Hilbert  $\mathcal{H}$  (modulo trasformazioni di fase) e le osservabili fisiche, ovvero le quantità misurabili, sono associate ad operatori **lineari** su  $\mathcal{H}$ . In quanto operatori, le osservabili in generale non commutano. Questo fatto tecnico costituisce la base matematica per il principio di indeterminazione di Heisenberg. L'esempio più semplice è la celebre relazione tra l'operatore posizione  $\hat{q}$  e l'operatore impulso  $\hat{p}$  di una particella puntiforme in una dimensione spaziale:

$$[\hat{q}, \hat{p}] = i\hbar \quad (2)$$

Questa relazione conduce direttamente alla celebre relazione di indeterminazione di Heisenberg  $\Delta q \Delta p \sim \hbar$ . Nel limite formale  $\hbar \rightarrow 0$  gli effetti non commutativi scompaiono. In questo senso il principio di corrispondenza è plausibile, ma rimane ancora da capire quale sia il legame quantitativo con i risultati della meccanica classica<sup>1</sup>. In particolare, quello che ci interessa chiarire è *quale sia il ruolo dell'azione  $S$  in un sistema quantistico*.

Storicamente, Dirac<sup>2</sup> mostrò nel 1933 come la meccanica classica possa emergere dalla meccanica quantistica sotto certe condizioni [1]. Traiettorie con azione grande  $S \gg \hbar$  e che non soddisfano le equazioni del moto sono soppresse dall'interferenza quantistica e sopravvivono solo le traiettorie estremali per le quali  $\delta S = 0$ . La trattazione di Dirac introduce già una somma sulle traiettorie e l'idea viene sviluppata successivamente da R. P. Feynman<sup>3</sup> nella sua tesi di dottorato del 1942. Il metodo di Dirac è oggi universalmente noto come tecnica dell'integrale sui cammini (*path-integral*) di Feynman.

Prima di discutere l'approccio di Dirac-Feynman al problema del limite classico in meccanica quantistica è utile concentrarci su due problemi che illustrano chiaramente le domande *difficili* alle quali vogliamo dare una risposta con questa tecnica. Entrambi gli esempi possono essere formulati in termini elementari per il sistema costituito da un punto materiale, ma le situazioni che discuteremo sono il prototipo per analoghi problemi in contesti più realistici<sup>4</sup>.

## Due problemi difficili

<sup>1</sup>Notiamo che il problema è molto più semplice in meccanica relativistica dove il parametro di deformazione rispetto alla meccanica galileiana è il rapporto tra la velocità caratteristica e la velocità della luce  $v/c$ . In questo caso, la *struttura* delle equazioni non cambia in maniera drastica in presenza di correzioni  $\sim v/c$ .

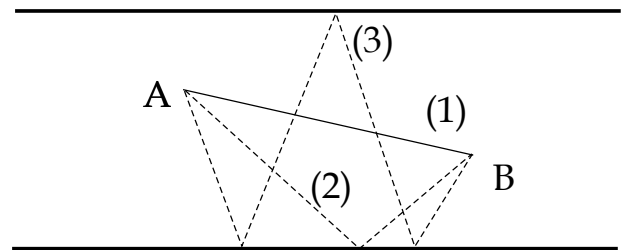
<sup>2</sup>Premio Nobel per la fisica nel 1933 insieme a Erwin Schrödinger.

<sup>3</sup>Premio Nobel per la Fisica nel 1965 insieme a Sin-Itiro Tomonaga e Julian Schwinger.

<sup>4</sup>Altri problemi molto interessanti legati al limite classico sono, ad esempio, il problema di sistemi a molti corpi o l'emergere del comportamento classico nel passaggio da sistemi microscopici a sistemi macroscopici. Altrettanto interessanti sono i legami tra lo spettro dell'energia e le proprietà ergodiche del moto classico.

## Topologia e molteplicità di soluzioni classiche

Il primo problema che consideriamo è quello di una particella quantistica che si muova all'interno di una striscia bidimensionale con pareti riflettenti, come mostrato nella Figura 1. Classicamente, ci sono infinite traiettorie che soddisfano  $\delta S = 0$  e che portano dal punto A al punto B nel tempo  $T$ . La più breve è rettilinea (1). Oltre a questa ce ne sono infinite altre che si riflettono sulle barriere come le traiettorie (2) e (3) della figura. Dal punto di vista quantistico ci interessa

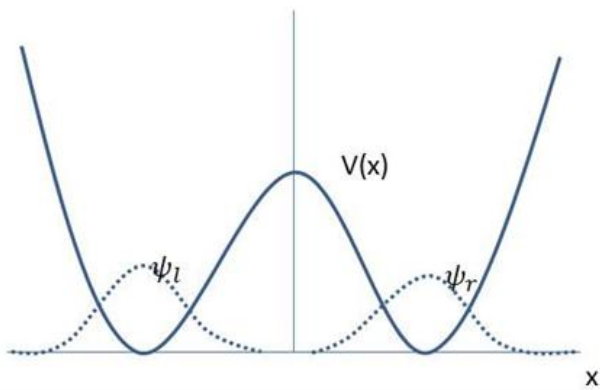


**Figura 1:** Propagazione classica da A a B nel tempo  $T$  in presenza di barriere.

determinare la probabilità che la particella presente in A al tempo  $t = 0$  si trovi in B al tempo  $t = T$ . Inoltre, la domanda che vogliamo porci è *in che modo la molteplicità di soluzioni classiche si riflette sulla propagazione quantistica*. In generale, un problema analogo si pone tutte le volte che la propagazione classica avviene all'interno di geometrie topologicamente non banali. In seguito, mostreremo come l'approccio di Feynman possa illustrare chiaramente il legame tra le traiettorie classiche e la propagazione quantistica almeno nel limite  $S/\hbar \gg 1$ .

## Effetti non perturbativi

Il secondo problema che consideriamo riguarda una particella che si muova nel potenziale simmetrico unidimensionale  $V(x) = V(-x)$  con due buche come in Figura 2. Gli stati classici di minima energia corrispondono ad una particella ferma in uno dei due minimi. In meccanica quantistica ci si potrebbe aspettare ingenuamente che a tali stati corrispondano funzioni d'onda localizzate intorno ai minimi, come nella figura. Come è noto questo non è possibile dato che lo stato quantico di minima energia deve obbedire alla simmetria del potenziale. Di fatto, esso risulta essere una combinazione lineare simmetrica



**Figura 2:** Particella unidimensionale che si muove nel potenziale simmetrico  $V(x) = V(-x)$ . La figura mostra due funzioni d'onda localizzate intorno ai due minimi classici. Come discusso nel testo, lo stato quantico di minima energia è simmetrico a sua volta ed è costruito come sovrapposizione quantistica simmetrica delle due funzioni d'onda.

delle due funzioni d'onda localizzate mentre il primo stato eccitato è invece associato alla combinazione antisimmetrica. La differenza di energia tra i due stati  $\Delta E = E_1 - E_0$  dipende tipicamente dai parametri del potenziale in modo non-perturbativo cioè si annulla a tutti gli ordini nello sviluppo perturbativo in termini di questi parametri. La ragione è che esiste un'azione nascosta  $S$ , caratteristica delle doppia buca, tale che  $\Delta E$  è espresso in termini di  $\exp(-S/\hbar)$ . Chiaramente lo sviluppo ingenuo in potenze di  $\hbar/S$  secondo le tecniche elementari della meccanica quantistica non è applicabile a questo problema. Di nuovo, mostreremo come l'approccio di Feynman consenta di trattare questo problema in modo molto semplice<sup>5</sup>.

## La formulazione di Feynman

Per capire in che cosa consista la formulazione di Feynman della meccanica quantistica [2] è utile considerare la quantità seguente

$$K(q'', q'; T) = \langle q'' | e^{-\frac{i}{\hbar} T \hat{H}} | q' \rangle \quad (3)$$

<sup>5</sup>Dal punto di vista matematico, il problema nasce dal fatto che il limite  $\hbar \rightarrow 0$  è una *perturbazione singolare*. I contributi legati ad  $\hbar$  che compaiono nell'equazione d'onda di Schrödinger sono associati ai termini con derivate di ordine massimo. Per chi ha familiarità con la fluidodinamica, la situazione è analoga al limite di viscosità zero delle equazioni di Navier-Stokes.

dove gli stati  $|q\rangle$  sono autostati (generalizzati) dell'operatore posizione e  $\hat{H}$  è l'hamiltoniana indipendente dal tempo del sistema. Feynman chiama questa espressione *propagatore* ed il suo significato fisico è quello di ampiezza di probabilità perché una particella nel punto  $q'$  al tempo  $t$  sia osservata nel punto  $q''$  al tempo  $t + T$ . Il problema è determinare in che modo il propagatore *veda* le traiettorie classiche. La domanda non è banale se si pensa alla risposta tipica che lo studente dei corsi elementari di meccanica quantistica fornisce per calcolare  $K$ . La procedura standard è questa: supponiamo di conoscere lo spettro di  $H$  che per semplicità assumeremo discreto con autovettori ortonormali  $|n\rangle$  associati agli autovalori  $E_n$ . Dunque possiamo scrivere

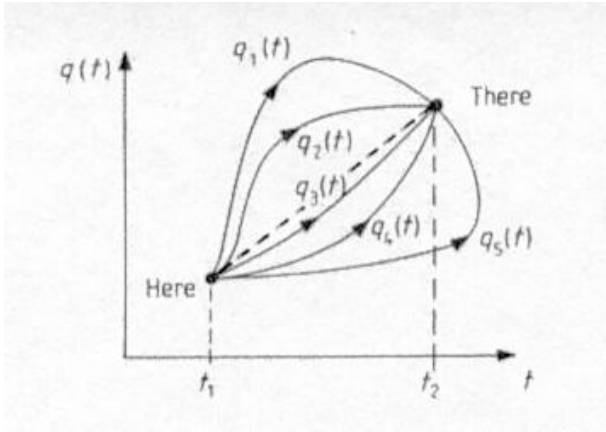
$$K(q'', q'; T) = \sum_n \langle q'' | n \rangle \langle n | q' \rangle e^{-\frac{i}{\hbar} T E_n} \quad (4)$$

Supponendo di poter calcolare la somma infinita otteniamo il risultato cercato che tuttavia è ben lontano dall'essere fisicamente trasparente. In particolare, *non è chiaro dove siano le traiettorie classiche*. Aggiungiamo inoltre che in generale lo spettro di  $H$  non è noto. L'idea di Feynman è quella di riscrivere l'espressione di  $K$  in modo differente, ma rigorosamente equivalente e più efficace da vari punti di vista, incluso il recupero del limite classico. Tralasciando la derivazione, il risultato di Feynman può essere scritto come integrale funzionale nella forma suggestiva:

$$K(q'', q'; T) = \int_{\substack{q(t')=q' \\ q(t'+T)=q''}} \mathcal{D}q(t) e^{\frac{i}{\hbar} S(q(t))} \quad (5)$$

In questa espressione il tempo iniziale  $t'$  è arbitrario poiché  $H$  non dipende dal tempo. La formula di Feynman è elegante, ma deve essere presa *cum grano salis*. L'espressione in forma di integrale sulle traiettorie è in realtà costruita attraverso una procedura di limite in modo analogo alla costruzione elementare dell'integrale di Riemann. Il simbolo  $\mathcal{D}q$  non rappresenta una misura nonostante la notazione, sebbene alcune proprietà siano quelle intuitive suggerite dalla notazione<sup>6</sup>. In termini descrittivi, la for-

<sup>6</sup>Una trattazione rigorosa dal punto di vista della costruzione di una misura funzionale è problematica poiché le traiettorie dominanti sono tipicamente ovunque non differenziabili e a causa della fase complessa nell'in-



**Figura 3:** Rappresentazione schematica della formula di Feynman come somma sui cammini classici pesati con una fase che dipende dall'azione. I cammini sono tutti quelli possibili senza necessariamente obbedire alle equazioni del moto classiche.

mula di Feynman esprime il propagatore come somma sui cammini classici pesati con una fase che dipende dall'azione, vedi Figura 3. Il ruolo di questa fase è quello di ricostruire i fenomeni di *interferenza quantistica*. Sottolineiamo che per cammini classici intendiamo funzioni continue del tempo che non hanno a che fare con gli operatori che agiscono su  $\mathcal{H}$ , ma non stiamo ancora parlando delle soluzioni delle equazioni del moto classiche. Tuttavia, la presenza esplicita dell'azione  $S(q(t))$  nella formula di Feynman suggerisce come procedere.

## Il limite classico $S/\hbar \gg 1$

Supponiamo di domandarci quale sia il comportamento del propagatore per un sistema in cui l'azione caratteristica sia grande rispetto alla costante di Planck. Formalmente, questo significa valutare la formula di Feynman nel limite  $\hbar \rightarrow 0$ . Considerazioni elementari portano alla congettura che l'integrale sia dominato dai punti in cui la fase nel fattore  $e^{\frac{i}{\hbar}S}$  è stazionaria. Ma questa è esattamente la condizione (1) che definisce le traiettorie classiche! Le correzioni successive al

tegrazione. Con opportune continuazioni analitiche, l'integrale di Feynman può essere trattato in modo matematicamente rigoroso e l'espressione del propagatore è legata alla cosiddetta formula di Kac [3]. In questa sede, questi aspetti non sono rilevanti dato che non cambiano la nostra discussione centrata sul legame con il limite classico.

metodo della fase stazionaria portano al calcolo delle correzioni quantistiche come fluttuazioni intorno alle soluzioni classiche. Chiaramente, la formulazione in termini di integrale di Feynman suggerisce sviluppi legati alla *teoria asintotica*. Ad esempio, l'esistenza di punti stazionari non fisici è alla base del calcolo istantonico come vedremo in seguito.

Mostriamo ora come il metodo di Feynman possa trattare agevolmente i due problemi *difficili* che abbiamo illustrato in precedenza.

## La particella libera sulla circonferenza

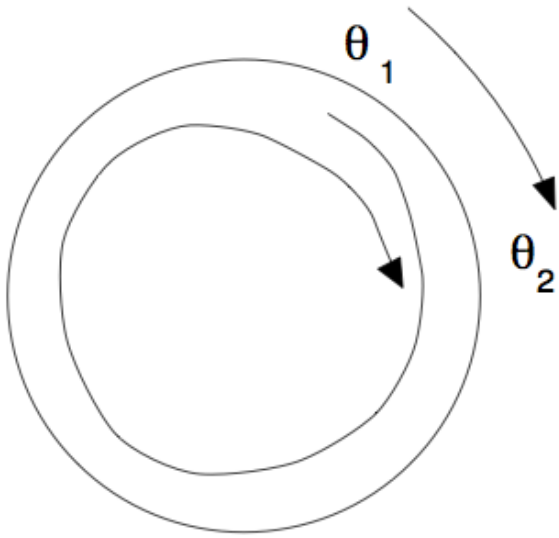
Riprendiamo in considerazione il primo problema discusso in precedenza. Vogliamo mostrare come il metodo di Feynman, nel caso di sistemi con molteplicità di soluzioni classiche, consenta di ritrovare il loro contributo dettagliato almeno nel regime quasi-classico in cui l'azione caratteristica  $S$  del problema sia grande rispetto ad  $\hbar$ . A questo scopo, analizziamo un problema molto semplice, ma che illustra il meccanismo generale. Si tratta di una particella di massa unitaria che si muove liberamente su una circonferenza di raggio  $R$  con traiettoria  $\vartheta(t)$  ed azione

$$S = \int dt \frac{1}{2} R^2 \dot{\vartheta}^2 \quad (6)$$

Poiché la circonferenza non è semplicemente connessa, è possibile andare da  $\vartheta_1$  a  $\vartheta_2$  nel tempo  $T$  in infiniti modi a seconda del numero di giri che si compiono intorno alla circonferenza durante il tragitto, come illustrato nella figura Figura 4 che mostra una coppia di traiettorie. Questa molteplicità delle soluzioni classiche deve avere una controparte quantistica come la formula di Feynman mostrerà agevolmente. Il calcolo secondo la formula (4) è molto semplice per questo problema. L'hamiltoniana della particella è

$$H = \frac{p_{\vartheta}^2}{2R^2} = -\frac{\hbar^2}{2R^2} \frac{d^2}{d\vartheta^2} \quad (7)$$

Consideriamo autofunzioni periodiche (in seguito commenteremo questa scelta). Esse sono onde



**Figura 4:** Particella libera su una circonferenza. Classicamente è possibile andare da  $\vartheta_1$  a  $\vartheta_2$  nel tempo  $T$  in infiniti modi a seconda del numero di giri che si compiono intorno alla circonferenza durante il tragitto.

piane

$$\psi_n(\vartheta) = \frac{1}{\sqrt{2\pi}} e^{in\vartheta}, \quad E_n = \frac{\hbar^2 n^2}{2R^2} \quad (8)$$

La formula (4) diventa

$$K(\vartheta_2, \vartheta_1; T) = \frac{1}{2\pi} \sum_{n \in \mathbb{Z}} e^{in(\vartheta_2 - \vartheta_1)} e^{-\frac{i\hbar n^2 T}{2R^2}} \quad (9)$$

In termini della funzione  $\Theta_3$  di Jacobi [4]

$$\Theta_3(z|\tau) = \sum_{n \in \mathbb{Z}} e^{i(2nz + \pi n^2 \tau)} \quad (10)$$

il propagatore si può scrivere

$$K(\vartheta_2, \vartheta_1; T) = \frac{1}{2\pi} \Theta_3\left(\frac{\vartheta_2 - \vartheta_1}{2} \middle| -\frac{\hbar T}{2\pi R^2}\right) \quad (11)$$

Usando a questo punto la proprietà di trasformazione modulare della funzione di Jacobi, ovvero l'identità di Poisson

$$\Theta_3(z|\tau) = (-i\tau)^{-1/2} e^{\frac{z^2}{i\pi\tau}} \Theta_3(z/\tau | -1/\tau) \quad (12)$$

la somma infinita può essere scritta

$$K(\vartheta_2, \vartheta_1; T) = \sum_{n \in \mathbb{Z}} K_n(\vartheta_2, \vartheta_1; T) \quad (13)$$

## A. Il propagatore per la particella libera

Il propagatore per la particella libera in una dimensione è

$$K_0(q'', q'; T) = \langle q'' | e^{-\frac{i}{\hbar} T \frac{\hat{p}^2}{2m}} | q' \rangle$$

dove l'operatore impulso è  $\hat{p} = -i\hbar \frac{d}{dq}$ . Data la forma semplice dell'hamiltoniana il propagatore non è altro che il nucleo dell'equazione del calore dopo un'opportuna continuazione analitica nella variabile temporale. Come è noto il risultato si scrive

$$K_0(q'', q'; T) = \left(\frac{m}{2\pi i \hbar T}\right)^{1/2} e^{\frac{im(q'' - q')^2}{2\hbar T}}$$

dove

$$K_n(\vartheta_2, \vartheta_1; T) = \sqrt{\frac{R^2}{2\pi i \hbar T}} e^{\frac{iR^2}{2\hbar} \frac{(\vartheta_2 - \vartheta_1 + 2\pi n)^2}{T}} \quad (14)$$

Notiamo che questa non è altro che la somma dei propagatori della particella libera su  $\mathbb{R}$  che percorre nel tempo  $T$  la distanza  $\vartheta_2 - \vartheta_1 + 2\pi n$  (vedi Inserto A).

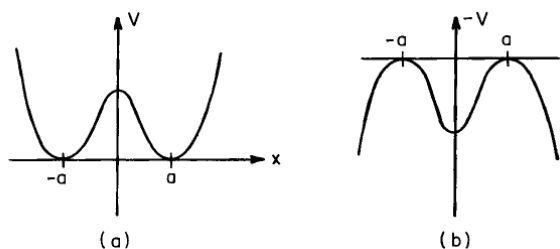
Dal punto di vista della formula di Feynman, questo risultato è ovvio. Infatti il calcolo del propagatore con il metodo della fase stazionaria è esatto per azioni quadratiche [5] e si riduce alla somma dei contributi indicati, praticamente senza nessun calcolo. Più in generale, per un'azione più complicata, come pure in dimensione maggiore di uno, la struttura del propagatore è simile e si ottiene analogamente sommando sulle soluzioni classiche topologicamente non equivalenti. Il metodo standard è inapplicabile in un caso generale (non libero) perché lo spettro non è noto e comunque manipolazioni come la trasformazione modulare della funzione di Jacobi sono state *ad hoc*.

Concludiamo commentando la scelta della periodicità  $\psi_n(0) = \psi_n(2\pi)$ . Più in generale, è consistente con la Meccanica Quantistica, assumere la condizione al contorno  $\psi_n(0) = e^{i\varphi} \psi_n(2\pi)$  per valori reali generici di  $\varphi$ . La ragione è che la periodicità fisica deve essere imposta su  $|\psi_n|^2$ . Il calcolo modificato porta semplicemente a modi-

ficare la somma (13) introducendo un peso  $e^{i\varphi n}$  nel contributo  $K_n$ . Questa modifica è del tutto accettabile nel quadro della teoria della formula di Feynman per sistemi con spazio delle configurazioni topologicamente non banale.

## Doppia buca e istantoni

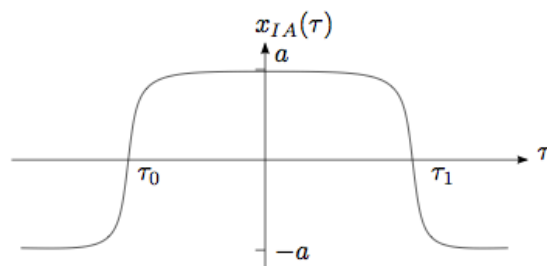
Passiamo adesso a discutere il problema del comportamento di un sistema con potenziale a doppia buca con il formalismo di Feynman. Consideriamo un potenziale simmetrico  $V(x)$  con minimi in  $x = \pm a$  come nella parte (a) della figura Figura 5. Effettuando una continuazione analiti-



**Figura 5:** (a) Potenziale a doppia buca (b) il suo opposto, rilevante dopo continuazione analitica della variabile tempo.

ca nella variabile tempo e passando al cosiddetto *tempo euclideo*, le equazioni del moto diventano quelle per il potenziale  $-V$  disegnato nella parte (b) di Figura 5. Una soluzione interessante è quella che interpola tra i due massimi (vedi inserto B). Questa soluzione è detta *istantone*<sup>7</sup>[6]. Naturalmente non si tratta di una soluzione fisica poiché è stata fatta una continuazione analitica nel tempo. Dunque, *la soluzione istantonica è analoga ai punti sella nel piano complesso che intervengono nella valutazione asintotica di integrali dipendenti da parametro*. Oltre a questa soluzione ne esistono infinite altre che effettuano un numero arbitrario di salti tra  $a$  e  $-a$  connettendo approssimativamente (ma correttamente nel limite semiclassico) un certo numero di istantoni elementari. Per esempio, nella Figura 6 è rappresentata una soluzione costituita da un istantone seguito da un anti-istantone. La soluzione parte dal minimo del potenziale in  $-a$  per  $\tau = -\infty$ , effettua una transizione al minimo in  $a$  al tempo  $\tau_0$ , torna indietro al tempo  $\tau_1$  finendo nel minimo

<sup>7</sup>Il nome inventato dal Nobel G. 't Hooft indica che si tratta di un solitone nella variabile tempo.



**Figura 6:** Soluzione costruita combinando un istantone seguito da un anti-istantone.

in  $-a$  per  $\tau \rightarrow +\infty$ . In prima approssimazione, questa soluzione si ottiene giustapponendo due soluzioni elementari che collegano  $-a$  con  $a$  e viceversa.

Il calcolo del propagatore tra i punti  $x = \pm a$  nel tempo (euclideo)  $T$  può essere fatta calcolando le fluttuazioni previste dalla formula di Feynman intorno alle soluzioni istantoniche. Il risultato del calcolo è dato dalla formula

$$\langle \pm a | e^{-\frac{\hat{H}T}{\hbar}} | -a \rangle = \left( \frac{\omega}{\pi \hbar} \right)^{1/2} e^{-\frac{1}{2}\omega T} \times \frac{\exp\left(K e^{-\frac{S}{\hbar}} T\right) \mp \exp\left(-K e^{-\frac{S}{\hbar}} T\right)}{2} \quad (15)$$

dove  $\omega$  è la pulsazione classica intorno ai minimi di  $V$ ,  $K$  è una combinazione di costanti (massa, parametri del potenziale, etc.),  $S$  è l'azione della soluzione istantonica elementare. Confrontando questa espressione con la forma (4) del propagatore concludiamo l'esistenza di uno stato fondamentale  $|+\rangle$  e di un primo stato eccitato  $|-\rangle$  con energie

$$E_{\pm} = \frac{\hbar\omega}{2} \mp \hbar K e^{-\frac{S}{\hbar}} \quad (16)$$

Dunque, la valutazione dell'integrale di Feynman intorno al punto sella istantonico (che è di fatto una soluzione del problema variazionale classico, per quanto non fisica) prevede due stati con energia pari all'energia di una particella che vede solo uno dei due minimi più un contributo esponenzialmente soppresso nel parametro che regola il regime di quasi-classicità  $S/\hbar \gg 1$ .

Come abbiamo già accennato, la valutazione di questo *splitting* energetico è notoriamente molto difficile con metodi perturbativi tradizionali, vedi Inserto B. Il calcolo può essere effettuato utilizzando il metodo WKB (Wentzel, Kramers,



## B. Un esempio esplicito di doppia buca

Consideriamo un esempio particolare di potenziale a doppia buca e cioè  $V(x) = \lambda(x^2 - a^2)^2$  per  $a, \lambda > 0$ . Introducendo il tempo euclideo  $\tau = -it$ , il fattore  $\exp(iS/\hbar)$  diventa  $\exp(-S_E/\hbar)$  con azione euclidea (le derivate sono ora rispetto a  $\tau$ )

$$S_E = \int_{\tau_1}^{\tau_2} d\tau \left( \frac{m}{2} x'^2 + V(x) \right)$$

Il principio variazionale porta alle equazioni del moto  $-m x'' + V'(x) = 0$ . Si noti che il potenziale euclideo ha segno opposto rispetto a quello in tempo reale. Le equazioni del moto sono equivalenti alla conservazione dell'energia che, per la soluzione istantonica  $x(\tau) = a \tanh\left(a \sqrt{\frac{2\lambda}{m}} \tau\right)$ , predice  $-\frac{m}{2} x'^2 + V(x) = 0$ . Quindi, cambiando variabile, l'azione istantonica risulta

$$S = \int_{-\infty}^{\infty} d\tau 2V = \int_{-a}^a dx \sqrt{2mV} = \frac{4\sqrt{2}}{3} a^3 \sqrt{m\lambda}$$

Se ora introduciamo la pulsazione intorno ai minimi  $\omega^2 = \frac{V''(a)}{m} = \frac{8a^2\lambda}{m}$ , possiamo scrivere l'azione istantonica come la seguente funzione di  $\lambda$  a  $\omega$  fissato

$$S_E = \frac{m^2 \omega^3}{12 \lambda}$$

Da questa formula risulta chiaro che l'approccio perturbativo in  $\lambda$  come piccola perturbazione rispetto al potenziale armonico con pulsazione  $\omega$  porta a correzioni  $\exp(-S_E/\hbar)$  che sono zero a tutti gli ordini in  $\lambda$ . Fisicamente, il limite  $\lambda \rightarrow 0$  con  $\omega$  fissato corrisponde ad un potenziale descritto da due buche approssimativamente armoniche situate a grande distanza e separate da una barriera di altezza  $\sim 1/\lambda$ . Il fattore esponenzialmente soppresso  $\exp(-S_E/\hbar)$  è legato alla probabilità di penetrazione di questa barriera.

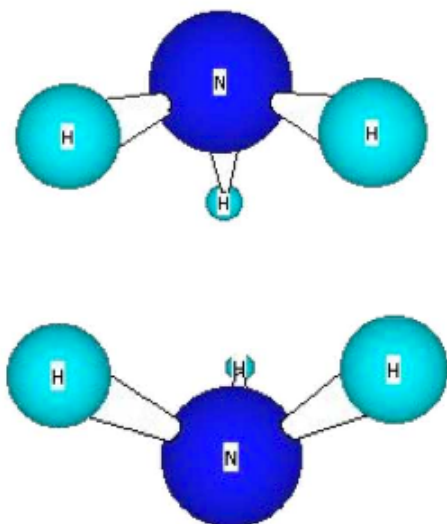
Brillouin) che tuttavia incontra notevoli problemi nell'essere esteso a più di una dimensione. La trattazione di Feynman è invece molto più potente e non ha queste limitazioni.

Concludiamo ricordando che un sistema unidimensionale con potenziale a doppia buca è un buon modello per la molecola di ammoniacca  $NH_3$ , Figura 7. Il fattore di correzione  $e^{-\frac{S}{\hbar}}$  è legato alla probabilità di transizione tra i due stati della molecola (vedi ancora Inserto B). La differenza di energia tra i due livelli più bassi è di  $9.84 \cdot 10^{-5}$  eV che corrisponde ad una frequenza  $\Delta E/h$  di circa 24 GHz che caratterizza il MASER all'ammoniacca.

## Conclusioni

La formulazione di Feynman della meccanica quantistica è rigorosamente equivalente all'ap-

proccio tradizionale, ma pone al centro dell'attenzione l'azione classica  $S$ . Di conseguenza, molte proprietà di natura classica legate alla soluzione del problema variazionale  $\delta S = 0$  hanno un ruolo importante. Abbiamo visto due esempi di questo aspetto generale. Nel primo, la topologia non-banale dello spazio delle configurazioni implica l'esistenza di infinite soluzioni globali del problema variazionale. Queste sono immediatamente presenti nella formulazione di Feynman e forniscono un'interpretazione a quantità quantistiche che sarebbero altrimenti decisamente oscure. Nel secondo esempio, il problema variazionale classico ammette soluzioni non fisiche ottenute per continuazione analitica. Di nuovo, la formula di Feynman ne tiene conto in maniera naturale e consente di calcolare in modo estremamente semplice quantità che sarebbero altrimenti elusive.



**Figura 7:** Rappresentazione schematica della molecola di ammoniaca nei due stati modellizzati dalle due buche del potenziale discusso nel testo.



- [1] P. A. M. DIRAC: "The Lagrangian in quantum mechanics", *Phys. Z. der Sowjetunion* **3** (1933) 64-71.

- [2] R. P. FEYNMAN: "The Space-Time Formulation of Non-relativistic Quantum Mechanics", *Reviews of Modern Physics* **20** (2) (1948) 367-387.
- [3] B. SIMON: *Functional Integration and Quantum Physics*, Academic Press (1979).
- [4] M. ABRAMOWITZ, I. A. STEGUN: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications (1972).
- [5] L. S. SCHULMAN: *Techniques and Applications of Path Integration*, Wiley (1981).
- [6] T. SCHAEFER, E. SHURYAK: "Instantons in QCD", *Rev. Mod. Phys.* **70** (1998) 323-426.



**Matteo Beccaria:** si è laureato in fisica nel 1991 presso l'Università di Pisa e ha conseguito il Diploma di Licenza in fisica presso la Scuola Normale Superiore. Nel 1995 ha conseguito il titolo di Dottore di Ricerca presso l'Università di Pisa. Dal 2002 è Professore Associato di Fisica Teorica all'Università del Salento dove insegna Teoria dei Campi Quantizzati. Si occupa di teoria delle interazioni fondamentali ed in particolare di teoria di stringa.

---

# Il principio variazionale nella fisica dei sistemi quantistici a molticorpi

**Giampaolo Co'**

Dipartimento di Matematica & Fisica "Ennio De Giorgi" - Università del Salento

---

**Il problema quantistico a molticorpi è affrontato ricercando soluzioni approssimate dell'equazione di Schrödinger. In questo ambito, il principio variazionale è uno dei metodi più usati. La sua applicazione più conosciuta, e semplice, è quella ad un sistema di particelle non interagenti. Questo conduce alla formulazione delle equazioni di Hartree-Fock. Il principio variazionale è utilizzato anche per descrivere sistemi di particelle in interazione tra loro. Le teorie che sono state sviluppate in questo ambito sono più elaborate e difficili da gestire, ma hanno assunto una grande rilevanza per la descrizione di liquidi quantistici.**

## Introduzione

Per descrivere un sistema composto è innanzitutto necessario scegliere le componenti fondamentali sulle quali basare la descrizione. In questa modellizzazione si trascurano le dimensioni e soprattutto la struttura interna di tali componenti. Ad esempio, la descrizione di un nucleo atomico, si basa normalmente sui nucleoni, nome

generico per indicare protoni e neutroni, considerandoli puntiformi, nonostante siano sistemi estesi composti da quark e gluoni. Analogamente, nella descrizione di un liquido quantistico si trascura la struttura interna delle molecole che lo compongono.

Una volta scelte le componenti fondamentali, che d'ora in poi chiamerò particelle, è richiesta la conoscenza della loro interazione reciproca. Normalmente questa interazione viene ottenuta dallo studio dei sistemi di due particelle, sistemi nucleone-nucleone per il caso nucleare e bi-molecolari per quello del liquido quantistico.

A questo punto il problema a molticorpi è ben definito. Nel caso in cui le energie cinetiche delle singole particelle siano molto inferiori al valore delle loro masse a riposo, il problema può essere trattato in termini di meccanica quantistica non relativistica, quindi si riduce alla ricerca di soluzioni dell'equazione di Schrödinger a molti corpi,

$$H(1, \dots, N)\Psi(1, \dots, N) = E\Psi(1, \dots, N) \quad (1)$$

Nell'equazione precedente, ho indicato con  $N$  il numero di particelle che compongono il sistema,

## Il principio variazionale: parte prima

Il principio variazionale afferma che la ricerca del minimo del funzionale dell'energia (3) è equivalente alla soluzione dell'equazione di Schrödinger (1). Possiamo riscrivere l'Eq. (3) moltiplicando entrambi i membri per il denominatore e quindi applicare la variazione

$$(\delta E[\Psi]) \langle \Psi | \Psi \rangle + E[\Psi] (\delta \langle \Psi | \Psi \rangle) = \delta \langle \Psi | H | \Psi \rangle$$

Imporre  $\delta E(\Psi) = 0$  equivale quindi a richiedere che

$$\delta \langle \Psi | (H - E) | \Psi \rangle = \langle \delta \Psi | (H - E) | \Psi \rangle + \langle \Psi | (H - E) | \delta \Psi \rangle = 0$$

I due termini dell'equazione precedente sono, di fatto, indipendenti uno dall'altro perché si possono far variare indipendentemente le parti reali ed immaginarie di  $\Psi$ . Questo implica che dobbiamo imporre separatamente

$$\langle \delta \Psi | (H - E) | \Psi \rangle = 0 \quad \text{e} \quad \langle \Psi | (H - E) | \delta \Psi \rangle = 0$$

Dato che è sempre possibile trovare una variazione di  $|\Psi\rangle$  diversa da zero, ognuna di queste equazioni corrisponde separatamente all'equazione di Schrödinger (1).

e con  $H$  l'operatore hamiltoniano

$$H = \sum_{i=1}^N -\frac{\hbar^2}{2m_i} \nabla_i^2 + \frac{1}{2} \sum_{i,j=1}^N v(i,j) \quad (2)$$

dove il primo termine rappresenta l'energia cinetica del sistema. Nell'equazione (1)  $\Psi$  è la funzione d'onda che descrive lo stato delle  $N$  particelle, e l'autovalore  $E$  rappresenta l'energia totale, non relativistica, del sistema. Gli argomenti di  $H$  e  $\Psi$  rappresentano le variabili che identificano ogni particella: posizione, spin ed, eventualmente, altro.

Nucleo	$Z$	$N - Z$	$\mathcal{N}_{\text{conf}}$
${}^3\text{H}$	1	2	24
${}^3\text{He}$	2	1	24
${}^4\text{He}$	2	2	96
${}^6\text{He}$	2	4	960
${}^6\text{Li}$	3	3	1280
${}^8\text{He}$	2	6	7168
${}^{12}\text{C}$	6	6	3784704
${}^{16}\text{O}$	8	8	$8.4 \cdot 10^8$
${}^{40}\text{Ca}$	20	20	$1.5 \cdot 10^{23}$
${}^{48}\text{Ca}$	20	28	$4.7 \cdot 10^{27}$

**Tabella 1:** Numero di configurazioni di spin e isospin per alcuni nuclei.

Anche se ben definito, il problema è di difficile soluzione. Un modo semplice per comprendere le difficoltà nel risolvere l'equazione (1) consiste nel moltiplicare l'equazione, a sinistra, per  $\Psi^*$  e poi integrare e sommare su tutte le variabili che identificano le particelle. In questo modo otteniamo un'espressione per l'energia

$$E = \frac{\int d1 \cdots dN \Psi^* H \Psi}{\int d1 \cdots dN \Psi^* \Psi} = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (3)$$

dove nell'ultimo termine ho utilizzato la notazione di Dirac. Il simbolo di integrazione nell'equazione precedente non indica solo l'integrazione sulle coordinate spaziali, ma anche la somma sulle coordinate discrete che definiscono le particelle. Ad esempio per i nucleoni sono spin, ed isospin. Per ottenere il valore dell'energia bisogna effettuare il calcolo di tanti integrali spaziali a  $3N$  dimensioni quante sono le diverse possibili configurazioni permesse dalle variabili discrete. Nel caso di un nucleo composto da  $Z$  protoni e  $N - Z$  neutroni, questo numero è

$$\mathcal{N}_{\text{conf}} = 2^N \frac{N!}{Z!(N-Z)!} \quad (4)$$

Per esempio, per il nucleo di  ${}^3\text{He}$  questo significa considerare che il primo nucleone è un protone con spin allineato all'insù, il secondo un neutro-

## Il principio variazionale: parte seconda

Esprimiamo uno stato di prova  $|\Phi\rangle$  come combinazione lineare degli autostati  $|\Psi_n\rangle$  dell'hamiltoniana  $H$

$$|\Phi\rangle = \sum_n A_n |\Psi_n\rangle$$

dove abbiamo indicato con  $A_n$  dei numeri complessi e

$$H|\Psi_n\rangle = E_n|\Psi_n\rangle$$

Il funzionale dell'energia dipendente dalle funzioni di prova  $|\Phi\rangle$  può essere scritto come

$$E[\Phi] = \frac{\langle \Phi | H | \Phi \rangle}{\langle \Phi | \Phi \rangle} = \frac{\sum_{nn'} \langle \Psi_{n'} | A_{n'}^* H A_n | \Psi_n \rangle}{\sum_{nn'} \langle \Psi_{n'} | A_{n'}^* A_n | \Psi_n \rangle} = \frac{\sum_{nn'} A_{n'}^* A_n \langle \Psi_{n'} | H | \Psi_n \rangle}{\sum_n |A_n|^2}$$

dove, nell'ultimo passaggio, ho utilizzato l'ortonormalità delle funzioni d'onda  $|\Psi_n\rangle$ . Considerando l'equazione di Schrödinger posso riscrivere l'ultima equazione in funzione degli autovalori dell'energia  $E_n$ .

$$E[\Phi] = \frac{\sum_n |A_n|^2 E_n}{\sum_n |A_n|^2} \geq \frac{\sum_n |A_n|^2 E_0}{\sum_n |A_n|^2} = E_0$$

Qui ho considerato che, in generale, i valori di  $E_{n>0}$  sono maggiori del valore di  $E_0$ , l'energia nello stato fondamentale. Il segno di uguale è valido solo, nel caso molto raro, in cui tutti gli autovalori dell'energia siano degeneri.

Il risultato ottenuto indica che la ricerca del minimo del funzionale  $E$  in uno spazio limitato a funzioni di prova permette di ottenere, al massimo, un limite superiore al vero valore dell'energia ovvero all'autovalore dell'hamiltoniano  $H$ .

ne con spin allineato all'ingiù e il terzo un neutrone con spin all'insù, e quindi fare l'integrale a 9 dimensioni. Il processo continua scambiando la terza componente dello spin ed i nucleoni per formare le possibili 24 configurazioni ed effettuare gli integrali a 9 dimensioni per ognuna delle configurazioni.

Nella Tabella 1 presento i valori del numero di configurazioni di spin e isospin per alcuni nuclei. Per un nucleo medio-pesante come il  $^{40}\text{Ca}$ , il numero di configurazioni è confrontabile con il numero di Avogadro. Anche supponendo di poter eseguire l'integrale a 120 dimensioni in un millisecondo, il tempo di calcolo richiesto è circa dieci volte quella che si stima sia l'attuale età dell'universo.

La tecnica oggi utilizzata per risolvere l'equazione (3) senza fare alcuna approssimazione è detta *Green's function Monte Carlo* [1]. Partendo da una funzione d'onda di prova  $|\Psi_T(t)\rangle$  definita ad un certo tempo  $t$ , questa tecnica fa evolvere

nel tempo il valore di  $E$  e nel limite di tempi molto grandi, infiniti teoricamente ma non numericamente, nella funzione d'onda  $|\Psi_T(t)\rangle$  rimane solo l'autostato di  $H$  con il valore più basso dell'energia, lo stato fondamentale.

Al momento, l'applicazione di questi metodi è stata possibile soltanto per studiare nuclei fino al  $^{12}\text{C}$  [2]. Si tratta di un grande successo dal punto di vista tecnico, e anche teorico, perché ha verificato la validità delle ipotesi che soggiacciono alla scelta di risolvere l'equazione di Schrödinger (1). È comunque evidente che questa metodologia non può essere utilizzata per svolgere uno studio sistematico della maggior parte dei nuclei, ed in generale di sistemi con numero di particelle superiore a poche decine.

Scopo delle teorie a molti corpi è quello di cercare delle soluzioni approssimate dell'equazione di Schrödinger (1). Le soluzioni sono tanto più valide quanto meglio sono controllate le approssimazioni.

Nella ricerca di soluzioni approssimate dell'equazione (1) il principio variazionale svolge un ruolo molto importante. L'idea fondamentale è quella di considerare l'eq. (3) come la definizione del funzionale dell'energia che dipende dalle funzioni d'onda  $\Psi$ . La soluzione del problema consiste così nel trovare il minimo del funzionale.

Nei due riquadri presento in maniera un po' più formale le due proprietà, a mio avviso, più rilevanti del principio variazionale. La prima indica che la ricerca del minimo del funzionale (3) equivale a risolvere l'equazione di Schrödinger (1). Questa proprietà è valida solo se, nella ricerca del minimo, non si pone alcuna restrizione all'espressione che la funzione d'onda  $\Psi$  deve assumere. In realtà il principio variazionale viene utilizzato proprio ipotizzando che  $\Psi$  abbia specifiche caratteristiche, quindi limitando la ricerca del minimo in un sottoinsieme, o meglio sottospazio, dell'insieme totale di tutte le possibili soluzioni del problema variazionale. A questo punto si inserisce la seconda proprietà del principio variazionale, illustrata nel secondo riquadro. L'energia ottenuta minimizzando il funzionale dell'energia in un sottospazio dello spazio delle funzioni d'onda  $\Psi$  fornisce un limite superiore al vero valore dell'energia dello stato fondamentale.

Le due proprietà che ho illustrato forniscono delle regole pragmatiche che chi fa calcoli variazionali ha sempre ben in mente. Di solito, ampliando il sottospazio delle  $\Psi$  nel quale si ricerca il minimo si ottiene un valore inferiore dell'energia. Può succedere che si ottenga anche un valore uguale, ma mai superiore. Dall'altro lato, un calcolo variazionale che ottiene un valore inferiore dell'energia, è migliore, nel senso che si avvicina di più al valore vero. Normalmente, risultati che non rispettano queste semplici regole empiriche hanno problemi, nel calcolo oppure nella definizione stessa del problema variazionale.

Nella sezione seguente presenterò l'applicazione del principio variazionale più conosciuta, ed utilizzata, quella ad un sistema di particelle non interagenti. Nella sezione successiva presenterò alcune applicazioni del principio variazionale al caso di sistemi di particelle interagenti.

## Particelle indipendenti

I sistemi fisici a multicorpi presentano alcune caratteristiche determinate quasi esclusivamente dalle proprietà delle particelle che li compongono, come se queste non interagissero tra loro. Queste evidenze empiriche hanno spinto alla formulazione di modelli a particelle indipendenti.

In questi modelli il sistema è descritto come insieme di particelle che si muovono indipendentemente una dall'altra. Nei modelli più elaborati le particelle sono soggette alla presenza di un campo di forza con il quale interagiscono individualmente e che serve a tenere legato il sistema. Il principio variazionale dà fondamento teorico a questi modelli.

Descrivere il sistema a multicorpi in termini di particelle indipendenti significa considerare che le caratteristiche del sistema sono ottenute come somma delle proprietà di ogni particella. In altre parole, l'operatore hamiltoniano  $H$  che descrive il sistema è espresso come somma di operatori hamiltoniani di ogni singola particella. Questo modello implica la trasformazione

$$H \longrightarrow H_{\text{MF}} = \sum_i h(i) \quad (5)$$

dove la somma è estesa a tutte le particelle, e questa è la convenzione che userò da questo momento in poi, dove il suffisso MF significa *mean-field*. Una possibile soluzione dell'equazione di Schrödinger con l'hamiltoniano  $H_{\text{MF}}$  è data dal prodotto degli autostati degli hamiltoniani  $h$  di singola particella

$$h(i)\phi_{\nu_i}(i) = \epsilon_i\phi_{\nu_i}(i) \quad (6)$$

dove  $\nu$  indica tutti i numeri quantici che caratterizzano lo stato di singola particella.

Nel caso in cui le particelle siano fermioni, cioè abbiano spin semi-intero, come elettroni o nucleoni, la funzione d'onda totale deve essere totalmente antisimmetrica per lo scambio di due particelle. Questo significa che la soluzione fisicamente accettabile del problema è data dalla combinazione lineare di prodotti antisimmetrizzati di funzioni d'onda di singola particella. Matematicamente queste funzioni d'onda a multicorpi possono essere espresse come determinanti di

prodotti di funzioni d'onda. Questi determinanti sono noti in letteratura come determinanti di Slater [3]

$$\Phi(1 \dots N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_{\nu_1}(1) & \dots & \phi_{\nu_N}(1) \\ \vdots & & \vdots \\ \phi_{\nu_1}(N) & \dots & \phi_{\nu_N}(N) \end{vmatrix}$$

Il fattore  $1/\sqrt{N!}$  è inserito in modo che la norma di  $\Phi$  sia 1.

Il principio variazionale è applicato cercando il minimo del funzionale dell'energia nel sottospazio formato dagli stati che sono singoli determinanti di Slater

$$\delta \left( \frac{\langle \Phi | H_{MF} | \Phi \rangle}{\langle \Phi | \Phi \rangle} - \sum_{i,j} \lambda_{ij} \langle \phi_{\nu_i} | \phi_{\nu_j} \rangle \right) = 0 \quad (7)$$

Il secondo termine nelle parentesi rappresenta la condizione che le funzioni d'onda di singola particella  $\phi_{\nu_i}$  devono essere ortonormali. Nell'equazione (7) ho indicato con  $\lambda_{ij}$  i moltiplicatori di Lagrange.

Inserendo nella (7) esplicitamente l'espressione del determinante di Slater ed effettuando la variazione sulle funzioni  $\phi_{\nu_i}$ , è possibile ricavare un sistema di equazioni integro-differenziali non-lineari le cui incognite sono le  $\phi_{\nu_i}$  e  $\lambda_{ij}$

$$-\frac{\hbar^2}{2m_i} \nabla^2 \phi_{\nu_i}(\mathbf{r}) + V(\mathbf{r})\phi_{\nu_i}(\mathbf{r}) - \int d^3 r' W(\mathbf{r}, \mathbf{r}') \phi_{\nu_i}(\mathbf{r}') = \lambda_{ij} \delta_{ij} \phi_{\nu_j}(\mathbf{r}) \quad (8)$$

dove  $\delta_{ij}$  è il simbolo di Kronecker. Nell'equazione precedente il potenziale locale  $V$ , detto di Hartree [4], è definito come

$$V(\mathbf{r}) = \int d^3 r' \sum_{\nu_i} v(\mathbf{r}, \mathbf{r}') \phi_{\nu_i}^*(\mathbf{r}') \phi_{\nu_i}(\mathbf{r}') \quad (9)$$

e il potenziale non locale  $W$ , detto di Fock-Dirac [5], ha l'espressione

$$W(\mathbf{r}, \mathbf{r}') = \sum_{\nu_i} v(\mathbf{r}, \mathbf{r}') \phi_{\nu_i}^*(\mathbf{r}) \phi_{\nu_i}(\mathbf{r}') \quad (10)$$

Si può dimostrare che gli autovalori  $\epsilon_i \equiv \lambda_{ij} \delta_{ij}$  corrispondono all'energia necessaria per estrarre la particella  $i$  dal sistema [6]. Questa è la definizione empirica dell'energia di singola

particella.

Le equazioni Hartree-Fock sono normalmente risolte numericamente con un metodo iterativo. Utilizzando un insieme di funzioni d'onda di prova  $\phi_{\nu_i}^T$  si calcolano i potenziali  $V$  e  $W$  e si inseriscono nella (8) che viene risolta con tecniche numeriche standard fornendo i valori di  $\lambda$  e  $\phi$  per ogni singola particella. Con questo nuovo insieme di stati di singola particella si ripete il processo fino a quando i risultati non convergono. Normalmente la convergenza viene cercata sui valori di  $\lambda$ .

Il metodo Hartree-Fock è quasi esclusivamente utilizzato per la descrizione dello stato fondamentale di sistemi finiti, ovvero sistemi le cui dimensioni globali sono confrontabili con le distanze tra due particelle, quindi sistemi come nuclei, atomi, molecole, cluster molecolari ecc.

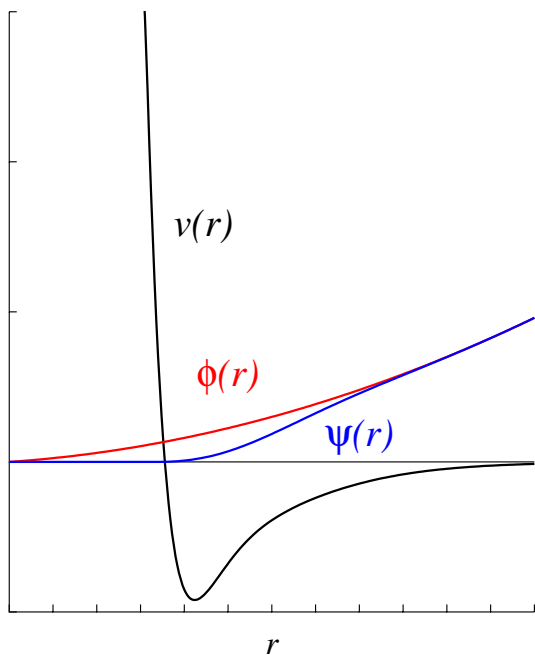
Il confronto tra i risultati ottenuti con il metodo Hartree-Fock con numerosi dati sperimentali è così soddisfacente da far spesso ritenere che il modello a campo medio sia la soluzione del problema a multicorpi. Metto ancora una volta in risalto che la scelta di applicare il principio variazionale a funzioni d'onda a multicorpi che possono essere espresse come singoli determinanti di Slater genera funzioni d'onda che non sono autostati dell'hamiltoniano  $H$ , ma solo di quella parte che può essere espressa come somma di operatori ad un corpo come indica la trasformazione (5). Il modello a particelle indipendenti trasforma, semplificandolo, il problema a molti corpi in molti problemi ad un corpo.

## Particelle interagenti

La parte dell'operatore hamiltoniano trascurata nel modello a particelle indipendenti fa interagire le particelle tra loro ed è la sorgente di tutti i fenomeni collettivi che il sistema presenta. L'applicazione del principio variazionale a sistemi di particelle interagenti implica una ricerca del minimo del funzionale dell'energia in un sottospazio più ampio di quello definito da singoli determinanti di Slater.

Un'espressione di funzioni d'onda di prova molto utilizzata è [7]:

$$\Psi_T(1, \dots, N) = F(1, \dots, N) \Phi(1, \dots, N) \quad (11)$$



**Figura 1:** La funzione  $\phi$  indica la funzione d'onda relativa tra due particelle non interagenti, la funzione  $\psi$  quella tra due particelle che interagiscono con il potenziale  $v$ .

dove  $\Phi$  è un determinante di Slater e  $F$  è detta funzione di correlazione a  $N$  corpi ed è definita come prodotto di correlazioni a due corpi  $f(ij)$

$$F(1, \dots, N) = \prod_{i < j} f(ij) \quad (12)$$

La motivazione fisica alla base di questa scelta della funzione d'onda di prova è legata alla forma del potenziale di interazione tra due particelle, che può essere schematizzata come indicato nella Figura 1. A grandi distanze il potenziale è nullo, a distanze intermedie è attrattivo e, a piccole distanze, presenta un *core* fortemente repulsivo. Questa forma del potenziale è comune a tutti i casi in cui le particelle che interagiscono hanno una struttura interna, indipendentemente dall'origine microscopica dell'interazione tra i componenti della particella. La forma del potenziale tra due nucleoni, le cui distanze sono dell'ordine di  $10^{-15}\text{m}$ , è analoga a quella del potenziale tra due atomi o tra due molecole che interagiscono a distanze centomila volte più grandi. Nonostante la variazione della scala delle distanze, i problemi del sistema a multicorpi da affrontare sono analoghi nel nucleo, nelle molecole, nei liquidi, e nei solidi.

Un modo per rendere evidenti questi problemi consiste nel far tendere all'infinito il valore del potenziale nella regione del *core* repulsivo. Alcune parametrizzazioni dell'interazione tra due nucleoni datate primi anni '70 del secolo scorso hanno questa caratteristica.

Affinché l'equazione di Schrödinger abbia significato fisico in tutto lo spazio, il prodotto tra operatore hamiltoniano, che contiene il potenziale  $v$ , e la sua autofunzione  $\psi$  deve sempre essere finito. Questo implica che se il potenziale tende all'infinito la funzione d'onda deve tendere a zero.

Le soluzioni dell'equazione di Schrödinger cercate utilizzando il principio variazionale con funzioni d'onda da modello a particelle indipendenti non sono nulle nella regione dove il potenziale tende all'infinito. La funzione di correlazione è costruita in modo che

$$\lim_{r_{ij} \rightarrow 0} f(r_{ij}) \phi_{\nu_i}(\mathbf{r}_i) \phi_{\nu_j}(\mathbf{r}_j) = 0 \quad (13)$$

dove  $r_{ij}$  è la distanza tra le due particelle. Questo significa che tutte le funzioni d'onda di prova (11) hanno il corretto comportamento a piccole distanze.

Il funzionale dell'energia ottenuto utilizzando la funzione di prova (11) è

$$E[\Psi_T] = \frac{\langle \Phi | F^+ H F | \Phi \rangle}{\langle \Phi | F^+ F | \Phi \rangle} \quad (14)$$

Confrontando questa espressione con quella del funzionale di campo medio (7) possiamo identificare

$$H_{MF} = F^+ H F \quad (15)$$

Questo aspetto mette in evidenza un fatto troppo spesso trascurato nelle presentazioni del metodo Hartree-Fock. L'interazione da utilizzare nella teoria Hartree-Fock non è quella che si ottiene dall'analisi dei sistemi a due particelle. Inserendo un potenziale che tende all'infinito nelle equazioni Hartree-Fock (8) ci si rende immediatamente conto che non si producono soluzioni finite. Il potenziale effettivo da usare in Hartree-Fock non deve avere *core* fortemente repulsivo nella zone delle piccole distanze di interazione.

Questo aspetto è ben noto in fisica nucleare dove la differenza tra interazione nuda, quella estratta dai sistemi a due particelle, ed effettiva,



quella da usare in calcoli Hartree-Fock, è notevole. La differenza tra queste due interazioni è legata alla rilevanza degli effetti prodotti dalla funzione di correlazione  $F$ . La funzione di correlazione modifica la struttura della funzione d'onda a piccole distanze relative tra le particelle. Quanto più grande è il numero di particelle nel volume caratterizzato dal core repulsivo, quantità detta densità relativa nel gergo della fisica a multicorpi, tanto più importanti sono gli effetti della funzione di correlazione.

La densità relativa di materia nucleare è 0.08 particelle, mentre quella dell'elio liquido di 2.4 particelle. L'elio liquido è relativamente più denso della materia nucleare. In atomi e molecole, le densità relative delle nubi elettroniche sono così basse che l'uso in calcoli Hartree-Fock del potenziale coulombiano nudo, anche se formalmente incorretto, è, pragmaticamente, accettabile.

Nel caso delle particelle indipendenti, trattato nella sezione precedente, la variazione avviene sulla forma delle funzioni d'onda  $\phi_{\nu_i}$  di singola particella. Nel caso della funzione di prova (11) la variazione viene effettuata anche sull'espressione delle funzioni di correlazione  $f$ .

Come si può facilmente immaginare la ricerca del minimo del funzionale (14) è molto più complessa che nel caso della teoria Hartree-Fock. In questo caso non è possibile ottenere un sistema chiuso di equazioni che risolve il problema senza alcuna approssimazione. Un metodo è quello di risolvere il problema del calcolo dell'integrale multidimensionale dell'equazione (14) utilizzando tecniche di integrazione numerica Monte Carlo. In questo caso non ci sono approssimazioni, ma ci si scontra con le difficoltà illustrate nell'introduzione, dovute al grande numero di configurazioni di spin-isospin da considerare.

Il problema della soluzione approssimata dell'equazione (14) è stato affrontato in prima battuta per sistemi infiniti, ovvero quei sistemi a multicorpi che hanno dimensioni totali molto maggiori delle distanze medie tra le particelle interagenti. Questi sistemi hanno invarianza traslazionale e le funzioni d'onda di singola particella sono autostati dell'impulso  $\mathbf{p}_i = \hbar\mathbf{k}_i$ , quindi la loro espressione funzionale è quella dell'onda piana

$$\phi_{\nu_i}(\mathbf{r}) = e^{i\mathbf{k}_i \cdot \mathbf{r}} \quad (16)$$

In questi sistemi la variazione avviene solo sulla funzione di correlazione e questo semplifica il problema. La soluzione del problema variazionale è stata affrontata sviluppando tecniche che analizzano il funzionale di energia come somma del contributo di gruppi di particelle legate dalla funzione di correlazione a due corpi  $f$ . Un'analisi topologica di questi cluster permette di scrivere un sistema chiuso di equazioni integrali che somma tutti i diagrammi di una particolare categoria.

La teoria, formulata inizialmente per sistemi bosonici, prende il nome di Hypernetted Chain (HNC), catene iperconnesse, ed è stata applicata con successo ai liquidi quantistici come l'elio e anche l'acqua [8]. La sua estensione a sistemi fermionici, Fermi Hypernetted Chain (FHNC) [9], è più recente ed ha permesso la sua applicazione allo studio della materia nucleare infinita e della materia di neutroni la cui equazione di stato è necessaria per la comprensione della stabilità delle stelle di neutroni. In anni più recenti la teoria FHNC è stata estesa per essere applicata a sistemi nucleari finiti, nello specifico a nuclei doppio magici [10].

Teorie come HNC e FHNC permettono di utilizzare le interazioni microscopiche tra due particelle perché sono costruite per gestire il problema del core repulsivo. Il confronto tra queste teorie microscopiche e quelle effettive permette di comprendere il ruolo delle correlazioni, ovvero di tutti quei fenomeni che non possono essere descritti dal modello a particelle indipendenti.

## Conclusioni

Il problema quantistico a multicorpi è ben definito: si tratta di risolvere l'equazione di Schrödinger che descrive il sistema. Dal punto di vista pragmatico, quando il numero di particelle del sistema supera le poche decine, il problema non è più risolvibile senza fare approssimazioni. Si cercano, quindi, soluzioni approssimate che siano il più possibile affidabili.

Rispetto ad altre metodologie, i vantaggi del principio variazionale sono legati alla sua semplicità concettuale, che, con poche e ben definite ipotesi permette di definire il problema senza ambiguità. Le difficoltà sono di carattere tecnico e computazionale, ma gli elementi con cui si

lavora, funzioni d'onda e di correlazione, sono ben definiti nell'ambito del formalismo e hanno un significato fisico ben preciso.

L'applicazione del principio variazionale nell'ipotesi che le particelle del sistema non interagiscano tra loro conduce alle equazioni Hartree-Fock. La teoria Hartree-Fock è il fondamento teorico sul quale si basano tutti i modelli a campo medio, ampiamente utilizzati, con successo, per la descrizione di molti osservabili.

D'altra parte, i modelli a campo medio non riescono, per definizione, a descrivere fenomeni collettivi del sistema, e devono essere estesi per considerare l'interazione tra le particelle. Le tecniche più note, ed utilizzate, per descrivere sistemi di particelle interagenti sono ispirate alla teoria dei campi [11], ma l'uso del principio variazionale è una valida alternativa. Le teorie HNC e FHNC, utilizzate per descrivere sistemi infiniti come materia nucleare e neutronica ed elio liquido, ma oggi sviluppate anche per sistemi finiti come nuclei e cluster metallici, sono ottenute applicando il principio variazionale.

Il confronto tra diverse teorie che studiano lo stesso problema da punti di vista differenti è estremamente utile e proficuo.

È parte della storia della fisica nucleare che, negli anni '80 del secolo scorso, l'inconsistenza tra i risultati ottenuti con il principio variazionale e quelli di altre teorie abbia permesso di evidenziare alcune inconsistenze nelle approssimazioni comunemente adottate nei calcoli fatti con queste ultime [12].

Ci sono altre importanti applicazioni del principio variazionale nella fisica dei molti corpi che non ho discusso, ad esempio la sua applicazione nella descrizione di sistemi superfluidi e superconduttivi. Ho scelto esempi che mi hanno permesso di mettere in evidenza l'utilità del principio variazionale nell'affrontare i problemi chiave di questo ambito di ricerca.



[1] M. H. KALOS: "Stochastic wave function for atomic helium", *J. Comp. Phys.* **2** (1967) 257–276.

- [2] S. C. PIEPER AND R. B. WIRINGA: "Quantum Monte Carlo calculations of light nuclei", *Annu. Rev. Nucl. Part. Sc.* **51** (2001) 53–90.
- [3] J. C. SLATER: "The theory of complex spectra", *Phys. Rev.* **34** (1929) 1293–1322.
- [4] D. R. HARTREE: "The distribution of charge and current in an atom consisting of many electrons obeying Dirac's equations", *Proc. Camb. Phil. Soc.* **25** (1929) 225–236.
- [5] V. FOCK: "Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems", *Z. Physik* **61** (1930) 126–148.
- [6] T. KOOPMAN: "Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms", *Physica* **1** (1934) 104–113.
- [7] R. JASTROW: "Many-Body Problem with Strong Forces", *Phys. Rev.* **98** (1955) 1479–1484.
- [8] J. E. MAYER AND M. G. MAYER: *Statistical mechanics*, Wiley, New York (1940).  
E. FEENBERG: *Theory of Quantum Fluids*, Academic Press, New York (1969).  
D. L. GOODSTEIN: *States of matter*, Prentice-Hall, Englewood Cliffs, NJ (1975).
- [9] S. FANTONI AND S. ROSATI: "Jastrow correlations and an irreducible cluster expansion for infinite boson or fermion systems", *Nuov. Cim.* **A20** (1974) 179–193.  
V. R. PANDHARIPANDE AND R. B. WIRINGA: "Variations on a theme of nuclear matter", *Rev. Mod. Phys.* **51** (1979) 821–861.  
S. ROSATI: *FHNC. Variational theory for strongly interacting Fermi systems*, From nuclei to particles, LXXIX Corso Soc. Italiana di Fisica, Bologna (1981).
- [10] F. ARIAS DE SAAVEDRA, C. BISCONTI, G. CO' AND A. FABROCINI: "Renormalized Fermi hypernetted chain approach in medium-heavy nuclei", *Phys. Rep.* **450** (2007) 1–95.
- [11] A. A. ABRIKOSOV, L. P. GORKOV AND I. E. DZIALOSHINSKI: *Methods of quantum field theory in statistical mechanics*, Prentice-Hall, Englewood Cliffs, NJ (1963).  
R. D. MATTUCK: *A guide to Feynmann diagrams in the many-body problem*, McGraw-Hill, New York (1967).  
A. L. FETTER AND J. D. WALECKA: *Quantum theory of many-particle systems*, McGraw-Hill, New York (1971).
- [12] R. GUARDIOLA AND J. ROS, Eds.: *The many-body problem. Jastrow correlations versus Brueckner theory*, Proc. Third Topical School, Granada 1980, Lecture Notes in Physics 138, Springer-Verlag Berlin Heidelberg (1981).

---

**Giampaolo Co':** Professore associato di Fisica Nucleare presso l'Università del Salento si occupa di teorie a molti corpi applicate allo studio della struttura dei nuclei atomici.

---

# Adam Smith, John Nash, il prezzo dell'anarchia e la decadenza della società moderna

**Vittorio Bilò**

*Dipartimento di Matematica e Fisica "Ennio De Giorgi", Università del Salento*

---

**È** idea diffusa tra numerosi studiosi di antropologia che la civiltà occidentale abbia oramai imboccato la strada senza ritorno del declino. Tra le varie cause individuate, spiccano l'eccessivo individualismo, l'indebolimento della fibra morale, la spettacolarizzazione dell'apparenza e dell'immagine a scapito dell'intelletto e della cultura. In questo articolo, proponiamo un'interpretazione socio-economica di questo fenomeno attraverso l'analisi di alcuni risultati teorici recentemente dimostrati nell'ambito della Teoria dei Giochi Algoritmica: una disciplina scientifica che si colloca all'intersezione tra la Teoria dei Giochi e l'Informatica Teorica.

Il mondo sta andando a rotoli. Chissà quanti di voi avranno pronunciato, specie negli ultimi tempi e in maniera più o meno convinta, queste parole! A chi scrive, perlomeno, capita di farlo

molto spesso e cioè ogni qualvolta si ritrovi a dover fronteggiare l'ennesima manifestazione di decadenza morale e culturale che oramai permea irrimediabilmente la nostra società.

"Che sia una nuova, e più subdola, manifestazione del processo darwiniano di selezione naturale?". Intendo dire: "Ci stiamo, forse, inconsapevolmente trasformando negli individui più adatti a sopravvivere sul pianeta Terra del terzo millennio?".

Con ogni probabilità sto traendo delle conclusioni assurde o, giudicate con più indulgenza, semplicemente simpatiche. Tuttavia, esse risultano in maniera del tutto naturale da alcuni risultati analitici dimostrati negli ultimi anni nell'ambito della Teoria dei Giochi Algoritmica: una giovane area scientifica che si colloca all'intersezione tra la Teoria dei Giochi e l'Informatica Teorica. E, seguendo quello "spirito di simpatia" che dovrebbe caratterizzare ogni rivista di divulgazione scientifica che si rivolga a un pubblico ampio ed eterogeneo, cercherò di spiegare perché le conclusioni raggiunte dal filosofo scozzese Adam Smith nella sua teoria della "mano invisibile",

opportunamente rinquadrare in un ambiente di applicazione più appropriato, possano ritenersi corrette e, soprattutto, indicative di un livello di degrado sociale che, pur già preoccupante di per sé, risulta persino meno scoraggiante di quello che potrebbe rivelarsi essere nella realtà.

## Adam Smith e il comportamento egoista

Adam Smith, filosofo scozzese del XVIII secolo, è universalmente ritenuto il padre della moderna Economia Politica e, in particolare, del Liberismo economico. Egli, infatti, da fiero sostenitore del libero mercato, asseriva che l'imposizione di una qualsiasi regolamentazione sull'attività dei singoli finisse con l'arrecare nocimento al fermento economico di una comunità, frenandone a tutti gli effetti la crescita, senza, per questo, apportarle alcun beneficio di rimando. A suggello del suo pensiero, egli elaborò il seguente postulato, noto col nome di **teoria della mano invisibile**:

*Nel libero mercato, se ogni individuo si adoperava per massimizzare il proprio tornaconto economico, allora anche il benessere di tutta la collettività risulta massimizzato.*

Tralasciando la visione machiavelliana, del fine che giustifica i mezzi, sulle modalità con le quali si possa operare al fine di massimizzare il proprio tornaconto, risulta subito lampante come la chimera del benessere collettivo venga degradata al ruolo di propulsore ed esaltatore del comportamento egoista degli individui, il tutto a scapito di ogni forma di sviluppo basata sulla collaborazione e sul mutuo sostegno tra le parti.

All'inizio del XX secolo, l'economista francese Léon Walras fornì una prova formale del postulato della mano invisibile dimostrando che, in condizioni di *concorrenza perfetta*, è possibile determinare un sistema di prezzi d'equilibrio che comporta l'eguaglianza tra domanda e offerta in tutti i mercati, nonché l'eguaglianza tra costo di produzione e prezzo di vendita per ciascun bene e per ciascun imprenditore.

Che il postulato proposto da Smith, poi, fosse vero o, in qualche forma, approssimativamente vero anche al di fuori del contesto del libero mercato in regime di concorrenza perfetta non era

facile da stabilire a quei tempi, data la mancanza di un'appropriata teoria matematica in grado di modellare e predire il comportamento strategico di soggetti razionali portatori di interessi socio-economici eterogenei. Soltanto nel secondo dopo guerra, la nascita della Teoria dei Giochi avrebbe posto le basi per lo studio analitico di tali situazioni.

## La teoria dei giochi e l'equilibrio di Nash

Quali saranno i percorsi scelti dagli abitanti di una città per recarsi al lavoro? In quale punto strategico conviene installare una nuova base militare? Qual è la maniera più remunerativa per vendere un bene a un'asta pubblica? Questi sono soltanto alcuni dei tanti problemi studiati nel tempo attraverso la Teoria dei Giochi. Essa nacque nel 1944, con la pubblicazione del libro *"Theory of Games and Economic Behavior"* di John von Neumann e Oscar Morgenstern, allo scopo di modellare situazioni in cui dei soggetti razionali (detti, per l'appunto, **giocatori**) si trovano a dover prendere delle decisioni strategiche e il risultato conseguito da ognuno di loro dipende sia dalla propria scelta che da quelle effettuate dagli altri. La teoria così sviluppata si limitava, tuttavia, a considerare situazioni in cui i giocatori potevano coordinare le proprie scelte al fine di ottenere risultati migliori: i cosiddetti *giochi cooperativi*.

Fu John F. Nash a introdurre, alla fine degli anni '40, la nozione di *gioco non cooperativo* [1], nel quale ogni giocatore agisce egoisticamente al solo scopo di ottenere il risultato migliore possibile per se stesso, senza comunicare, né tantomeno cooperare con nessun altro giocatore, o gruppi di giocatori. Inizialmente, le idee di Nash furono snobbate dallo stesso von Neumann, il quale riteneva che il futuro della Teoria dei Giochi si sarebbe basato sui giochi cooperativi, ma con il passare degli anni l'attenzione del mondo economico si riversò sempre più sullo studio della variante proposta da Nash, tant'è che nel 1994 egli venne insignito del premio Nobel per l'economia per la sua "analisi pionieristica degli equilibri nella teoria dei giochi non cooperativi".

Secondo la visione di Nash, che può essere a tutti gli effetti riconosciuta come un'estensione di quella teorizzata da Smith, il comportamento di un giocatore è modellato sulla base di tre ipotesi fondamentali:

1. **egoismo:** il giocatore ha come unico obiettivo l'ottimizzazione del proprio tornaconto personale (massimizzazione dei propri profitti o minimizzazione dei propri costi) e non si preoccupa del fatto che il suo comportamento possa andare a scapito degli altri soggetti che partecipano al gioco;
2. **razionalità:** il giocatore, nel perseguire il proprio obiettivo, si comporta in maniera logica e intelligente, effettuando sempre le scelte più ragionevoli;
3. **non cooperativismo:** il giocatore, nel perseguire il proprio obiettivo, non può creare accordi né coalizioni con altri soggetti che partecipano al gioco.

È chiaro che l'ipotesi di razionalità non possa ammettere una formalizzazione rigorosa e che la sua interpretazione debba essere demandata, di volta in volta, all'intuizione del momento.

Formalmente, un gioco non cooperativo (in forma strategica) è definito come segue.

**Definizione 1.** *Un gioco non cooperativo in forma strategica è una tripla  $SG = ([n], \Sigma_{i \in [n]}, U_{i \in [n]})$ , dove  $[n] := \{1, 2, \dots, n\}$  è l'insieme dei giocatori e, per ogni  $i \in [n]$ ,  $\Sigma_i$  è l'insieme delle scelte strategiche a disposizione del giocatore  $i$  e  $U_i : \times_{i \in [n]} \Sigma_i \rightarrow \mathbb{R}$  è la sua funzione di utilità.*

Chiamiamo **insieme dei profili strategici** l'insieme  $\Sigma = \times_{i \in [n]} \Sigma_i$  e, analogamente, chiamiamo **profilo strategico** un qualsiasi elemento  $\sigma \in \Sigma$ . Intuitivamente, un profilo strategico è il risultato che si ottiene quando ogni giocatore ha effettuato una scelta strategica. Ad esempio, scriveremo  $\sigma = (\sigma_1, \dots, \sigma_n)$  per modellare la situazione in cui ogni giocatore  $i \in [n]$  ha scelto la strategia  $\sigma_i \in \Sigma_i$  dando vita al profilo strategico  $\sigma$ . Com'è possibile vedere, la definizione delle  $U_i$ , essendo dipendente da  $\Sigma$ , fa sì che l'utilità ottenuta dal giocatore  $i$  nel profilo strategico  $\sigma$ , ossia il valore  $U_i(\sigma)$ , non dipenda soltanto dalla sua scelta strategica, ma anche da quelle effettuate dagli altri. Assumeremo implicitamente che ogni giocatore

cercherà di minimizzare la propria utilità, intesa quindi, d'ora in avanti, come misura di un costo che ognuno deve inevitabilmente sostenere.

Una volta stabilite quelle che sono le variabili in gioco (il modo di dire è decisamente appropriato in questo contesto), siamo interessati a predire, in maniera più o meno esatta, il comportamento che verrà tenuto da ognuno dei giocatori.

Siano dati un profilo strategico  $\sigma$  e una strategia  $\tau \in \Sigma_i$  per il giocatore  $i$ . Con la notazione  $\sigma_{-i} \diamond \tau$  indichiamo il profilo strategico  $(\sigma_1, \dots, \sigma_{i-1}, \tau, \sigma_{i+1}, \dots, \sigma_n)$  ottenuto da  $\sigma$  per effetto del cambiamento di strategia da  $\sigma_i$  a  $\tau$  da parte del giocatore  $i$ . Si noti come il profilo strategico  $\sigma_{-i} \diamond \tau$  differisca da  $\sigma$  solo per la scelta di tale giocatore.

**Definizione 2.** *Dato un profilo strategico  $\sigma \in \Sigma$ , la strategia  $\tau \in \Sigma_i$  è una **mossa migliorativa** per il giocatore  $i$  in  $\sigma$  se  $U_i(\sigma_{-i} \diamond \tau) < U_i(\sigma)$ .*

Una mossa migliorativa è, quindi, un cambio di strategia che porta un beneficio al giocatore che la effettua.

Siano dati ora un profilo strategico  $\sigma$  e una mossa migliorativa  $\tau$  per il giocatore  $i$  in  $\sigma$ . Sulla base delle ipotesi fatte sulla natura dei giocatori, diventa naturale aspettarsi che il profilo strategico  $\sigma$  non sarà mai un risultato plausibile del gioco in quanto il giocatore  $i$  vorrà sempre cambiare la propria scelta da  $\sigma_i$  a  $\tau$  e migliorare la propria situazione. Su questa semplice osservazione si basa la definizione di Equilibrio di Nash in strategie pure che è uno dei concetti che meglio di tutti incarnano la nozione di soluzione di un gioco non cooperativo.

**Definizione 3.** *Il profilo strategico  $\sigma \in \Sigma$  è un **Equilibrio di Nash in strategie pure** se per ogni  $i \in [n]$  e per ogni  $\tau \in \Sigma_i$ , vale  $U_i(\sigma_{-i} \diamond \tau) \geq U_i(\sigma)$ .*

Un Equilibrio di Nash in strategie pure è quindi un profilo strategico in cui nessun giocatore possiede una mossa migliorativa.

Un altro concetto fondamentale, fortemente legato a quello di mossa migliorativa, è la nozione di *contromossa migliore*. Il modo più semplice per illustrare questo concetto lo si ottiene pensando che il gioco si svolga a turni (come per i giochi da tavolo in genere) e che, a ogni turno, un giocatore decida se e come cambiare la propria

scelta strategica. Ovviamente, ci aspettiamo che il giocatore modificherà la propria scelta se si trova in un profilo strategico per il quale possiede una mossa migliorativa, ma come si comporterà quando le mosse migliorative a sua disposizione sono più di una? L'ipotesi di razionalità ci porta a pensare che egli vorrà effettuare la scelta che gli dia il massimo beneficio nell'immediato, ossia scegliere la mossa migliorativa che gli consenta di migliorare la propria utilità il più possibile (in questo caso si dice che il giocatore è **miope**). Formalmente, possiamo enunciare la seguente definizione.

**Definizione 4.** Sia dato un profilo strategico  $\sigma \in \Sigma$ . Una strategia  $\tau \in \Sigma_i$  è una **contromossa migliore** per il giocatore  $i$  in  $\sigma$  se  $U_i(\sigma_{-i} \diamond \tau) = \min_{\tau' \in \Sigma_i} \{U_i(\sigma_{-i} \diamond \tau')\}$ .

A volte si è tentati di pensare che una contromossa migliore per il giocatore  $i$  in  $\sigma$  sia la migliore tra le sue mosse migliorative in  $\sigma$ . Questo, tuttavia, risulta vero solo quando l'insieme delle mosse migliorative a disposizione del giocatore  $i$  in  $\sigma$  è non vuoto. Infatti, quando il giocatore  $i$  non ha mosse migliorative in  $\sigma$ , questo erroneo presupposto ci porterebbe a dire che egli non ha una contromossa migliore in  $\sigma$  (l'insieme vuoto non ha elementi), mentre la strategia  $\sigma_i$  da egli correntemente adottata in  $\sigma$  costituisce una contromossa migliore sulla base della Definizione 4. Non è difficile vedere che un profilo strategico  $\sigma$  è un Equilibrio di Nash in strategie pure se, per ogni  $i \in [n]$ , la strategia  $\sigma_i$  è una contromossa migliore per il giocatore  $i$  in  $\sigma$ .

L'Equilibrio di Nash in strategie pure è un concetto tanto semplice quanto efficace per caratterizzare quelli che possono essere ritenuti comportamenti plausibili con le ipotesi di egoismo, razionalità e non cooperativismo dei giocatori. Tuttavia, esso non gode di alcune di quelle proprietà fondamentali, quali *esistenza* e *unicità*, che, di solito, ci si attende dal concetto di soluzione di un dato problema.

Nash risolse la questione dell'esistenza di soluzioni all'equilibrio dando ai giocatori la possibilità di utilizzare strategie probabilistiche (dette *strategie miste*) e dimostrando che, sotto queste nuove ipotesi, ogni gioco finito (ossia con un numero finito di giocatori e un numero finito di

strategie per giocatore) ammette sempre almeno un *Equilibrio di Nash in strategie miste* [2].

L'Equilibrio di Nash in strategie miste, pur essendo ormai universalmente riconosciuto e utilizzato come concetto standard di soluzione per i giochi non cooperativi, è stato da sempre oggetto di forti critiche, sia per il fatto di assumere strategie probabilistiche, sia perché esso richiede che ogni giocatore possieda un'informazione completa su tutte le possibili strategie a disposizione degli altri partecipanti al gioco. Per questi motivi, l'Equilibrio di Nash in strategie pure, che non necessita di alcuna di queste assunzioni, è il concetto di soluzione da preferirsi in tutte le situazioni in cui esso esiste.

Concludiamo questa sezione illustrando i concetti appena introdotti attraverso quello che, con ogni probabilità, risulta essere l'esempio più famoso di gioco non cooperativo: il *Dilemma del Prigioniero*.

**Esempio 1.** La polizia ha arrestato due noti pregiudicati fortemente sospettati di aver commesso un omicidio. Gli inquirenti hanno prove schiaccianti per farli condannare per traffico di stupefacenti, ma non hanno abbastanza prove per ottenere una condanna per omicidio, a meno che uno dei due non tradisca l'altro confessando il reato commesso. Per questi motivi, gli investigatori propongono a ognuno dei due sospettati (che sono tenuti in celle separate e, quindi, non possono fare patti preventivi) lo stesso tipo di accordo. Nel caso in cui uno solo dei due confessasse, egli otterrebbe un sensibile sconto di pena, venendo condannato a un solo anno di carcere, mentre all'altro ne toccherebbero venti. Tuttavia, se entrambi confessassero, lo sconto sarebbe molto più modesto, così che entrambi finirebbero con lo scontare quindici anni di prigione. Infine, nel caso in cui nessuno dei due parlasse, essi verrebbero condannati a due anni di carcere per traffico di stupefacenti.

La rappresentazione grafica del gioco in questione è riportata nella Figura 1.

Dall'analisi delle utilità riportate in figura, è possibile notare come nel profilo strategico (*Tace, Confessa*) il primo giocatore disponga di una mossa migliorativa: la strategia di confessare. In questo caso, infatti, la sua utilità scende da 20 a 15, anche se ciò comporta un peggioramento dell'utilità del secondo giocatore che passa da 1 a 15. Tale strategia, essendo l'unica mossa miglio-

	Tace	Confessa
Tace	(2,2)	(20,1)
Confessa	(1,20)	(15,15)

**Figura 1:** Rappresentazione grafica del gioco del Dilemma del Prigioniero.

rativa per il primo giocatore, costituisce anche una contromossa migliore. Il secondo giocatore, invece, non possiede alcuna mossa migliorativa, ma possiede, ovviamente, una contromossa migliore, ossia la strategia di confessare che è quella da lui correntemente utilizzata. Infine, il profilo strategico (*Confessa, Confessa*) costituisce l'unico Equilibrio di Nash in strategie pure del gioco.

È interessante notare (e in questo risiede la significatività di questo gioco) come il profilo strategico (*Tace, Tace*), che costa a entrambi i giocatori soltanto 2 anni di prigione, sia di gran lunga migliore di quello all'equilibrio. Tuttavia, le ipotesi di egoismo e razionalità non rendono possibile il raggiungimento di una simile situazione ideale. Infatti, anche se i due sospettati si accordassero preventivamente sul comportamento da tenere (possibilità che abbiamo escluso a priori in quanto in presenza di giocatori non cooperativi e che richiederebbe comunque che essi siano informati delle condizioni proposte prima di essere separati), l'idea di evitare un anno di prigione porterebbe comunque ognuno di loro a tradire la parola data.

## Il prezzo dell'anarchia

Come abbiamo visto alla fine della sezione precedente, il concetto di inefficienza delle soluzioni all'equilibrio rispetto a soluzioni idealmente ottimali, ma che non sono realizzabili in pratica poiché in contrasto con le ipotesi di egoismo, razionalità e non cooperativismo dei giocatori, risulta noto sin dai tempi dell'introduzione del gioco del Dilemma del Prigioniero: nella soluzione all'equilibrio i due prigionieri finiscono con lo scontare molti più anni di carcere di quelli che avrebbero fatto se si fossero accordati sulla strate-

gia comune di non confessare il loro reato. Ritornando indietro alla visione della mano invisibile di Adam Smith, in questo caso siamo in presenza di una situazione in cui il comportamento egoista dei giocatori non porta alla massimizzazione del loro benessere.

Gli economisti hanno da sempre accettato questa situazione come un dato di fatto, senza porsi ulteriori domande, in particolare sull'entità di tale inefficienza nei vari giochi, specie in quelli che modellano situazioni di comportamento reale. È stato a seguito dell'interessamento alla Teoria dei Giochi da parte della comunità scientifica degli informatici, da sempre preoccupati delle prestazioni dei sistemi utilizzati nella pratica, che il problema di quantificare la distanza tra le soluzioni all'equilibrio e quelle ottimali che si possono ottenere rilassando le ipotesi di egoismo, razionalità e non cooperativismo dei giocatori è diventato oggetto di studio sistematico.

Dato un gioco non cooperativo SG, una **funzione sociale** per SG è una funzione  $SF : \Sigma \rightarrow \mathbb{R}$  che associa ad ogni profilo strategico  $\sigma \in \Sigma$  un valore  $SF(\sigma)$ , detto **valore sociale** di  $\sigma$ , che esprime il grado di soddisfazione percepito dall'insieme dei giocatori in  $\sigma$ . La funzione sociale più naturale per esprimere il benessere raggiunto dalla collettività dei giocatori è la funzione  $SUM(\sigma) = \sum_{i \in [n]} U_i(\sigma)$ , ossia la somma delle utilità realizzate da tutti i giocatori nel profilo strategico  $\sigma$ : tanto più basso sarà questo valore, tanto più alto sarà il benessere sociale della collettività.

Dato un gioco non cooperativo SG, detti  $PNE(SG)$  l'insieme degli Equilibri di Nash in strategie pure di SG e  $\sigma^*(SG)$  il profilo strategico che minimizza la funzione sociale  $SUM$ , si definisce **prezzo dell'anarchia** (PoA) di SG il rapporto tra il valore sociale del peggior Equilibrio di Nash in strategie pure e il valore sociale dell'ottimo [3], ossia

$$PoA(SG) = \max_{\sigma \in PNE(SG)} \left\{ \frac{SUM(\sigma)}{SUM(\sigma^*(SG))} \right\}$$

Il prezzo dell'anarchia quantifica, quindi, la massima perdita in efficienza che i giocatori *possono* subire a causa del loro comportamento egoista: più esso cresce e più il benessere sociale garantito dalla peggior soluzione all'equilibrio diminuisce. Tale nozione si estende naturalmente a un'intera

classe di giochi considerando il caso peggiore su tutte le possibili istanze della classe. Data una classe di giochi  $\mathcal{C}$ , definiamo

$$\text{PoA}(\mathcal{C}) = \sup_{\text{SG} \in \mathcal{C}} \{\text{PoA}(\text{SG})\}$$

## I giochi di congestione

Introduciamo, in questa sezione, una classe di giochi molto studiata in pratica: la classe dei giochi di congestione.

**Definizione 5.** Un **gioco di congestione** è una quadrupla  $\text{SG} = ([n], R, \Sigma_{i \in [n]}, \ell_{j \in R})$ , dove  $[n]$  è l'insieme dei giocatori,  $R$  è un insieme di risorse, e l'insieme delle scelte strategiche a disposizione di ogni giocatore  $i$  è un qualsiasi sottoinsieme non vuoto dell'insieme delle parti di  $R$ , ossia  $\Sigma_i \subseteq \mathcal{P}(R) \setminus \emptyset$ . Dato un profilo strategico  $\sigma$ , definiamo **congestione** della risorsa  $r_j$  in  $\sigma$ , e lo indichiamo con  $n_j(\sigma)$ , il numero di giocatori che utilizzano  $r_j$  quando  $\sigma$  viene realizzato. Formalmente,  $n_j(\sigma) = |\{i \in [n] : r_j \in \sigma_i\}|$ . Ogni risorsa  $r_j \in R$  ha una **funzione di costo**  $\ell_j : [n] \rightarrow \mathbb{R}_{\geq 0}$  che dipende esclusivamente dalla propria congestione. L'utilità realizzata dal giocatore  $i$  in  $\sigma$  è definita come  $U_i(\sigma) = \sum_{r_j \in \sigma_i} \ell_j(n_j(\sigma))$ .

I giochi di congestione costituiscono una classe di giochi molto importante e molto studiata in letteratura grazie al fatto che essi ammettono sempre Equilibri di Nash in strategie pure [4].

Andremo, ora, a illustrare tre risultati teorici, decisamente sorprendenti e controintuitivi, che si manifestano nelle seguenti sottoclassi dei giochi di congestione:

- **giochi di congestione lineari**,
- **giochi di condivisione dei costi su reti di comunicazione multicasting** (ai quali ci riferiremo d'ora in avanti come *giochi multicast*),
- **giochi di taglio**.

### Giocatori altruisti nei giochi di congestione lineari

I giochi di congestione lineari corrispondono alla sottoclasse dei giochi di congestione in cui, per ogni risorsa  $r_j \in R$ , la funzione di costo  $\ell_j$  risulta essere lineare nella sua congestione. È stato

dimostrato in [5] che il prezzo dell'anarchia di questa classe di giochi è uguale a  $5/2$ , il che vuol dire che ogni possibile soluzione all'equilibrio generata da giocatori egoisti risulta avere un grado di benessere collettivo pari almeno al 40% di quello ottimo. Una domanda che nasce spontanea, quindi, è quella di capire se giocatori maggiormente altruistici e cooperativi, che abbiano un occhio di riguardo anche al benessere collettivo oltre che al proprio tornaconto personale, possano produrre equilibri migliori abbassando ulteriormente il prezzo dell'anarchia.

Diciamo che il giocatore  $i \in [n]$  è  $\gamma$ -altruista, con  $\gamma \in [0, 1]$ , se è interessato a minimizzare la seguente funzione di utilità:

$$U_i(\sigma) = \gamma \sum_{k \in [n]: k \neq i} \sum_{r_j \in \sigma_k} \ell_j(n_j(\sigma)) + (1 - \gamma) \sum_{r_j \in \sigma_i} \ell_j(n_j(\sigma))$$

Sulla base di tale definizione, un giocatore  $\gamma$ -altruista vorrà minimizzare una combinazione lineare tra il suo costo personale e quello sostenuto da tutti gli altri giocatori. In questa maniera, un giocatore 0-altruista è un classico giocatore egoista, mentre un giocatore 1-altruista è un giocatore puramente altruista che si adopera soltanto per migliorare il benessere degli altri senza prendere in considerazione il proprio costo personale.

È stato dimostrato in [6] che il prezzo dell'anarchia nei giochi di congestione lineari con giocatori  $\gamma$ -altruisti è uguale a  $\frac{5-\gamma}{2-\gamma}$  per  $\gamma \in [0, 1/2]$  e  $\frac{2-\gamma}{1-\gamma}$  per  $\gamma \in [1/2, 1]$ . Com'è possibile notare, quindi, il prezzo dell'anarchia cresce dal valore  $5/2$  al valore 3 per  $\gamma$  che cresce da 0 a  $1/2$ , per poi iniziare ad aumentare in maniera molto più marcata fino a tendere verso l'infinito per  $\gamma$  che tende a 1. Ne consegue che la situazione che produce equilibri aventi il miglior grado di benessere sociale nel caso peggiore è quella in cui i giocatori sono puramente egoisti. Abbiamo, quindi, scoperto un nuovo contesto applicativo in cui la teoria della mano invisibile di Adam Smith trova una giustificazione analitica.

### Giocatori ignoranti nei giochi *multicast*

I giochi *multicast* sono una sottoclasse dei giochi di congestione che modellano la creazione spon-



tanea di reti di comunicazione. Esiste un nodo speciale nella rete, detto sorgente, che possiede un'informazione appetita dai vari giocatori (ad esempio, un film, un concerto, un aggiornamento software). Ogni giocatore vuole connettersi alla sorgente, in modo da ricevere l'informazione desiderata, attraverso un insieme di link trasmissivi che possono essere instaurati a un certo costo. Quando uno stesso link risulta condiviso da più giocatori, il suo costo viene ripartito equamente tra ognuno di essi.

Il prezzo dell'anarchia di questa classe di giochi è molto alto, essendo pari al numero di giocatori  $n$ . Ci chiediamo nuovamente se sussiste la possibilità di ottenere equilibri migliori cambiando leggermente la natura dei giocatori. In questo caso, ci concentreremo sul grado di conoscenza posseduta da ognuno di loro nella seguente maniera: associamo, a ogni giocatore  $i \in [n]$ , l'insieme  $K_i \subseteq [n]$  di tutti i giocatori dei quali  $i$  conosce la scelta strategica effettuata. Il giocatore  $i$ , non disponendo di una conoscenza completa del profilo strategico  $\sigma$  in cui si viene a trovare, non è più in grado di valutare correttamente la congestione  $n_e(\sigma)$  per ogni arco  $e \in E$ , ma ne potrà stimare soltanto un valore approssimato  $n_e^i(\sigma) \leq n_e(\sigma)$ . In questo contesto, la sua utilità in  $\sigma$  verrà, quindi, ridefinita come

$$U_i(\sigma) = \sum_{e \in \sigma_i} \ell_j(n_e^i(\sigma))$$

È stato dimostrato in [7] che, se ogni giocatore  $i \in [n]$  conosce soltanto le scelte strategiche effettuate dall'insieme dei giocatori  $K_i = \{1, \dots, i\}$ , allora il prezzo dell'anarchia scende a  $O(\log^2 n)$ , mostrando come il benessere sociale aumenti di un fattore esponenziale in presenza di giocatori ignoranti.

## Giocatori lungimiranti nei giochi di taglio

I giochi di taglio, infine, sono una sottoclasse dei giochi di congestione che possono essere definiti attraverso la seguente, semplice, metafora politica. Ogni giocatore  $i \in [n]$  deve decidere se schierarsi col partito di destra o con quello di sinistra. Per ogni coppia di giocatori  $i, j \in [n]$ , il valore  $c_{ij}$  quantifica l'insoddisfazione che  $i$  e  $j$  provano quando vengono a trovarsi nello stesso

partito (possiamo pensare al valore  $c_{ij}$  come al livello di rivalità politica che intercorre tra i due giocatori). La funzione di utilità del giocatore  $i$  è dunque definita come

$$U_i(\sigma) = \sum_{j \in [n]: \sigma_j = \sigma_i} c_{ij}$$

ossia, ogni giocatore vuole stare nel partito (non importa se di destra o di sinistra) in cui incontra il minor livello di rivalità politica.

Piuttosto che considerare gli Equilibri di Nash in strategie pure per questo gioco, prendiamo in esame le soluzioni che si generano nella seguente maniera: a partire dalla situazione in cui nessuno ha ancora effettuato scelte strategiche, ogni giocatore, a turno, decide a quale partito afferire conoscendo le scelte già effettuate dai suoi predecessori. Il primo giocatore potrà, quindi, scegliere indifferentemente uno dei due schieramenti e supponiamo opti per quello di destra; il secondo giocatore, nel caso in cui  $c_{12} > 0$  sceglierà il partito di sinistra; il terzo giocatore sceglierà di andare a sinistra se  $c_{23} < c_{13}$ , e così via. In generale, ogni giocatore  $i \in [n]$  sceglierà la strategia che costituisce una contromossa migliore nel profilo strategico ristretto alle scelte effettuate dai primi  $i - 1$  giocatori (giocatore miope). Il prezzo dell'anarchia delle soluzioni così generate è pari a 2.

Un limite intrinseco dei giocatori miopi è che ognuno di loro decide di effettuare la scelta strategica in grado di garantirgli il minor costo possibile nell'immediato, senza ponderare il fatto che la sua scelta potrebbe essere fortemente condizionata da quelle effettuate dai suoi successori. In quest'ottica, si possono prendere in considerazione giocatori **lungimiranti** che valutino tutte le possibili situazioni che si possono creare a seguito della propria scelta strategica e, sulla base di una tale analisi esaustiva, scelgano la loro contromossa migliore.

È stato dimostrato in [8] che il prezzo dell'anarchia relativo alle soluzioni generate da giocatori lungimiranti sale a 3, mostrando, quindi, come il benessere sociale diminuisca all'aumentare del livello di razionalità dei giocatori.

## Conclusioni

Dall'analisi di alcuni risultati teorici relativi al prezzo dell'anarchia in tre giochi non cooperativi di rilevante applicazione pratica, è possibile concludere che nella società del terzo millennio, in cui il faro guida dell'umanità è la ricerca spasmodica del benessere personale e collettivo, trovi piena esaltazione l'*homo oeconomicus*: l'evoluzione egoista, ignorante e miope dell'*homo sapiens*. Ma non biasimiamoci: non è colpa nostra, ma soltanto della mano invisibile teorizzata da Adam Smith. Una mano invisibile che, nell'indicarci la strada che porta a una fantomatica ricchezza di massa, ci spinge inesorabilmente verso il baratro della decadenza sociale.



- [1] J. F. NASH: "Non-Cooperative Games", *Annals of Mathematics*, **54(2)** (1951) 286–295.
- [2] J. F. NASH: "Equilibrium Points in  $n$ -Person Games", *Proceedings of the National Academy of Sciences of the United States of America*, **36** (1950) 48–49.
- [3] E. KOUTSOPIAS, C. PAPADIMITRIOU: "Worst-case Equilibria", *Proceedings of the 16th International Symposium on Theoretical Aspects of Computer Science (STACS)*, **LNCS 1653** (1999) 404–413.

- [4] R. W. ROSENTHAL: "A Class of Games Possessing Pure-Strategy Nash Equilibria", *International Journal of Game Theory*, **2** (1973) 65–67.
- [5] G. CHRISTODOULOU, E. KOUTSOPIAS: "The Price of Anarchy of Finite Congestion Games", *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC)*, (2005) 67–73.
- [6] I. CARAGIANNIS, C. KAKLAMANIS, P. KANELLOPOULOS, M. KYROPOULOU, E. PAPAIOANNOU: "The Impact of Altruism on the Efficiency of Atomic Congestion Games", *Proceedings of the 5th International Symposium on Trustworthy Global Computing (TGC)*, **LNCS 6084** (2010) 172–188.
- [7] V. BILÒ, A. FANELLI, M. FLAMMINI, L. MOSCARDELLI: "When Ignorance Helps: Graphical Multicast Cost Sharing Games", *Theoretical Computer Science*, **411(3)** (2010) 660–671.
- [8] V. BILÒ, M. FLAMMINI, G. MONACO, L. MOSCARDELLI: "Some Anomalies of Farsighted Strategic Behavior", *Proceedings of the 10th Workshop on Approximation and Online Algorithms (WAOA)*, **LNCS 7846** (2012) 229–241.



**Vittorio Bilò:** è un ricercatore in informatica presso l'Università del Salento. I suoi interessi di ricerca spaziano nell'ambito della Teoria degli Algoritmi e della Complessità Computazionale, con particolare riferimento alla Teoria dei Giochi Algoritmica.

---

# La fine del ghiaccio artico ed il futuro del clima

**Francesco Paparella**

Dipartimento di Matematica & Fisica "Ennio De Giorgi" - Università del Salento

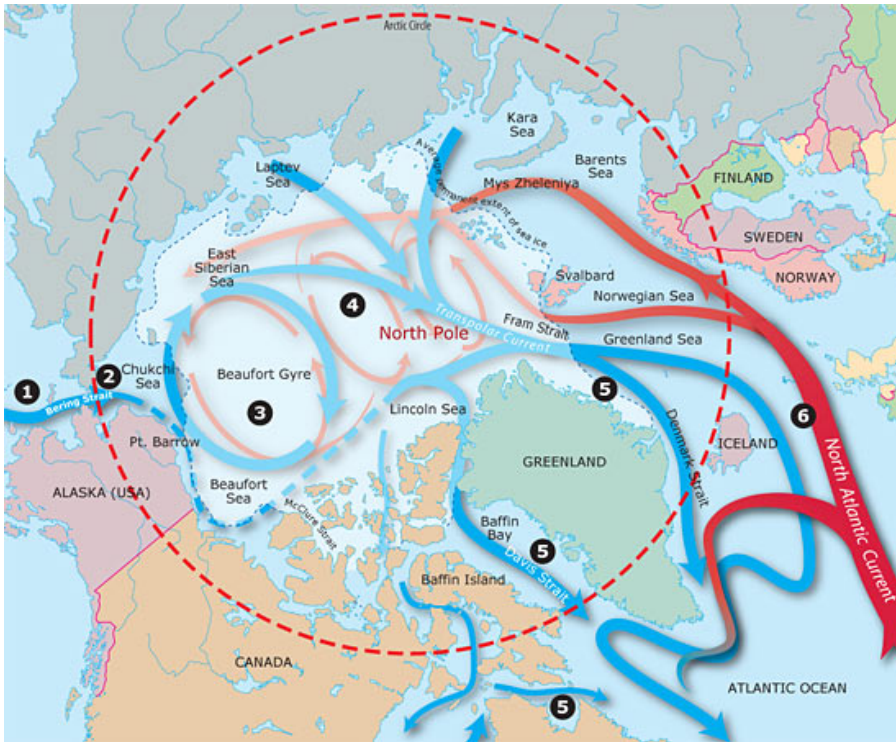
---

**I**l dibattito riguardo ai cambiamenti climatici furoreggia sui mezzi di comunicazione di massa. All'indomani di alluvioni, periodi di siccità, tempeste di neve o ondate di calore c'è chi paventa i cambiamenti climatici. Ma questo genere di disastri è sempre esistito, e nessuno di essi, preso singolarmente, può dare informazioni riguardo al clima. Esiste dunque un evento che, se si verificasse, sarebbe da solo la prova inconfutabile che il nostro pianeta sta per entrare in un nuovo regime climatico? Un simile evento esiste, ed il fatto che verosimilmente si verificherà entro qualche decennio non è una buona notizia.

Riguardo all'uragano che ha investito New York alla fine di Ottobre 2012, l'autorevolissimo New York Times scrive: *Sandy non può essere considerato un disastro stagionale o una coincidenza regionale, ma è un ulteriore segnale delle calamità che ci aspettano in un'epoca di cambiamenti climatici* [1]. Siamo abituati a sentire commenti analoghi dopo ogni evento meteorologico estremo che abbia causato vittime o ingenti danni. Nell'immediatezza dell'evento può capitare che amministratori e politici si autoassolvano da miopi scelte del passato usando i cambiamenti climatici come utile capro espiatorio, salvo poi omettere di intraprendere costose e forse impopolari decisioni sulla

gestione del territorio che mitighino i danni di altri eventi del genere. Atteggiamenti di questo tipo, privi di qualunque giustificazione scientifica, rendono forti le argomentazioni di coloro che negano che sia davvero in atto un cambiamento climatico: i clima-scettici. A New York nessuno ha ancora dimenticato Sandy, ma tutti hanno osservato che durante quest'anno si sono formate solo tredici grandi tempeste atlantiche (contro diciannove nel 2012) e solo due di esse sono state un uragano [2] (contro dieci dell'estate scorsa). Quest'anno le notizie tragiche sono arrivate dalle Filippine, dove il passaggio del tifone Haiyan / Yolanda ha causato una orrenda catastrofe. Ma se anche questo tifone, come molti altri prima di lui avesse proseguito la sua corsa rimanendo sul Pacifico, anziché attraversare una zona densamente popolata, di lui sarebbe rimasta solo una riga negli annali della meteorologia. Per i clima-scettici è facile dichiarare che il re è nudo, e denunciare come ipocrite le campagne degli attivisti ambientali [3].

In realtà, solo il lento accumularsi dell'evidenza statistica può mostrare se la frequenza degli eventi meteorologici estremi stia aumentando oppure no, e questo processo si scontra quotidianamente con la necessità di confrontare le statistiche relative al presente con quelle assai lacunose che riguardano il passato. Si tratta di un lavoro certosino che non si adatta a essere strillato sulle prime pagine dei giornali e che lascia ampi margini alle critiche di chi chiede interval-



**Figura 1:** Mappa dell'Oceano Artico e delle sue principali correnti. Tratta da <http://www.divediscover.who.edu/arctic/circulation.html>

li di fiducia più stretti e maggiori evidenze sui rapporti tra cause ed effetti.

Eppure esistono ottime ragioni per sostenere che il clima stia cambiando, e che lo stia facendo per cause *antropogeniche* (cioè legate alle attività degli esseri umani). Fra i segnali che meno lasciano adito a dubbi sulla realtà del cambiamento ce ne sono alcuni che sono molto evidenti, anche se non molto pubblicizzati, in quanto non avvengono a ridosso delle grandi città, ma in un luogo remoto: l'Oceano Artico.

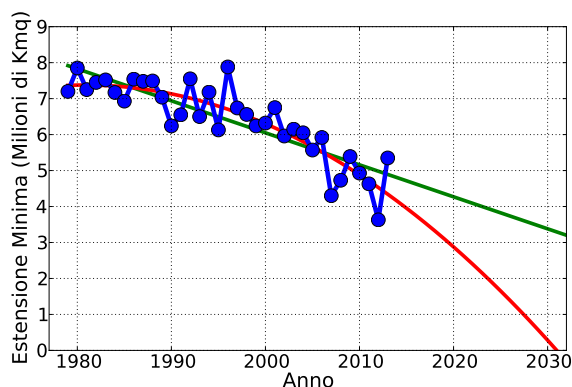
## Il declino dei ghiacci artici

I due poli del nostro pianeta sono geograficamente assai diversi. Il Polo Sud è occupato da un continente (l'Antartide) sulle cui terre si stende una sterminata calotta glaciale. Solo una piccola parte di essa si estende oltre la linea di costa e galleggia sull'oceano, rimanendo precariamente agganciata alla terra ferma. Il Polo Nord è occupato da un oceano (l'Oceano Artico), chiuso a tenaglia fra il Nord America e l'Eurasia. Da un lato esso è collegato con il Pacifico tramite lo Stretto di Bering, dall'altro la Groenlandia e le isole Svalbard delimitano lo stretto di Fram, che collega l'Artico con l'Atlantico (Figura 1).

Gran parte della superficie dell'Artico è coperta da ghiacci marini, la *banchisa*. Non si tratta,

nel suo complesso, di un singolo blocco rigido che si estende per migliaia di chilometri, bensì di un coacervo di innumerevoli lastroni di ghiaccio che galleggiano a contatto gli uni con gli altri, trasportati da venti e correnti, e che in continuazione si scontrano, si frammentano, si risaldano. La copertura non è mai perfetta: anche durante il più freddo degli inverni, i ghiacci occasionalmente si separano fino a lasciare temporaneamente libere superfici di acqua liquida.

I forti venti artici muovono grandi masse d'acqua, che si auto-organizzano in complessi sistemi di correnti, e trasportano la banchisa che fluttua su di esse. La principale è la ricircolazione di Beaufort (*Beaufort gyre*): una corrente che si richiude ad anello su sé stessa, localizzata fra le coste del Canada e quelle della Siberia orientale. I maggiori scambi d'acqua avvengono tra l'Oceano Artico e l'Oceano Atlantico. Le ultime propaggini della Corrente Nord Atlantica entrano nell'Artico scorrendo lungo la Scandinavia e la Siberia occidentale, mentre la Corrente Transpolare si forma non lontano dal Polo Nord dall'unione di più correnti minori e si getta nell'Atlantico scorrendo lungo la Groenlandia. Questa corrente trasporta verso sud anche grandi quantità di ghiaccio, che si sciolgono prontamente appena raggiungono latitudini inferiori. Ciò vuol dire che la quantità di ghiaccio presente nell'Artico



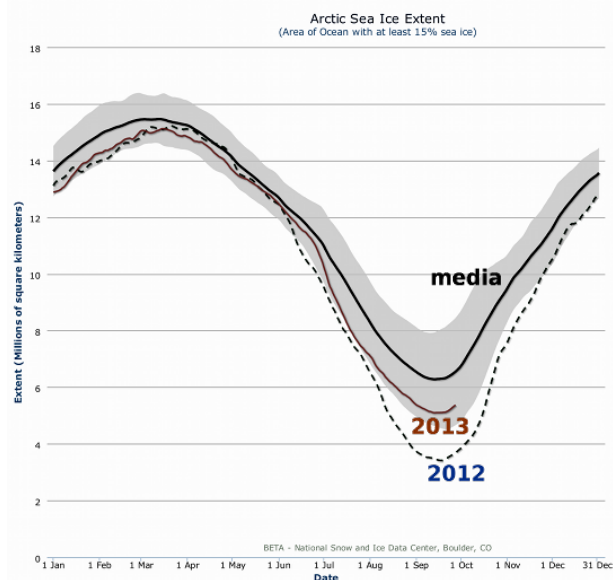
**Figura 2:** Estensione media dei ghiacci artici durante il mese di settembre (linea blu). La linea verde è una interpolazione lineare dei dati e raggiunge lo zero attorno al 2070. La linea rossa è una interpolazione quadratica. Dati tratti da [4].

dipende anche dal modo in cui i venti modulano le correnti marine che entrano ed escono da esso.

L'area coperta dai ghiacci artici è soggetta ad una notevole escursione stagionale. La massima estensione della banchisa si registra a marzo, la minima a settembre, ed è meno della metà della massima. A partire dal 1979 l'estensione dei ghiacci marini artici è stata misurata, con notevole accuratezza, sulla base di immagini satellitari. Il minimo assoluto in questa serie di dati è stato registrato nel settembre 2012, con una estensione inferiore a quattro milioni di chilometri quadrati (Figura 3). Nel 2013 il minimo non è stato così pronunciato, pur rimanendo molto al di sotto del valore medio.

La serie storica delle misure satellitari mostra che l'estensione dei ghiacci artici ha avuto una generale tendenza alla diminuzione. In particolare, i valori dell'estensione media durante il mese di settembre mostrano una tendenza accelerata al ribasso (Figura 2). Una semplice interpolazione lineare dei dati suggerisce che la banchisa estiva sparirà non oltre il settembre 2070. Una più accurata interpolazione quadratica sposta questa data a poco dopo il 2030.

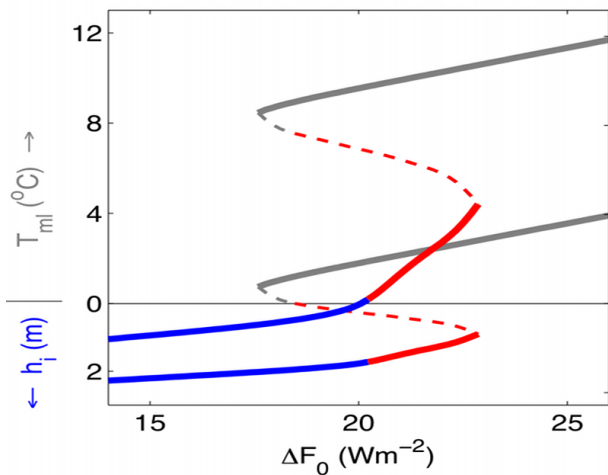
A rendere ancora più preoccupante la tendenza al ribasso c'è anche il fatto che la percentuale di ghiaccio con età maggiore di un anno è anch'essa in diminuzione. Il ghiaccio con meno di un anno (poco spesso e facilmente frammentabile) è passato dal 50% al 75% della copertura. Il ghiaccio con più di quattro anni è passato dal 25% al 2% [4].



**Figura 3:** Estensione dei ghiacci artici nel 2012 (linea tratteggiata) e fino al settembre 2013 linea continua marrone. La linea nera rappresenta l'estensione media negli anni 1981-2010. L'area grigia corrisponde a più o meno due deviazioni standard intorno alla media. Tratto da [4].

È del tutto ovvio che previsioni effettuate estrapolando in avanti nel tempo mere interpolazioni dei dati si limitano a suggerire delle tendenze, ma non hanno una base fisica. Una descrizione fisica dei ghiacci artici deve includere tre elementi: una parte termodinamica che descriva la formazione e lo scioglimento del ghiaccio sulla base della temperatura e dei flussi di calore locali; una parte fluidodinamica, che descriva il trasporto del ghiaccio da parte delle correnti e dei venti; ed infine la retroazione ghiaccio-albedo. Quest'ultimo punto è particolarmente importante. Poiché il ghiaccio è bianco, esso riflette verso l'alto buona parte della radiazione incidente: il ghiaccio ha una alta albedo. Al contrario, l'acqua liquida lascia penetrare la maggior parte della radiazione incidente, assorbendola e scaldandosi (la superficie del mare infatti è blu scura): l'acqua liquida ha una bassa albedo. Questo effetto apre la possibilità di avere due stati stabili corrispondenti alla stessa forzante termica: uno stato caldo, con acqua liquida, in cui la radiazione solare è in gran parte assorbita ed usata per scaldare l'ambiente; ed uno freddo, con estesa copertura di ghiacci, che riflette nello spazio una frazione consistente della radiazione solare.

Un semplice modello climatico illustra bene



**Figura 4:** Equilibri stabili (linea continua) ed instabili (linea tratteggiata) in un semplice modello di ghiaccio artico. La linea superiore si riferisce all'inverno, quella inferiore all'estate. Le linee blu corrispondono allo stato in cui è presente una copertura perenne di ghiacci (situazione odierna), quelle rosse allo stato in cui la copertura dei ghiacci appare solo d'inverno, quelle grigie allo stato senza mai ghiacci. L'asse verticale riporta lo spessore medio dei ghiacci (sotto lo zero) o la temperatura media dell'acqua (sopra lo zero). Il parametro  $\Delta F_0$  (asse orizzontale) è l'incremento del flusso superficiale di calore rispetto ai livelli attuali. Tratto da [5].

questa idea. In Figura 4 le linee indicano lo spessore dei ghiacci, o la temperatura dell'acqua (quando i ghiacci sono assenti) in funzione del parametro  $\Delta F_0$  che misura l'incremento del riscaldamento dell'Artico rispetto ai livelli attuali. Le due linee blu rappresentano la situazione in cui la banchisa è perenne. All'aumentare del riscaldamento si passa in un altro regime, in cui il ghiaccio marino esiste solo d'inverno (linee rosse). Questo cambio di regime è reversibile, in quanto una diminuzione del riscaldamento riporta subito il sistema allo stato con una banchisa perenne. Tuttavia, aumentando ancora il riscaldamento (quando  $\Delta F_0 \simeq 23 \text{ Wm}^{-2}$ ), il sistema salta in modo brusco nello stato del tutto privo di ghiacci rappresentato dalle linee grigie. Questa ulteriore transizione è irreversibile, nel senso che, a quel punto, una piccola diminuzione del riscaldamento non riporta il sistema nello stato con ghiacci invernali, ma lo sposta a sinistra lungo le linee grigie: la retroazione ghiaccio-albedo mantiene caldo l'Oceano Artico, e per riformare i ghiacci è necessario una diminuzione del

riscaldamento fino a  $\Delta F_0 \simeq 18 \text{ Wm}^{-2}$ , quando il sistema ripiomba nello stato di copertura glaciale perenne. L'impossibilità di recuperare lo stato precedente ad un cambio di regime semplicemente riportando il parametro di controllo al suo valore originale, è ciò che, con gergo tecnico, si chiama *isteresi*.

Modelli come quello appena discusso sono utili perché indicano quali possano essere i cambiamenti di regime e suggeriscono la possibilità di fenomeni di isteresi. Tuttavia, essi sono troppo semplici per permettere di formulare previsioni riguardo alla data in cui l'Artico potrebbe rimanere privo di ghiacci, sia perché i valori critici del riscaldamento ai quali avvengono le varie transizioni dipendono da parametri non noti con esattezza, sia perché il valore stesso del riscaldamento è una incognita che un modello completo dovrebbe poter prevedere. Per avere un quadro più realistico e dettagliato, i climatologi utilizzano modelli matematici costituiti da enormi sistemi di equazioni che descrivono il moto dell'atmosfera e dell'oceano, la dinamica dei ghiacci, il trasporto degli aerosol, ed in alcuni casi perfino l'evoluzione degli ecosistemi forestali terrestri e la biogeochimica marina. Questi *modelli del sistema Terra*, le cui equazioni sono risolte in modo approssimato da programmi di simulazione che girano su potenti supercalcolatori, hanno l'ambizione di descrivere in modo esauriente le interazioni fra tutti i processi che influenzano il clima, inclusi quelli di natura biologica.

Anche con modelli così complessi ed onnicomprensivi, per poter calcolare l'ammontare del riscaldamento è necessario specificare quali saranno le concentrazioni future dei gas serra in atmosfera. Da questo punto di vista previsioni vere e proprie non se ne possono fare: la concentrazione futura dei gas serra sarà influenzata da decisioni di tipo politico, economico e sociale, che sono soggette al libero arbitrio degli esseri umani. Pertanto, i climatologi hanno preparato molti *scenari* diversi fra loro, corrispondenti a diversi possibili andamenti della concentrazione dei gas serra, ipotizzando molteplici futuri sviluppi socio-economici. Alcuni di questi scenari sono stati codificati formalmente, e sono usati dall'*Intergovernmental Panel on Climate Change* per formulare valutazioni di rischio riguardo al clima del XXI secolo. In particolare, lo scenario

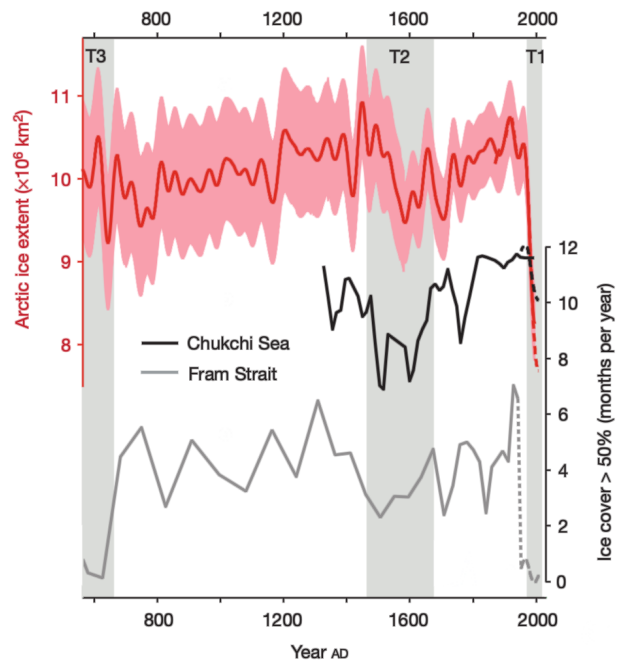
RCP8.5 prescrive un continuo aumento dei gas serra, con ritmi di crescita analoghi a quelli degli ultimi decenni, mentre lo scenario RCP4.5 immagina una crescita assai moderata dei gas serra nella prima parte del secolo, seguita da una successiva stabilizzazione su valori che, pur essendo più alti degli attuali, si spera possano non essere catastrofici.

La Figura 5 riporta l'estensione dei ghiacci marini artici a settembre ricostruita dal 1980 fino al 2005 (linea nera tratteggiata), e quella prevista dal 2006 fino al 2100 mediando fra loro i risultati di 39 diversi modelli (37 per lo scenario RCP8.5, linee tratteggiate blu e rossa). Restringsendo la media ai cinque modelli che individualmente meglio riproducono il passato, si ottengono gli andamenti mostrati dalle linee a tratto continuo, mentre le aree ombreggiate mostrano la banda entro cui fluttuano le simulazioni dei cinque modelli.

Ciò che è immediatamente evidente è l'enorme incertezza insita in queste stime. Riguardo all'estensione dei ghiacci artici, modelli diversi tendenzialmente mostrano futuri diversi (ciò non è vero per altre variabili climatologiche, per le quali c'è un ottimo accordo fra i modelli). Ancor più disarmante è che anche i modelli considerati "migliori" sottostimano fortemente sia l'estensione dei ghiacci realmente osservata (linea viola), sia il suo tasso di decrescita.

È pur vero che nell'ultimo decennio i modelli climatici hanno mostrato enormi progressi nella loro capacità di simulare una dinamica realistica dei ghiacci artici. Purtroppo, visti i risultati, è necessario riconoscere che è ancora prematuro considerare i modelli come guida affidabile riguardo ai dettagli del futuro dei ghiacci artici [7]. In effetti, se una conclusione può essere tratta, è che gli attuali modelli del sistema terra sovrastimano la data di sparizione del ghiaccio artico [8].

Per capire se ciò che si osserva è una genuina tendenza al ribasso che porterà in un futuro prossimo alla perdita completa della banchisa artica durante l'estate, oppure se si tratta di una temporanea fluttuazione insita nella naturale variabilità del sistema climatico terrestre, è necessario prendere in considerazione i dati osservativi relativi ad un passato progressivamente più lontano. Infatti, per quanto strano possa sembrare, esisto-



**Figura 6:** Ricostruzione dell'estensione del ghiaccio artico negli ultimi 1450 anni (linea rossa). Le bande rosa indicano il margine di confidenza del 95%. La risoluzione nominale è di 40 anni. Durata, in mesi all'anno, della copertura glaciale nel mare di Chukchi (linea nera) e nello stretto di Fram (linea grigia) ricostruite da carote di sedimento marino. Le linee tratteggiate indicano osservazioni dirette.

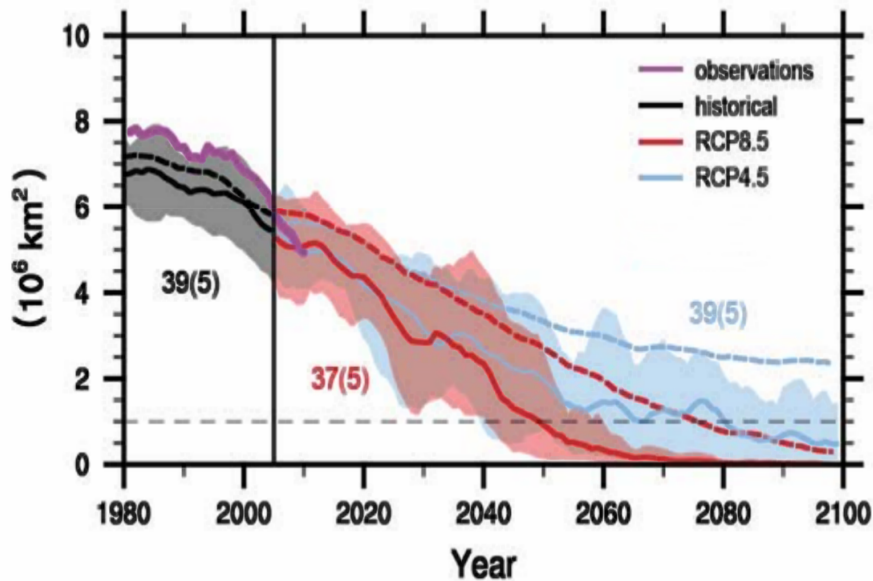
no molti indicatori indiretti, ma misurabili accuratamente, che possono fornirci indizi affidabili riguardo allo stato dell'Artico nel passato.

## L'Artico nel passato

È opinione comune che la promozione pubblicitaria di un prodotto sia un fenomeno moderno. In realtà, già intorno all'anno 985 lo scaltro capo vichingo Erik il Rosso aveva impiegato tecniche analoghe a quelle dei moderni pubblicitari per convincere un gran numero di suoi concittadini islandesi a seguirlo in una terra che egli intendeva colonizzare. Come è scritto nelle saghe, Erik riteneva che "gli uomini sarebbero stati più prontamente persuasi a seguirlo se la nuova terra avesse avuto un nome attraente" [9]. Perciò nel linguaggio dei vichinghi, egli chiamò la colonia *Terra Verde*: Groenlandia!

Preso alla lettera, Erik non mentiva: oggi come allora il racconto di fiordi costieri coperti di morbidi prati verde smeraldo che digradano verso il mare blu cobalto corrisponde alla realtà. Ciò che

## September sea-ice extent



**Figura 5:** Estensione dei ghiacci marini artici durante il mese di settembre simulata da modelli del sistema terra. In blu: scenario RCP4.5; in rosso: scenario RCP8.5; in nero: ricostruzione del passato. Le linee tratteggiate sono la media fra tutti i modelli. Le linee continue sono la media fra i cinque modelli che meglio riproducono le osservazioni (linea continua viola). Le aree ombreggiate indicano il margine di variabilità dei cinque modelli. Le serie di dati sono state smussate con una media mobile quinquennale. Tratto da [6].

Erik aveva ommesso, da bravo pubblicitario, era di sottolineare che questo paesaggio idilliaco dura al più per un paio di mesi, nel pieno dell'estate. Durante il resto dell'anno, lungo le coste della Groenlandia, il colore dominante nel paesaggio non è certo il verde.

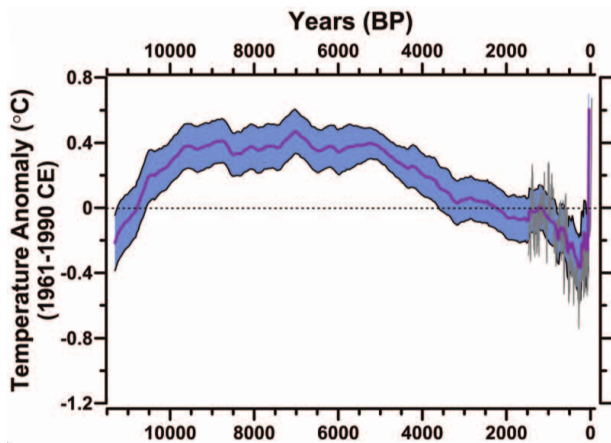
Che i vichinghi della fine del primo millennio abbiano creduto a Erik il Rosso può far sorridere. Ma che ci credano anche molte persone colte del XXI secolo è sconcertante. Capita spesso, infatti, di sentir dire che il declino osservato dei ghiacci artici non è poi così preoccupante: mille anni fa la Groenlandia era una "terra verde", e quindi sicuramente molto più calda di oggi (v. per esempio [10]).

In realtà, quando si cerca di ricostruire il passato dell'Artico usando dati oggettivi, si ottengono risultati simili a quelli della Figura 6. Lunghi cilindri di fango (chiamati, in gergo, *carote*) estratti perforando i sedimenti del fondo dell'Artico, contengono informazioni dettagliate sul clima del passato. Esaminando una carota è possibile misurare l'abbondanza degli scheletri calcarei e silicei delle cellule fitoplanctoniche vissute molti secoli fa e poi cadute sul fondo, sedimentando in strati progressivamente più antichi, a mano a mano che si segue la carota nella sua lunghezza. Poiché il fitoplancton ha serie difficoltà a sopravvivere sotto la banchisa (per il semplice motivo che il ghiaccio filtra buona parte della luce neces-

saria a questi organismi) una bassa quantità di fitoplancton segnala una banchisa pressoché perenne, grande abbondanza è indice di un luogo libero dai ghiacci, e valori intermedi permettono di calcolare per quanti mesi all'anno un certo tratto di mare era coperto dalla banchisa. La tecnica del carotaggio è usata anche sui ghiacciai perenni (presenti in Groenlandia e in molte altre isole artiche). Le carote di ghiaccio mantengono informazioni concernenti la temperatura atmosferica, l'umidità ed il livello di aerosol marini, tutti indicatori che sono legati all'estensione dei ghiacci artici. Tarando con le osservazioni moderne la procedura di analisi dei dati è possibile ricostruire con ragionevole affidabilità l'estensione dei ghiacci artici negli ultimi millecinquecento anni.

La ricostruzione di Figura 6 mostra una apprezzabile variabilità naturale. In particolare sono stati identificati tre periodi caratterizzati da estensioni della banchisa relativamente ridotte (le tre bande grigie). La colonizzazione della Groenlandia da parte dei vichinghi effettivamente avvenne poco dopo la fine del primo di questi periodi miti, e la loro dipartita (poco dopo il 1450) coincide con la massima espansione dei ghiacci. Ma è evidentissimo che queste fluttuazioni (pur ben evidenti) sono piccole se confrontate con la fortissima e repentina diminuzione dell'estensione dei ghiacci osservata negli ultimi





**Figura 7:** Ricostruzione della temperatura media globale degli ultimi 11000 anni (linea viola) con bande di incertezza larghe una deviazione standard (area azzurra). La linea grigia è la ricostruzione di Mann et al. relativa agli ultimi 2000 anni.

decenni.

Andando ancora più indietro nel tempo le evidenze diventano più indirette e frammentarie. È certo che intorno ad 8000 anni fa, all'alba della civiltà umana, la temperatura media globale fosse leggermente più alta che nel XX secolo. Una recente ricostruzione, basata principalmente su carote di sedimento marino, ma contenente anche fonti di dati di origine terrestre, mostra una differenza di circa  $0.4^{\circ}\text{C}$  (Figura 7) [12]. Restrungendo l'attenzione solo all'Artico, è verosimile che le differenze di temperatura fra quell'epoca ed il XX secolo fossero più marcate: probabilmente l'Artico era più caldo di alcuni gradi. Per valutare l'estensione dei ghiacci artici così indietro nel tempo è stato trovato un metodo ingegnoso: si cercano resti di ossa di balena e pezzi di legno spiaggiati lungo le coste delle isole artiche. Quando per buona parte dell'anno la banchisa arriva a lambire la terra è quasi impossibile che le onde portino a secco la carcassa di una balena o un tronco galleggiante. Eventi del genere, invece, sono relativamente comuni quando l'acqua liquida spazza la battigia per tutto l'anno. Integrando queste evidenze indirette con il conteggio del plancton nelle carote di sedimento è stato possibile dedurre che tra 8500 e 6000 anni fa il cosiddetto *passaggio a Nord-Ovest* (la rotta che congiunge l'Atlantico al Pacifico passando a nord del Canada) era spesso libero dai ghiacci, ma l'ipotesi che l'intero Oceano Artico fosse totalmente privo di banchisa, sia pure solo d'estate,

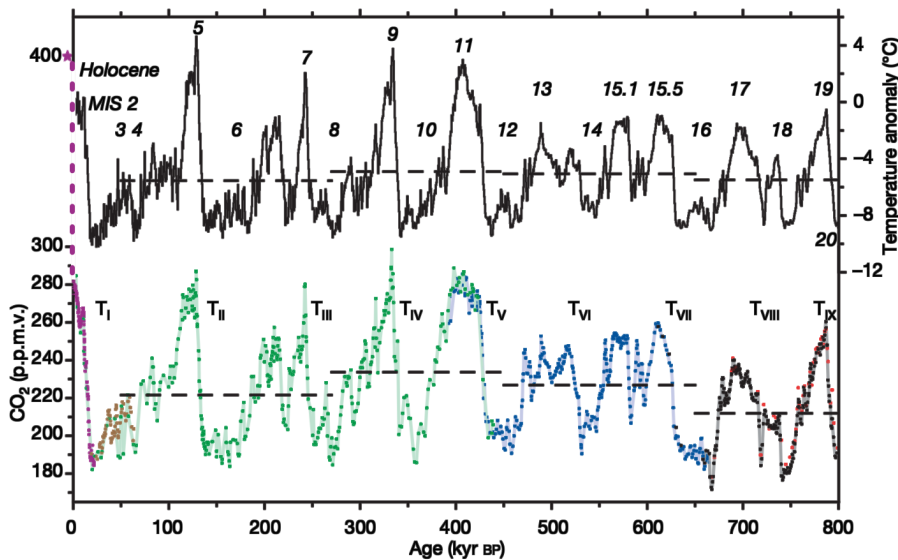
non sembra essere sostenuta dai dati [13].

In effetti, più che la differenza tra la situazione odierna ed il passato, ciò che davvero sbalordisce nella ricostruzioni delle Figure 6 e 7 è la drammatica rapidità con cui una lenta tendenza al raffreddamento arresta il suo corso e si trasforma in riscaldamento con ritmi di crescita senza precedenti.

I geologi fissano a 11700 anni fa l'inizio dell'Olocene (la più recente epoca geologica). L'evento che marca il confine con il precedente Pleistocene è il termine dell'ultima fase glaciale. A partire da circa tre milioni di anni fa il nostro pianeta si trova in una era glaciale, durante la quale si alternano delle fasi glaciali vere e proprie (come quella terminata alla fine del Pleistocene) e delle più miti fasi interglaciali (come quella che stiamo vivendo).

Uno sguardo d'insieme all'alternanza di fasi fredde e fasi più miti può essere ottenuto dalle carote di ghiaccio dell'Antartide, le uniche così profonde da permetterci di risalire con continuità così indietro nel tempo da vedere l'alternanza delle fasi glaciali. I paleoclimatologi misurano il rapporto fra l'abbondanza nelle molecole d'acqua di due isotopi stabili dell'ossigeno:  $^{16}\text{O}$  e  $^{18}\text{O}$ . Poiché le molecole contenenti  $^{16}\text{O}$  evaporano un po' più facilmente di quelle contenenti  $^{18}\text{O}$ , questo rapporto isotopico costituisce una misura indiretta della temperatura dell'ambiente in cui si è formato il ghiaccio. Inoltre, piccole bolle d'aria intrappolate nelle carote permettono di misurare la concentrazione di  $\text{CO}_2$  presente in atmosfera al momento della formazione della bolla.

In questo modo si ottengono ricostruzioni come quella mostrata in Figura 8, eseguita a partire da carote di ghiaccio estratte da un altopiano dell'Antartico chiamato *Dome C*. È evidente una fortissima correlazione fra temperatura e concentrazione di  $\text{CO}_2$ . Altrettanto evidente è l'alternanza tra fasi glaciali ed interglaciali, con cicli che durano approssimativamente centomila anni. La ricostruzione mostra che durante alcuni dei precedenti interglaciali i picchi di temperatura hanno raggiunto valori paragonabili a quelli dell'Olocene. La domanda ovvia è se una banchisa artica permanente fosse presente durante quelle fasi interglaciali. La risposta non è affatto scontata. La ricostruzione delle temperature



**Figura 8:** Differenze di temperatura rispetto alla media dell'ultimo millennio (linea superiore, scala a destra) e concentrazione della CO<sub>2</sub> atmosferica (linea inferiore, scala a sinistra) negli ultimi ottocentomila anni, ricostruite dalle carote di ghiaccio antartico del progetto EPICA. La linea tratteggiata quasi verticale mostra l'incremento di CO<sub>2</sub> dell'ultimo secolo. Le "T" indicano il termine delle fasi glaciali. I numeri indicano la successione degli stadi marini isotopici. Adattato da [14].

sulla base dei rapporti isotopici dell'ossigeno è una procedura assai delicata, soggetta ad enormi incertezze: molti altri fattori, oltre alla temperatura, influenzano quel rapporto. Ma soprattutto la ricostruzione si riferisce alle condizioni presenti nel luogo dove si è formato il ghiaccio, ovvero l'Antartide. Ben pochi dettagli possiamo dedurre sulle condizioni presenti agli antipodi, nell'Artico.

Maggiori indizi possono essere ricavati dalle evidenze geologiche sull'altezza del livello del mare. Durante una fase glaciale si accumula così tanta acqua nei ghiacciai che il livello del mare si abbassa sensibilmente. Si stima che esso sia calato di almeno 120 metri durante l'ultima fase glaciale. Analogamente, evidenze geologiche di un innalzamento del livello del mare segnalerebbero una fase con calotte glaciali più piccole di quelle odierne. Si stima che se la calotta glaciale della Groenlandia si sciogliesse il livello del mare potrebbe innalzarsi di quasi dieci metri. È bene ricordare che il livello del mare non è influenzato dallo scioglimento della banchisa, in quanto, galleggiando sull'acqua, essa già sposta un volume di liquido pari a quello che avrebbe se si sciogliesse (come è facile verificare osservando un bicchiere contenente acqua e cubetti di ghiaccio: il livello rimane invariato mentre i cubetti si sciolgono). Tuttavia, se la calotta di ghiaccio della Groenlandia dovesse sciogliersi è impensabile che nell'Oceano Artico possano mantenersi condizioni tali da consentire la formazione di una banchisa. Recenti studi [15] mostrano che durante l'interglaciale 11 (Figura 8) il livello del mare

potrebbe essere salito proprio di una decina di metri. Questo dato è soggetto a notevolissime incertezze, e potrebbe essere smentito da studi successivi. Di fatto, è l'unica evidenza che suggerisce uno scioglimento completo dei ghiacci artici in uno dei precedenti interglaciali. L'ipotesi che la copertura glaciale artica sia perdurata ininterrottamente per tutto il Pleistocene è sicuramente ancora nell'ambito del possibile.

## Che cosa ci riserva il futuro?

Il futuro è sempre nebuloso, ma chi sa leggere il passato riesce a vedere più lontano degli altri. Oscillazioni di circa cento parti per milione (fra 180 e 280 ppm) nella concentrazione atmosferica di CO<sub>2</sub> sono associate all'alternarsi delle fasi glaciali ed interglaciali. Nell'ultimo secolo, a causa delle emissioni legate al consumo di combustibili fossili ed alle altre attività umane, la concentrazione di CO<sub>2</sub> è balzata a 400 ppm (Figura 8). Ad oggi, non ostante il Protocollo di Kyoto e gli altri accordi internazionali, l'incremento dei gas serra non accenna ad arrestarsi. È evidente che ci stiamo avventurando in una terra incognita.

Il primo segnale di un regime climatico senza precedenti nell'Olocene sarà il completo scioglimento estivo della banchisa artica. È del tutto verosimile che ciò avverrà nell'arco delle nostre vite. Già oggi l'Artico è un luogo molto diverso rispetto a pochi decenni fa. In epoca sovietica il *passaggio a Nord-Est* (cioè la rotta che collega la Scandinavia allo Stretto di Bering passando lungo la costa nord siberiana) era dominio di

pochi rompighiaccio a propulsione nucleare. A cominciare dal 2002, intermittenemente, il passaggio è rimasto libero dai ghiacci alla fine dell'estate. Nel 2009 è cominciato lo sfruttamento commerciale del passaggio. In estate decine di navi mercantili, inizialmente in convoglio con un rompighiaccio in testa, poi anche da sole, risparmiano migliaia di miglia marine viaggiando dal Nord Europa all'Asia orientale attraverso l'Artico, anziché passando per l'Oceano Indiano. Inoltre, le condizioni assai meno rigide di un tempo hanno permesso l'inizio di una intensa attività di esplorazione petrolifera. La prima piattaforma di trivellazione è stata inaugurata nel 2012 (suscitando le proteste di Greenpeace). Molte altre la seguiranno.

Il ritiro della banchisa ha anche gravi conseguenze ecologiche. Quando un ambiente si trasforma con questa rapidità le reti alimentari non riescono ad adattarsi. La tundra Artica sta cambiando aspetto e comincia ad essere colonizzata da alberi e cespugli. Fioriture estive di alghe fitoplanctoniche rendono disponibili grandi quantità di cibo per lo zooplancton e per consumatori di rango superiore, ma in modo più intermittenente che in precedenza. La funzione regolatrice dei mammiferi predatori (foche, volpi e orsi polari) viene meno col declinare del loro numero [16].

Questi cambiamenti, tutti già in atto, possono ancora essere considerati locali, limitati all'Artico. Dobbiamo preoccuparci? Gli orsi polari e le foche sono animali interessanti e buffi, ma pochi di noi riterrebbero giusto cambiare profondamente il proprio stile di vita per salvarli. Ma se ciò che sta succedendo è solo l'inizio di una catena di eventi che avranno ripercussioni globali, l'intera questione deve essere vista sotto una luce diversa.

Una prima minaccia risiede nel profondo della tundra. Quando un organismo muore e viene inglobato nel permafrost, il metano prodotto dai processi di decomposizione rimane intrappolato nel terreno ghiacciato. Analogamente, il metano prodotto dai processi di decomposizione che avvengono sul fondo del mare è sequestrato in sostanze gelatinose note come *clatrati*. Riscaldando il permafrost o il fondo del mare, si produce un rilascio di metano in atmosfera. Sfortunatamente, il metano è un gas serra ancora più potente della CO<sub>2</sub>. Questo costituisce una pericolosa re-

troazione positiva sul riscaldamento dell'Artico: maggiore è il livello del riscaldamento, maggiore è la quantità di metano rilasciata in atmosfera, che, a sua volta, incrementa ulteriormente il riscaldamento. Il livello di pericolo dipende da quanto metano esattamente è immagazzinato nella tundra e nell'oceano artico. Ma questo è un dato assai difficile da stimare [17].

Sarebbe in un certo senso tranquillizzante poter appurare con certezza che durante l'interglaciale 11 la banchisa artica è davvero scomparsa e che la calotta glaciale della Groenlandia si è sciolta. Infatti, questo evento non ha impedito poi al pianeta di ripiombare nella successiva fase glaciale, e quindi indicherebbe una relativamente facile reversibilità degli effetti di un eventuale rilascio di metano. D'altra parte, le carote di ghiaccio non mostrano bruschi aumenti delle concentrazioni di metano all'apice dell'interglaciale 11, probabilmente perché un processo di riscaldamento diluito sull'arco di diecimila anni non può aver prodotto concentrazioni di metano in atmosfera altrettanto alte di quelle che potrebbe produrre un riscaldamento di pari intensità che avvenga in uno o due secoli. I pezzi di questo rompicapo ancora non collimano bene insieme.

Una seconda minaccia è dovuta all'effetto dell'acqua dolce che si riverserebbe nell'Oceano Atlantico in seguito ad un massiccio scioglimento dei ghiacci. L'Atlantico si distingue nettamente dal Pacifico in quanto al suo estremo settentrionale l'acqua fredda e salata diviene così densa da affondare e produrre masse d'acqua profonde, che poi si spostano verso sud, in un viaggio che dura molti secoli. Questa *circolazione meridionale* potrebbe interrompersi in tempi relativamente rapidi (qualche decennio) se un improvviso afflusso di acqua dolce dovesse diluire troppo la salinità del Nord Atlantico. Il collasso della circolazione meridionale produrrebbe una diminuzione del trasporto di calore dall'equatore ai poli operato dall'Atlantico, ed un severa riorganizzazione della sua circolazione più superficiale, in particolare la corrente del Nord Atlantico, che è il prolungamento della famosa corrente del Golfo che riscalda le isole britanniche. Cambiamenti di questa portata modificherebbero il clima dell'Europa in modo drastico, e renderebbero i luoghi in cui viviamo assai diversi da come sono oggi. Per quanto ipotetica, la possibilità di un collasso

della circolazione meridionale dell'Atlantico è così verosimile che la ricerca di segnali precursori è un attivo campo di studio [18].

La terza minaccia è legata all'innalzamento del livello dei mari. Se si dovesse arrivare allo stadio in cui i ghiacci artici fossero assenti anche d'estate, l'Artico salterebbe verso uno stato molto più caldo dell'attuale (Figura 4). Lo scioglimento della calotta glaciale della Groenlandia, che già oggi perde più ghiaccio di quanto non se ne accumuli con le nevicate, accelererebbe fino a completarsi nel giro di qualche secolo. Quand'anche si scoprisse che questo nuovo stato è reversibile, il conseguente innalzamento dei mari avrebbe comunque cambiato la geografia del nostro pianeta.

In conclusione, se fra qualche decennio in estate la banchisa artica si sarà sciolta completamente, non si tratterà di una curiosità oceanografica. Sarà il segnale che di lì a poco un nuovo mondo potrebbe sostituirsi al vecchio.



- [1] Disponibile in rete alla pagina <http://www.nytimes.com/2013/08/24/opinion/the-next-hurricane-and-the-next.html>
- [2] Dati disponibili sul sito <http://www.hurricanes.gov>
- [3] Disponibile in rete alla pagina <http://www.chron.com/opinion/outlook/article/Neeley-Lack-of-hurricanes-helps-climate-change-4803578.php>
- [4] National Snow & Ice Data Center: <http://nsidc.org/arcticseaicenews/>
- [5] I. EISENMAN, J. S. WETTLAUER: "Nonlinear threshold behavior during the loss of Arctic sea ice", *Proceedings of the National Academy of Sciences* **106** (2009) 28-32.
- [6] Bozza definitiva del quinto rapporto di valutazione (AR5) dell'IPCC *The Physical Science Basis*. Disponibile in rete all'indirizzo <http://www.ipcc.ch/report/ar5/wg1/#.U1FxpN8nHiM>
- [7] W. MASLOWSKI, J. C. KINNEY, M. HIGGINS, A. ROBERTS: "The Future of Arctic Sea Ice", *Annual Review Earth Planetary Sciences* **40** (2012) 625-654.
- [8] J. E. OVERLAND, M. WANG: "When will the summer Arctic be nearly sea ice free?", *Geophysical Research Letters* **40** (2013) 2097-2101.

- [9] La saga di Erik il Rosso è disponibile all'indirizzo <http://www.americanjourneys.org/aj-056/index.asp>. Una buona introduzione alla storia della colonia vichinga in Groenlandia è: JARED DIAMOND: *Collapse: How Societies Choose to Fail or Succeed*, Viking Press (2005). Tr. it.: *Collasso. Come le società scelgono di morire o vivere*, Einaudi (2005).
- [10] <http://www.weeklystandard.com/Content/Protected/Articles/000/000/000/427zgwb.asp>
- [11] C. KINNARD, C. M. ZDANOWICZ, D. A. FISHER, E. ISAKSSON, A. DE VERNAL, L. G. THOMPSON: "Reconstructed changes in Arctic sea ice over the past 1,450 years", *Nature* **479** (2011) 509-513.
- [12] S. A. MARCOTT, J. D. SHAKUN, P. U. CLARK, A. C. MIX: "A Reconstruction of Regional and Global Temperature for the Past 11,300 Years", *Science* **339** (2013) 1198-1201.
- [13] L. POLYAK ET AL.: "History of sea ice in the Arctic", *Quaternary Science Reviews* **29** (2010) 1757-1778.  
S. FUNDER ET AL.: "A 10,000-Year Record of Arctic Ocean Sea-Ice Variability—View from the Beach", *Science* **333** (2011) 747-750.
- [14] D. LÜTHI ET AL.: "High-resolution carbon dioxide concentration record 650,000–800,000 years before present", *Nature* **453** (2008) 379-382.
- [15] M. E. RAYMO, J. X. MITROVICA: "Collapse of polar ice sheets during the stage 11 interglacial", *Nature* **483** (2012) 453-456.
- [16] E. POST ET AL.: "Ecological Consequences of Sea-Ice Decline", *Science* **341** (2013) 519-524.
- [17] F. M. O'CONNOR ET AL.: "Possible role of wetlands, permafrost, and methane hydrates in the methane cycle under future climate change: a review", *Reviews of Geophysics* **48** (2010) RG4005.
- [18] T. M. LENTON: "Early warning of climate tipping points", *Nature Climate Change* **1** (2011) 201-209.



**Francesco Paparella:** Laureato in Fisica presso l'Università di Torino, ha conseguito un dottorato in Geofisica presso l'Università di Genova. È attualmente ricercatore presso il Dipartimento di Matematica e Fisica dell'Università del Salento dove insegna Istituzioni di Fisica Matematica. Si occupa di meccanica dei fluidi (applicata alla geofisica) e di sistemi dinamici (applicati alla meccanica dei fluidi ed all'ecologia).

# L'equazione di Eulero–Lagrange

*La teoria attrae la pratica come il magnete attrae il ferro.*

— Carl Friedrich Gauss

**Rocco Chirivì**

Dipartimento di Matematica e Fisica “Ennio De Giorgi” - Università del Salento

---

**L'**equazione di Eulero–Lagrange è uno dei primi risultati del calcolo delle variazioni classico. In questa lezione la useremo per determinare l'espressione analitica della brachistocrona, della catenaria e per studiare le geodetiche di semplici superfici.

Per illustrare di cosa si occupi la branca della matematica detta “calcolo delle variazioni” vediamo per prima cosa tre fra i problemi che hanno portato alla sua nascita. Si tratta di problemi di ambito fisico, ma mentre i primi due sono stati studiati nel periodo classico del calcolo delle variazioni, l'ultimo è alquanto più complesso ed è tipico della teoria del ventesimo secolo.

1. Qual è la curva che unisce due punti fissati, lungo cui una massa cade in tempo minimo?
2. Quale forma assume una corda sospesa ai suoi due estremi?
3. Se si immerge un filo di ferro, piegato lungo una certa curva chiusa, in acqua saponata si forma una superficie che ha il filo di ferro per bordo; come si descrive questa superficie?

Vedremo in seguito che in tutti e tre i casi si tratta della ricerca di un certo elemento, sia esso una curva o una superficie, che rende minimo un qualche valore. In questi termini il problema è simile al calcolo del minimo di una funzione su un intervallo. La differenza è che il valore da minimizzare non dipende da una variabile numerica ma da un oggetto più complesso: una curva o una superficie. Questa differenza è essenziale e rende il calcolo delle variazioni così interessante e difficile.

## La brachistocrona

Il primo dei problemi sopra enunciati è detto della *brachistocrona* ed è stato posto come sfida alla comunità dei matematici, cosa usuale al tempo, da Johann Bernoulli nella rivista “Acta Eruditorum” del 1696. Esso fu risolto da Newton, Leibniz, L'Hospital, dallo stesso Johann Bernoulli e dal fratello Jakob Bernoulli; queste soluzioni apparvero in un numero della stessa “Acta Eruditorum” nell'anno successivo. Una curiosità: la soluzione di Newton arrivò anonima, cosa giustificata dalla contesa in corso tra i matematici del tempo per la paternità del calcolo infinitesimale;

Johann Bernoulli parteggiava infatti per Leibniz e Newton non voleva rispondergli direttamente.

La soluzione di Johann Bernoulli si basava sulla legge di Snell per la rifrazione e il principio di Fermat secondo cui la luce viaggia lungo il cammino più breve tra due punti. Quella di Jakob Bernoulli, invece, era più geometrica e rappresentò un primo approccio alla soluzione sistematica di questo tipo di problemi.

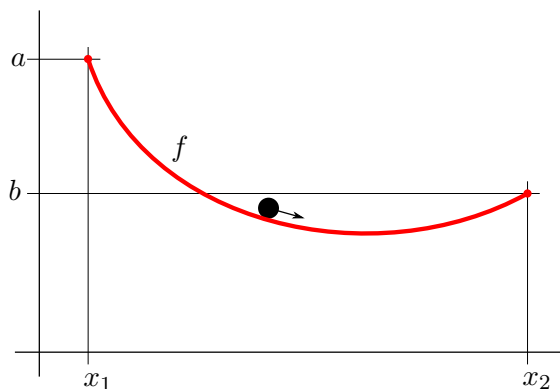


Figura 1: Una brachistocrona.

Supponendo che i punti di arrivo e partenza abbiano coordinate  $(x_1, a)$  e  $(x_2, b)$  rispettivamente e che la curva sia rappresentabile come il grafico di una funzione derivabile  $f : [x_1, x_2] \rightarrow \mathbb{R}$ , il tempo impiegato a percorrere la curva sotto la spinta della gravità è

$$J(f) \doteq \frac{1}{\sqrt{2g}} \int_{x_1}^{x_2} \left( \frac{1 + f'(x)^2}{a - f(x)} \right)^{1/2} dx$$

dove  $g$  è l'accelerazione di gravità. Per ricavare questa espressione osserviamo che in un generico intervallo infinitesimale  $[x, x + dx]$ , la curva è lunga  $(1 + f'(x)^2)^{1/2} dx$  mentre la velocità a cui è percorsa è  $(2g(a - f(x)))^{1/2}$  come si trova subito dalla conservazione dell'energia.

Quindi, se  $\Omega$  è l'insieme delle funzioni derivabili su  $[x_1, x_2]$  con  $f(x_1) = a$  e  $f(x_2) = b$ , abbiamo il funzionale su  $\Omega$

$$\Omega \ni f \xrightarrow{J} J(f) \in \mathbb{R}$$

e il problema della brachistocrona è equivalente a cercare un minimo per tale funzionale.

Si noti però che, sebbene dal punto di vista fisico si possa intuire che esiste una traiettoria che risolve il problema posto, non è assolutamente chiaro come mai il funzionale ammetta un minimo. E, inoltre, la scelta della classe delle funzioni

derivabili come insieme in cui cercare il minimo potrebbe non essere quella naturale per il funzionale  $J$ . Infatti tale funzionale potrebbe non avere minimo in questa classe ed averne in una classe più ampia, di funzioni meno regolari. D'altro canto, in una classe più ampia  $J$  potrebbe avere più minimi e questi potrebbero non essere accettabili fisicamente. Torneremo su questo punto ancora in seguito.

## La catenaria

Il secondo dei problemi enunciati è detto della *catenaria*. Il primo ad occuparsene fu Galileo Galilei che diede una soluzione sbagliata, egli pensava che la catenaria fosse un arco di conica. Il problema venne poi risolto in maniera corretta da Huygens, Leibniz e dai fratelli Bernoulli.

Anche questo problema è riconducibile alla ricerca di un minimo. Supponiamo che la corda sia appesa nei punti  $(x_1, a)$  e  $(x_2, b)$  e che la sua forma sia rappresentabile come il grafico di una funzione derivabile  $f : [x_1, x_2] \rightarrow \mathbb{R}$ . All'equilibrio l'energia della corda sarà solo energia potenziale e dovrà essere minima. Tale energia è esprimibile, a meno di costanti, come

$$J(f) \doteq \int_{x_1}^{x_2} f(x)(1 + f'(x)^2)^{1/2} dx$$

Infatti, analogamente al caso precedente, abbiamo che in un intervallo infinitesimale  $[x, x + dx]$  la curva ha una massa proporzionale alla sua lunghezza  $(1 + f'(x)^2)^{1/2} dx$  ed un'altezza data da  $f(x)$ .

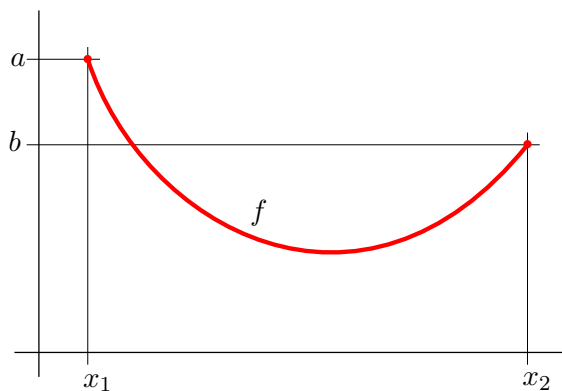
Osserviamo però che, a differenza del problema precedente, la lunghezza della corda, cioè

$$L(f) \doteq \int_{x_1}^{x_2} (1 + f'(x)^2)^{1/2} dx$$

è fissata, diciamo  $L(f) = \ell$ ; abbiamo quindi un vincolo che deve essere soddisfatto da  $f$ .

La catenaria è allora caratterizzata come la funzione  $f$  dell'insieme  $\Omega$  delle funzioni derivabili su  $[x_1, x_2]$  tali che  $f(x_1) = a$  e  $f(x_2) = b$  per cui  $L(f) = \ell$  e che rende minimo  $J$ .

Come per la brachistocrona, non è per nulla ovvio che  $J$  ammetta un (solo) minimo; benché, anche qui, l'intuito fisico ci porta a pensare che



**Figura 2:** Una catenaria.

la corda dovrà pur disporsi in qualche modo e questa forma è necessariamente un minimo.

## Il problema di Plateau

Il terzo problema prende il nome dal fisico belga Joseph Plateau, che studiò la tensione superficiale. Un primo risultato fu ottenuto da Eulero nel 1744 che dimostrò come la superficie di rotazione di area minima tra due circonferenze parallele si ottenga da una catenaria.

In generale, data un'unione finita  $\Gamma$  di curve chiuse in  $\mathbb{R}^3$ , consideriamo la classe  $\Omega$  di tutte le superfici  $S$  in  $\mathbb{R}^3$  per cui: sia possibile definire l'area  $J(S)$ ,  $S$  abbia certe fissate condizioni di regolarità e tali che il bordo di  $S$  sia  $\Gamma$ . Possiamo allora considerare il funzionale  $\Omega \ni S \mapsto J(S) \in \mathbb{R}$  che associa ad  $S$  la sua area.

Una pellicola di acqua saponata che si formi su un filo di ferro lungo  $\Gamma$  si svilupperà secondo una superficie  $S$  che minimizza l'area  $J(S)$ . Ma, come visto con i problemi precedenti e, anzi, in modo sostanzialmente più "selvaggio", la superficie  $S$  potrà essere molto poco regolare anche se  $\Gamma$  è molto regolare.

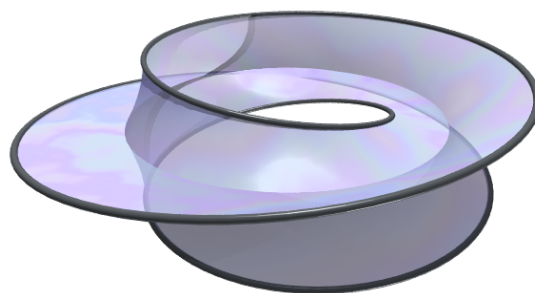
D'altra parte, come al solito, se ammettiamo superfici poco regolari in  $\Omega$  allora potrebbero esistere molti minimi per  $J$ . Ed ancora: la stessa definizione di area  $J(S)$  per una superficie poco regolare potrebbe essere problematica da definire (vedi l'articolo di Alessio Figalli in questo stesso numero di Ithaca).

Come esempio di quanto detto consideriamo la Figura 3: una superficie di area minima con bordo  $\Gamma$  regolare (diffeomorfo ad una circonferenza), ha una curva di punti singolari in cui localmente si incontrano tre superfici ad ango-

lo di  $2\pi/3$  radianti. (Si veda il sito [4] per altre interessanti illustrazioni di superfici di area minima.)

Questo esempio rende chiaro quanto la classe delle possibili superfici tra cui cercare il minimo sia problematica da definire; anzi, è questo un punto centrale per i risultati sulle superfici minime.

Come osservato per la prima volta da Meusnier nel 1785, una superficie minimizzante ha curvatura media nulla. Non è però vero il viceversa, cioè non tutte le superfici con curvatura media nulla di bordo  $\Gamma$  assegnato minimizzano l'area. Infatti la condizione di avere curvatura media è condizione necessaria per avere un minimo locale per  $J$  ma essa non è sufficiente. Tale condizione è esattamente ciò che si ottiene imponendo che la superficie considerata soddisfi l'equazione di Eulero–Lagrange argomento della prossima sezione.



**Figura 3:** Una superficie minima con punti singolari e bordo regolare.

## L'equazione di Eulero–Lagrange

In quanto segue ci poniamo un obiettivo molto modesto: vogliamo dimostrare il primo risultato classico del calcolo delle variazioni noto come equazione di Eulero–Lagrange. Il teorema che vedremo esprimerà una condizione necessaria per avere un minimo di un funzionale definito in termini integrali; tale condizione è data da un'equazione differenziale.

Sia  $[x_1, x_2] \subset \mathbb{R}$  un intervallo fissato, con  $x_1 < x_2$  e sia

$$\mathbb{R}^3 \ni (x, y, z) \mapsto F(x, y, z) \in \mathbb{R}$$

di classe  $C^2$ . Definiamo

$$J(f) \doteq \int_{x_1}^{x_2} F(x, f(x), f'(x)) dx$$

e consideriamo il funzionale

$$\Omega \ni f \mapsto J(f) \in \mathbb{R}$$

sulla classe  $\Omega$  delle funzioni  $f$  di classe  $C^1$  su  $[x_1, x_2]$  per cui  $f(x_1) = a$  e  $f(x_2) = b$ . Ci proponiamo di trovare una condizione necessaria su  $f$  per avere un minimo locale di  $J$  in  $\Omega$ .

Il primo passo è dimostrare il seguente risultato, detto *lemma fondamentale del calcolo delle variazioni*. Esso garantisce che se una funzione annulla un'ampia classe di funzionali integrali allora essa è identicamente nulla.

**Lemma 1.** *Se  $f$  è continua in  $[x_1, x_2]$  e vale*

$$\int_{x_1}^{x_2} f(x)h(x) dx = 0$$

per ogni  $h$  di classe  $C^1$  in  $[x_1, x_2]$  con  $h(x_1) = h(x_2) = 0$ , allora  $f(x) = 0$  per ogni  $x \in [x_1, x_2]$ .

*Dimostrazione.* Supponiamo per assurdo che  $f$  non sia identicamente nulla e sia  $x_0 \in [x_1, x_2]$  tale che  $f(x_0) > 0$  (il caso  $f(x_0) < 0$  è analogo, basta considerare  $-f$  al posto di  $f$ ). Essendo  $f$  continua, essa sarà positiva in un intorno di  $x_0$ . Questo ci permette di supporre che  $x_1 < x_0 < x_2$  senza perdita di generalità; inoltre, per lo stesso motivo, esistono  $\delta, \epsilon > 0$  per cui  $f(x) > \epsilon$  per ogni  $x \in (x_0 - \delta, x_0 + \delta) \subset [x_1, x_2]$ .

Consideriamo ora la funzione  $h : [x_1, x_2] \rightarrow \mathbb{R}$  definita come  $(x - x_0 + \delta)^2(x - x_0 - \delta)^2$  se  $x_0 - \delta < x < x_0 + \delta$  e 0 fuori da questo intervallo. Essa rientra nelle ipotesi del lemma ma si vede subito che

$$\int_{x_1}^{x_2} f(x)h(x) dx > \epsilon \int_{x_0 - \delta}^{x_0 + \delta} h(x) dx > 0$$

e abbiamo quindi un assurdo.  $\square$

Prima di enunciare il teorema di questa lezione vediamo una definizione: diciamo che  $f \in \Omega$  è

un *minimo locale* di un funzionale  $J : \Omega \rightarrow \mathbb{R}$  se per ogni  $h : [x_1, x_2] \rightarrow \mathbb{R}$  di classe  $C^1$  con  $h(a) = h(b) = 0$  esiste un  $\epsilon > 0$  per cui  $J(f + th) \geq J(f)$  per ogni  $0 \leq t < \epsilon$ .

**Teorema 2.** *Se la funzione  $f$  di classe  $C^2$  in  $[x_1, x_2]$  è un minimo locale per il funzionale  $J$  allora  $f$  risolve l'equazione di Eulero–Lagrange:*

$$\frac{\partial F}{\partial y}(x, f(x), f'(x)) - \frac{d}{dx} \frac{\partial F}{\partial z}(x, f(x), f'(x)) = 0$$

*Dimostrazione.* Sia  $h$  fissata di classe  $C^1$  in  $[x_1, x_2]$  con  $h(x_1) = h(x_2) = 0$ ; allora per ogni reale  $t$  la funzione  $x \mapsto f(x) + th(x)$  è in  $\Omega$ . Consideriamo la funzione

$$\mathbb{R} \ni t \mapsto J(f + th) \in \mathbb{R}$$

Essendo  $f$  un minimo locale di  $J$ ,  $\varphi$  deve avere un minimo locale in 0. Inoltre  $J(f + th)$  è un integrale su un compatto di una funzione  $C^1$ , quindi  $\varphi$  è derivabile (rispetto a  $t$ ); possiamo scambiare la derivata e l'integrale per calcolare  $\varphi'$ . Abbiamo quindi

$$\begin{aligned} 0 &= \varphi'(0) \\ &= \frac{d}{dt} \Big|_{t=0} \int_{x_1}^{x_2} F(x, f(x) + th(x), f'(x) + th'(x)) dx \\ &= \int_{x_1}^{x_2} \frac{d}{dt} \Big|_{t=0} F(x, f(x) + th(x), f'(x) + th'(x)) dx \\ &= \int_{x_1}^{x_2} \left( \frac{\partial F}{\partial y}(x, f(x), f'(x))h(x) + \frac{\partial F}{\partial z}(x, f(x), f'(x))h'(x) \right) dx \\ &= \int_{x_1}^{x_2} \left( \frac{\partial F}{\partial y}(x, f(x), f'(x)) + \frac{d}{dx} \frac{\partial F}{\partial z}(x, f(x), f'(x)) \right) h(x) dx + \left[ \frac{\partial F}{\partial z}(x, f(x), f'(x))h(x) \right]_{x_1}^{x_2} \\ &= \int_{x_1}^{x_2} \left( \frac{\partial F}{\partial y}(x, f(x), f'(x)) + \frac{d}{dx} \frac{\partial F}{\partial z}(x, f(x), f'(x)) \right) h(x) dx \end{aligned}$$

deve abbiamo usato l'integrazione per parti e osservato che le ipotesi su  $h(x)$  annullano il termine fuori dall'integrale nell'ultimo passaggio.

Usando la generalità di  $h$  abbiamo la tesi dal Lemma fondamentale.  $\square$

## Brachistocrona e cicloide

Vediamo ora come dall'equazione di Eulero–Lagrange possiamo derivare che la brachistocrona è un arco di cicloide.



Per questo problema il funzionale  $J$  è associato alla funzione

$$F(x, y, z) = \left( \frac{1 + z^2}{y - a} \right)^{1/2}$$

Possiamo supporre, a meno di una traslazione, che  $x_1 = 0$  e  $a = 0$ . Allora l'equazione di Eulero–Lagrange è, dopo varie semplificazioni,

$$f''(x) = -\frac{1 + f'(x)^2}{2f(x)}$$

Ora, una soluzione di questa equazione può essere trovata in forma parametrica come

$$\begin{cases} x &= \lambda(t - \sin t) \\ y &= \lambda(1 - \cos t) \end{cases}$$

con  $\lambda$  un numero reale fissato. È noto che questa è la parametrizzazione di una cicloide. Il valore della costante  $\lambda$  si determina imponendo la condizione che la curva passi per il punto  $(x_2 - x_1, b - a)$ .

Per verificare che abbiamo una soluzione consideriamo la funzione  $t \mapsto \lambda(t - \cos t)$ , osserviamo che  $\varphi$  è invertibile e che possiamo quindi definire  $f(x)$  come  $\lambda(1 - \cos \varphi^{-1}(x))$ .

Usando la formula per la derivata della funzione composta abbiamo

$$\begin{aligned} \frac{d}{dx} f|_{x_0=\varphi(t_0)} &= \lambda \sin t_0 \cdot \frac{d}{dx} \varphi^{-1}|_{x_0=\varphi(t_0)} \\ &= \frac{\sin t_0}{1 - \cos t_0} \end{aligned}$$

e analogamente troviamo

$$\frac{d^2}{dx^2} f|_{x_0=\varphi(t_0)} = -\frac{1}{\lambda(1 - \cos t_0)}$$

da ciò segue subito che la  $f$  così definita risolve l'equazione differenziale di Eulero–Lagrange. Per procedere osserviamo, senza dimostrazione, che

1. la soluzione trovata è l'unica soluzione dell'equazione differenziale in questione,
2. il funzionale  $J$  ammette almeno un minimo locale.

Sebbene l'equazione di Eulero–Lagrange sia solo una condizione necessaria, usando questi due fatti possiamo concludere che la soluzione data è il minimo cercato.

## Catenaria e coseno iperbolico

L'equazione di Eulero–Lagrange permette di trovare l'espressione analitica della catenaria e risolvere così il secondo problema. In realtà abbiamo bisogno di una piccola variazione (che non dimostreremo) in quanto, come abbiamo notato, è questo un problema con vincolo.

Analogamente a quanto visto per le funzioni, introduciamo un moltiplicatore di Lagrange: sia  $\Phi(f) \doteq L(f) - \ell$ , in modo che il vincolo è  $\Phi(f) = 0$ , e studiamo il funzionale senza vincoli

$$\begin{aligned} \tilde{J}(f) &\doteq J(f) - \lambda \Phi(f) \\ &= \int_{x_1}^{x_2} (f(x) - \lambda)(1 + f'(x)^2)^{1/2} dx + \lambda \ell \end{aligned}$$

dove  $\lambda$  sarà determinato nella soluzione imponendo il vincolo (oltre ad imporre  $f(x_1) = a$  e  $f(x_2) = b$ ).

Il funzionale  $\tilde{J}$  è associato alla funzione

$$F(x, y, z) \doteq (y - \lambda)(1 + z^2)^{1/2}$$

da cui ricaviamo subito che l'equazione di Eulero–Lagrange è

$$(1 + f'(x)^2)^{1/2} = \frac{d}{dx} \frac{f'(x)(f(x) - \lambda)}{(1 + f'(x)^2)^{1/2}}$$

È possibile risolvere esplicitamente questa equazione differenziale e trovare che

$$f(x) = \alpha \cosh \left( \frac{x}{\alpha} + \beta \right) + \lambda$$

è una soluzione, dove  $\alpha$ ,  $\beta$  e  $\lambda$  sono tre costanti; esse si ricavano imponendo le condizioni:  $f(x_1) = a$ ,  $f(x_2) = b$  e  $L(f) = \ell$ .

Non proviamo, come nel caso della brachistocrona, che si tratta dell'unica soluzione dell'equazione di Eulero–Lagrange e che il funzionale ammette almeno un minimo locale.

Usando ciò possiamo concludere che la soluzione trovata è il minimo cercato, cioè l'espressione analitica della catenaria.

## Geodetiche

In questa ultima sezione vogliamo far vedere come l'equazione di Eulero–Lagrange possa essere usata per determinare le geodetiche di semplici superfici. Per ragioni di spazio saremo me-

no rigorosi delle sezioni precedenti; il nostro scopo è mostrare come l'equazione di Eulero–Lagrange abbia applicazioni a problemi geometrici e non solo fisico–analitici come visto nelle sezioni precedenti.

Sia  $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}$  una funzione  $C^\infty$  e sia  $S \doteq \{x \in \mathbb{R}^3 \mid \Phi(x) = 0\}$  la superficie di  $\mathbb{R}^3$  definita da  $\Phi$ . Fissati due punti  $p, q$  di  $S$  sia  $\Omega$  lo spazio dei cammini da  $p$  a  $q$ , cioè l'insieme delle applicazioni  $\gamma$  di classe  $C^\infty$  da  $[0, 1]$  in  $S$  con  $\gamma(0) = p$  e  $\gamma(1) = q$ . Definiamo la *lunghezza* del cammino  $\gamma$  come

$$L(\gamma) \doteq \int_0^1 |\gamma'(t)| dt$$

e l'*energia* del cammino  $\gamma$  come

$$E(\gamma) \doteq \int_0^1 |\gamma'(t)|^2 dt$$

Diciamo che un cammino  $\gamma \in \Omega$  è una *geodetica* di  $S$  da  $p$  a  $q$  se è un minimo locale in  $\Omega$  del funzionale lunghezza.

Come provato in [3] (Lemma 12.1) il funzionale energia assume il suo minimo sulle geodetiche di lunghezza minima con parametro proporzionale alla lunghezza d'arco (ciò segue essenzialmente dalla disuguaglianza di Schwarz). L'energia è un funzionale preferibile alla lunghezza in quanto, di solito, porta ad equazioni differenziali più semplici.

Si noti che siamo interessati al funzionale  $E$  in  $\Omega$ , cioè consideriamo solo i cammini da  $[0, 1]$  in  $S$  e non tutte le applicazioni  $C^\infty$  da  $[0, 1]$  in  $\mathbb{R}^3$ . Dobbiamo quindi usare la versione dell'equazione di Eulero–Lagrange con vincolo come visto nel caso della catenaria.

In particolare, osserviamo che il vincolo  $\gamma(t) \in S$  per ogni  $t \in [0, 1]$  può essere espresso come

$$\int_0^1 \Phi^2(\gamma(t)) dt = 0$$

Modifichiamo quindi il funzionale energia definendo

$$\tilde{E}(\gamma) \doteq \int_0^1 (|\gamma'(t)|^2 - \lambda \Phi^2(\gamma(t))) dt$$

dove  $\lambda \in \mathbb{R}$  è una costante da determinare. È allora immediato verificare che l'equazione di

Eulero–Lagrange diventa

$$\gamma''(t) = -\frac{\lambda}{2} \nabla \Phi(\gamma(t))$$

cioè, la derivata seconda di una geodetica è parallela al gradiente della funzione che definisce la superficie, cioè è parallela alla normale alla superficie. Si noti inoltre che  $\gamma''$  risulta quindi essere parallela alla normale alla curva.

Vediamo ora alcuni esempi. Cominciamo con il caso semplicissimo delle geodetiche del piano, considerando il piano come l'ipersuperficie di  $\mathbb{R}^3$  definita da  $z = 0$ .

È chiaro che in questo caso le geodetiche sono segmenti, infatti l'equazione differenziale ora scritta diventa

$$\gamma''(t) = -\lambda(0, 0, 1)$$

e quindi  $\gamma(t) = -\lambda/2(0, 0, 1)t^2 + at + b$ . Imponendo che  $\gamma(t)$  sia nel piano definito  $z = 0$  per ogni  $t$  abbiamo  $\lambda = 0$  e imponendo che il cammino sia tra  $p$  e  $q$  abbiamo  $b = p$  e  $a = q - p$ . Abbiamo quindi ritrovato che la curva più breve, e l'unica geodetica, che congiunge  $p$  e  $q$  è il segmento.

Consideriamo ora il cilindro  $S = \{(x, y, z) \mid x^2 + y^2 = 1\}$  con asse parallelo all'asse  $z$  e raggio 1. Siano, ad esempio,  $p = (1, 0, 0)$  e  $q = (1, 0, 1)$ . L'equazione di Eulero–Lagrange in questo caso diventa

$$\begin{cases} \gamma''_x(t) = -\lambda \gamma_x(t) \\ \gamma''_y(t) = -\lambda \gamma_y(t) \\ \gamma''_z(t) = 0 \end{cases}$$

e quindi abbiamo

$$\begin{cases} \gamma_x(t) = A \cos(\omega t + \varphi) \\ \gamma_y(t) = B \sin(\omega t + \varphi) \\ \gamma_z(t) = at + b \end{cases}$$

con  $\omega = \pm\sqrt{\lambda}$  e  $A, B, \varphi \in \mathbb{R}$ . Imponendo le condizioni  $\gamma(t) \in S$ ,  $\gamma(0) = p$  e  $\gamma(1) = q$  otteniamo  $a = 1$ ,  $b = 0$ ,  $\varphi = 0$ ,  $A = B = 1$  e  $\lambda = 4k^2\pi$ ,  $k$  naturale.

Si noti come in questo caso *non* abbiamo unicità delle geodetiche: esistono infinite geodetiche date da eliche sul cilindro che partono da  $p$  e compiono più giri intorno al cilindro prima di raggiungere  $q$ . È però vero che c'è una sola geodetica di lunghezza (ed energia) minima

corrispondente a  $k = 0$ , il segmento verticale  $[0, 1] \ni t \mapsto (1, 0, t) \in S$ .

Come ultimo esempio consideriamo le geodetiche su una sfera  $S$  di raggio 1 in  $\mathbb{R}^3$  tra i punti  $p = (1, 0, 0)$  e  $q = (0, 1, 0)$ . Il vincolo è dato da  $\Phi(x, y, z) = x^2 + y^2 + z^2 - 1$  e il gradiente è quindi  $\nabla\Phi(x, y, z) = 2(x, y, z)$ ; l'equazione per una geodetica è allora

$$\gamma''(t) = -\lambda\gamma(t)$$

le cui soluzioni sono

$$\begin{cases} \gamma_x(t) = A_x \cos(\omega t + \varphi_x) \\ \gamma_y(t) = A_y \cos(\omega t + \varphi_y) \\ \gamma_z(t) = A_z \cos(\omega t + \varphi_z) \end{cases}$$

con  $\omega = \pm\sqrt{\lambda}$ . Imponendo le condizioni iniziali e finali ricaviamo  $A_x = 1, A_y = 1, A_z = 0, \varphi_x = 0, \varphi_y = \pi/2$  ( $\varphi_z$  qualsiasi) e  $\lambda = 4k^2\pi$  con  $k$  naturale non nullo.

Vediamo che anche in questo caso *non c'è* unicità per le geodetiche, ma tutte le geodetiche sono circonferenze di raggio massimo, cioè nel piano generato da  $p$  e  $q$ , che passano varie volte da  $p$  e  $q$  in entrambi i sensi di rotazione.

Ovviamente c'è solo una geodetica di lunghezza minima assoluta, cioè l'arco  $[0, 1] \ni t \mapsto (\cos \pi t/2, \sin \pi t/2, 0) \in S$  corrispondente a  $k = 1$ .



- [1] Giovanni Alberti:  
<http://www.dm.unipi.it/~alberti/files/didattica/divulgazione/calcvartreccani.pdf>
- [2] MORRIS KLINE: *Storia del pensiero matematico*, Einaudi, Torino, Volumi 1 & 2 (1999).
- [3] JOHN MILNOR: *Morse theory*, Princeton University Press (1963).
- [4] Emanuele Paolini:  
<http://web.math.unifi.it/~paolini/diletto/minime/>



**Rocco Chirivì:** Laureato in matematica all'Università di Pisa e alla Scuola Normale Superiore. Ha conseguito il dottorato di ricerca presso la Scuola Normale Superiore. È stato ricercatore in Algebra presso l'Università La Sapienza di Roma e l'Università di Pisa. Da maggio 2012 è ricercatore presso il dipartimento di Matematica e Fisica dell'Università del Salento. Si occupa di teoria delle rappresentazioni di algebre e gruppi di Lie.

Numero II Anno 2013

Viaggio nella Scienza

# Ithaca



---

Il calcolo delle variazioni

---

