Introduzione al problema dell'ottimizzazione

Marco Inversi

11 dicembre 2019

1 Introduzione

Un problema di ottimizzazione è un problema di minimo. Siano dati un insieme $\mathbb{X}$ ed una funzione $F : \mathbb{X} \to [0, +\infty)$; vogliamo studiare il problema di minimo associato ad F: F assume un valore minimo nell'insieme $\mathbb{X}$? Qual è questo valore? Quali sono i punti che lo realizzano? In caso di risposta affermativa, il valore minimo si chiama valore ottimo, i punti che realizzano tale valore si chiamano punti di ottimo (ecco perchè problema di ottimizzazione).

2 Semplici esempi

Esempio 2.1 (Parabola).
Una parabola di equazione

$$y(x) = ax^2 + bx + c,$$

con $a > 0$ raggiunge il valore minimo nel suo vertice (vedi Figura 1)

Esempio 2.2 (Funzione lineare su un poliedro convesso).
Sia data una funzione lineare

$$f(x, y) = ax + by;$$

un poliedro convesso $\mathcal{P}$ nel piano è l'intersezione di semipiani (delimitati da rette). Per cercare il minimo di f in $\mathcal{P}$ si possono guardare le linee di livello di f, cioè fissato $k \in \mathbb{R}$ vogliamo studiare l'insieme

$$A_k = \{(x, y) \in \mathbb{R}^2 \mid ax + by = k\},$$

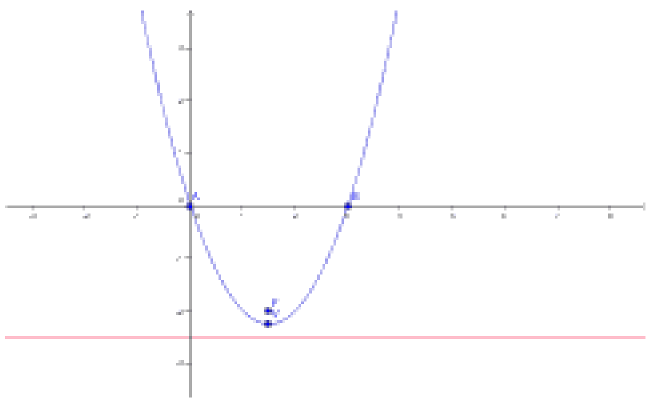

Figura 1: Parabola con la concavità verso l'alto

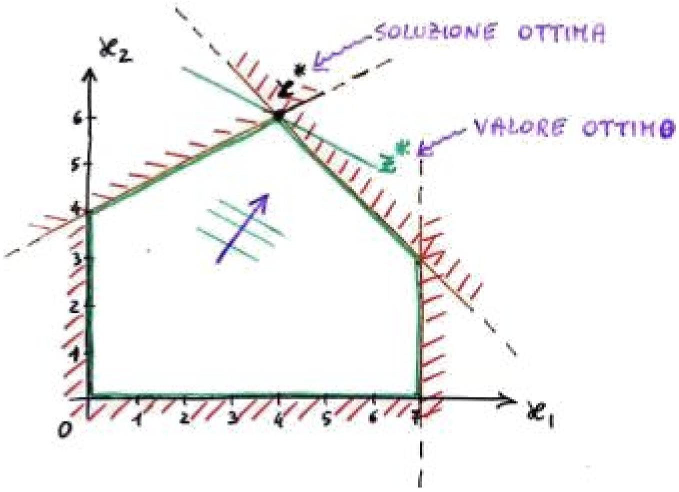

Figura 2: Poliedro convesso

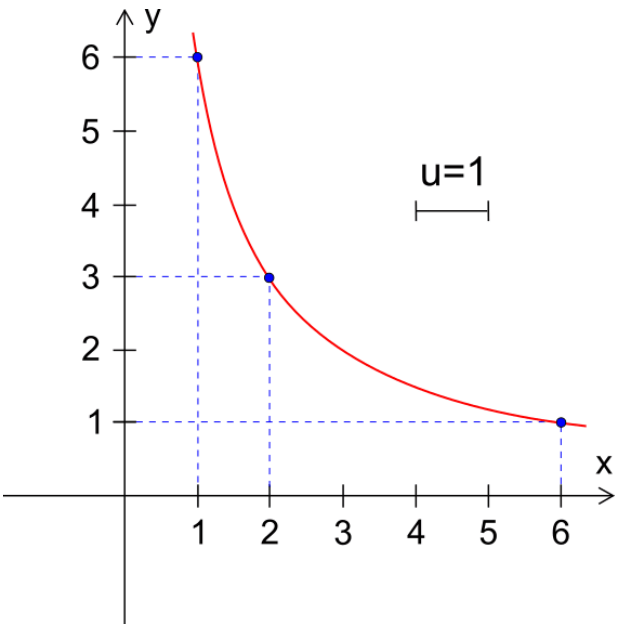

Figura 3: Iperbole

ovvero il luogo geometrico in cui f assume il valore costante k. La collezione degli insiemi $\{A_k\}_{k\in\mathbb{R}}$ è una famiglia di rette parallele che folia il poliedro $\mathcal{P}$. Allora per cercare il minimo di f su $\mathcal{P}$: basta individuare la retta tangente a $\mathcal{P}$ nella famiglia di insiemi di livello A_k che abbia il più piccolo k possibile (vedi Figura 2).

Esempio 2.3 (Iperbole).
Consideriamo l'iperbole

$$y(x) = \frac{1}{x}$$

definita sull'insieme $(0, +\infty)$. Su tale insieme l'iperbole non assume valore minimo: prende soltanto valori positivi e sempre più vicini a 0 che, purtroppo, non viene mai raggiunto. Questo è un esempio (semplice) di funzione che non ammette minimo, pur essendo limitata dal basso (vedi Figura 3).

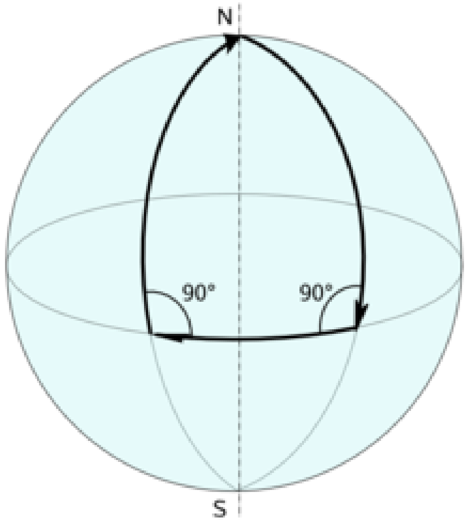

Figura 4: Triangolo geodetico

3 Una motivazione pratica

A cosa serve tutto ciò? Ci sono motivazioni valide per studiare questa classe di problemi oppure è solo un esercizio più o meno divertente di matematica?

Le applicazioni che il mondo quotidiano ci suggerisce sono molteplici. Lo sviluppo tecnologico che rende la nostra vita molto confortevole è stato possibile perchè i matematici hanno formulato e studiato problemi (come quelli di ottimizzazione) in termini del tutto astratti.

Per esempio, supponiamo di essere proprietari di un'azienda tessile. Possiamo produrre due filati: sappiamo che la produzione di una tonnellata di ciascun prodotto ha un costo e porta un certo guadagno. Tuttavia abbiamo risorse limitate (operai, macchine, materiali, tempo, ...) e vogliamo massimizzare il guadagno, cioè capire quale sia la maniera più intelligente di distribuire le risorse nella produzione dei due filati. Sotto certe assunzioni, si può tradurre questa situazione nel problema di minimizzare una funzione lineare su un poliedro convesso.

Un altro celebre esempio è il problema delle geodetiche sulla sfera. Dati due punti sulla sfera il cammino che minimizza la distanza non è una retta, ma un arco di cerchio massimo (si può vedere tendendo uno spago tra due punti su una sfera). Ecco perchè gli aerei vanno verso i poli nei viaggi trans-oceanici (vedi 4).

C'è anche un principio fisico (minima azione) secondo cui i sistemi tendono a disporsi in modo da minimizzare la loro energia. Cos'è l'energia di un sistema? Potremmo dire che è una quantità che si associa ad ogni configurazione, in modo da misurare quanto un sistema è stabile, in qualche senso. Ovviamente, questa è un'astrazione matematica e nessuno dice che la natura funzioni veramente secondo questo principio: in termini filosofici, l'ottimizzazione è un principio che gli scienziati utilizzano per ordinare il mondo. Noi formuliamo modelli che validiamo tramite le esperienze (si veda il pensiero epistemologico di Karl Popper). In questo processo, utilizziamo il linguaggio matematico per descrivere la natura, soltanto perchè questo è più comodo, preciso e prolifico di tutti gli altri che ora conosciamo. Questo è uno dei principali motivi di sviluppo della matematica. Non si può escludere che in un futuro più o meno lontano si utilizzi un linguaggio completamente diverso dalla matematica. Le mente umana ha una capacità creativa davvero impressionante.

4 Il problema isoperimetrico nel piano

4.1 Obiettivo

Per iniziare, dichiariamo il nostro obiettivo: tra le figure piane di area 1, ne esiste una che ha perimetro minimo? In caso affermativo, qual è la figura che realizza tale minimo?

La questione sembra interessante e abbastanza naturale: come è possibile circondare una quantità di area fissata a con la minor quantità possibile di contorno?

4.2 Breve storia del problema

Fin dall'antichità l'uomo si è posto questa domanda (certamente in termini diversi): l'ottimizzazione sembra un principio del tutto naturale secondo cui ordinare la nostra realtà. La matematica è un linguaggio efficace per formulare correttamente e risolvere questo problema.

Il primo a fornire una dimostrazione è stato Zenodoro, vissuto durante le fasi conclusive dell'Ellenismo, il massimo periodo di fioritura della civiltà greca (tutto il III secolo a.C. e la prima metà del II secolo a.C.). Le conquiste scientifiche e tecnologiche della civiltà ellenistica sono state di altissimo livello: basti pensare ai nomi di Archimede, Eratostene, Aristarco, Apollonio (l'elenco dovrebbe continuare, essendoci arrivate testimonianze indirette dell'attività di molti altri scienziati di indubbio valore, le cui opere sono però andate completamente perdute).

L'opera più importante di Zenodoro è il trattato perduto *Sulle figure isoperimetriche*, contenente in particolare la dimostrazione delle proprietà isoperimetriche dei poligoni regolari e del cerchio. Precisiamo che quella di Zenodoro è la prima opera scientifica dedicata al problema isoperimetrico, della quale si abbiano notizie storiche certe. Precedente a tale opera è solo la tradizione leggendaria legata alla figura della regina Didone. Virgilio scrive questi versi nel Libro I (584-589):

Giunsero in questi luoghi, ov'or vedrai
Sorger la gran cittade e l'alta ròcca
De la nuova Cartago, che dal fatto
Birsa nomossi, per l'astuta merce
Che, per fondarla, fêr di tanto sito
Quanto cerchiar di bue potesse un tergo.

Secondo la leggenda, Didone (regina di Tiro), costretta all'esilio dal fratello Pigmalione, si rifugiò presso il re Iarba nel Nord Africa, per chiedere asilo. Iarba le promise che le avrebbe dato tanto terreno quanto poteva abbracciarne una pelle di toro. Didone non si scoraggiò: tagliò la pelle in striscioline sottili e le unì in modo da formare una corda, con cui recintò lo spazio nel quale sarebbe nata Cartagine. Didone si è chiesta quale forma dare alla sua alla sua corda per abbracciare la massima area possibile.

In ogni caso, a questa leggenda non corrispondono testimonianze di trattati filosofico-scientifici sull'argomento che risultino antecedenti a quello di Zenodoro. Che al termine del III secolo a.C. sia emerso il bisogno, da parte degli scienziati dell'epoca, di trattare il problema isoperimetrico (in altri termini, che i tempi fossero maturi per questo), sembra abbastanza ragionevole se si tiene presente la popolarità che in quel periodo dovette avere l'opera di Archimede *Sulla misura del cerchio*, nella quale tra le altre cose era provata la fondamentale relazione che fornisce la doppia area di un cerchio come prodotto del perimetro per il raggio. Purtroppo nessuna delle opere di Zenodoro ci è pervenuta: conosciamo il suo lavoro per via indiretta.

Dopo secoli di silenzio quasi assoluto, l'interesse moderno peri problemi isoperimetrici riemerge contemporaneamente alla nascita del Calcolo delle Variazioni con i fratelli Bernoulli alla fine del 1600. Inoltre, la prima trattazione "geometrica" moderna (potremmo dire, nello stile degli antichi geometri) del classico problema isoperimetrico nel piano è solo del 1842, ad opera di Jacob Steiner: la prima, significativa generalizzazione delle idee geometriche contenute nel trattato di Zenodoro è stata ottenuta ben 2000 anni dopo.

La dimostrazione di Zenodoro utilizza la geometria sintetica ed è sorprendentemente precisa. I greci avevano l'idea di cosa fosse una dimostrazione rigorosa, esattamente nei termini attuali, cioè rispondente ai principi della logica aristotelica. Per essere precisi, Zenodoro dimostra che un cerchio ha perimetro minore di tutti i poligoni convessi aventi la sua stessa area.

Il problema isoperimetrico è stato risolto nella massima generalità possibile con gli strumenti (molto avanzati) dell'Analisi Matematica della seconda metà del '900 (molto difficile). Tra i tanti nomi, si ricorda quello di Ennio De Giorgi (matematico italiano attivo in tutto il '900).

4.3 Formulazione del problema

Se proviamo a formulare il problema, incontriamo difficoltà a capire cosa siano l'area e il perimetro di un insieme E contenuto nel piano. Si può rispondere a questa domanda a vari livelli:

- se E è un poligono, è intuitivamente chiaro cosa siano la sua area (la misura della parte di piano che racchiude) e cosa sia il suo perimetro (la somma delle lunghezze dei lati, che sono segmenti);

- se E è un insieme "regolare", possiamo definire la sua area allo stesso modo e il suo perimetro come la lunghezza di una corda che serve per circondarlo.

É possibile definire queste nozioni per classi più vaste di insiemi? La risposta è affermativa; occorre introdurre le definizioni giuste (misura di Lebesgue e di Hausdorff, funzioni $\mathcal{BV}$, teorema di rappresentazione di Riesz, derivate distribuzionali, insiemi di perimetro finito, ecc...). Quindi la questione è molto complicata e non possiamo affrontarla. Tuttavia, per motivi misteriosi e intrinseci nella natura delle cose, questa è l'ambientazione più "naturale" del problema.

Supponiamo di avere una nozione solida e abbastanza generale di area e perimetro di un insieme. Chiamiamo $\mathcal{C}$ la classe degli insiemi nel piano per i quali sono definiti l'area e il perimetro. In altri termini, sono definite delle funzioni

$$\mathrm{Per}\,(\cdot) : \mathcal{C} \to [0, +\infty] \quad \mathrm{Area}\,(\cdot) : \mathcal{C} \to [0, +\infty].$$

Fissiamo $M > 0$ e chiamiamo

$$\mathcal{C}_M = \{E \in \mathcal{C} \mid \mathrm{Area}\,(E) = M\}.$$

Il problema isoperimetrico corrisponde a indagare l'esistenza del minimo della funzione $\mathrm{Per}\,(\cdot)$ sull'insieme $\mathcal{C}_M$ e, poi, la forma del profilo ottimale.

4.4 Esistenza del profilo ottimale

Fissiamo $M > 0$. Per noi, che sappiamo poco o nulla di matematica, non è nemmeno chiaro che il problema isoperimetrico ammetta soluzione, cioè che esista una figura geometrica nella classe $\mathcal{C}_M$ che abbia minor perimetro a parità di area. Ricordiamoci anche che le definizioni di perimetro e area che (non) abbiamo coincidono quelle classiche soltanto nel caso di figure "ragionevoli". Quindi, vorremmo dimostrare l'esistenza di un profilo ottimale e il fatto che sia un insieme "regolare".

Nonostante l'apparente sconforto, molti intuitivamente disegnerebbero un cerchio (del raggio giusto) e avrebbero ragione: perchè? Il mondo fisico suggerisce che non può essere altrimenti: il cerchio è la figura più simmetrica, liscia e regolare che ci sia.

In seguito dimostreremo alcuni fatti; comunque, per essere coerenti, riportiamo il seguente teorema.

Teorema 4.1. *Sia $M > 0$ fissato. Esiste un unico insieme $E_0 \in \mathcal{C}$ avente $Area(E_0) = M$ e tale che E_0 realizza il minimo della funzione perimetro nell'insieme C_M, ovvero per ogni $E \in C_M$ vale che*

$$Per(E_0) \leq Per(E)$$

e l'uguaglianza vale se e solo se $E = E_0$.

4.5 Sulla forma del profilo ottimale

Una volta garantita l'esistenza del profilo ottimale, possiamo studiarne la sua forma o almeno cercare di approssimarla. L'esistenza di E_0 si prova in maniera completamente astratta. La dimostrazione non è costruttiva: non ci dà alcuna informazione su come sia fatto il profilo ottimale, ci dice soltanto che esiste (e non è poco). Con semplici argomenti di geometria piana, possiamo provare a indagare la forma di E_0.

Procediamo con delle riduzioni successive fino ad identificare la circonferenza come profilo ottimale.

4.5.1 Intuizioni euristiche

L'insieme E_0 dovrebbe essere connesso, cioè formato "da un unico pezzo". Supponiamo che sia formato da due pezzi distinti F, G. Allora vale che

$$\text{Area}\,(E) = \text{Area}\,(F) + \text{Area}\,(G) \quad \text{Per}\,(E) = \text{Per}\,(F) + \text{Per}\,(G)\,.$$

Notiamo che l'area e il perimetro di un insieme sono grandezze invarianti per traslazione, cioè non cambiano se spostiamo l'insieme con un movimento rigido. Allora, possiamo immaginare di far avvicinare F e G fino a farli toccare in un punto, oppure in un segmento, ottenendo un risparmio di perimetro a parità di area.

Tuttavia, non è detto che ciò si possa sempre fare: gli insiemi nel piano possono essere osceni e non è chiaro che si possano individuare due punti nei quali ci può essere tangenza. Tuttavia, la nostra intuizione è più che ragionevole.

L'insieme E_0 non dovrebbe avere angoli concavi e convessi, per il semplice motivo che conviene tagliarli: in questo modo si riduce di "parecchio" il perimetro. Attenzione: anche il valore dell'area risulta modificato, pur cambiando di molto poco rispetto al risparmio ottenuto sul perimetro. Allora bisognerebbe individuare un tratto liscio sul bordo di E_0, in cui aggiungere o togliere un pochino di area spingendo la frontiera verso l'esterno o verso l'interno. In questo modo il perimetro aumenta di molto poco rispetto al guadagno iniziale.

Questa operazione chirurgica risulta abbastanza difficoltosa e tecnicamente molto complicata. Anche in questo caso la nostra intuizione è del tutto ragionevole, ma abbiamo difficoltà a tradurla nel linguaggio matematico.

Viste le difficoltà che si presentano, dobbiamo cambiare approccio e procedere in maniera diversa.

4.5.2 Regolarizzazione per simmetrie

Consideriamo un retta s nel piano che divide l'insieme E_0 in due parti di uguale area. Questa retta individua due semipiani e divide l'insieme E_0 in due parti E_r (a destra) e E_l (a sinistra). Chiamiamo F_r l'insieme ottenuto simmetrizzando E_r rispetto alla retta s e F_l l'insieme ottenuto simmetrizzando E_l rispetto alla retta s. Osserviamo che

$$\mathrm{Area}\,(F_l) = \mathrm{Area}\,(F_r) = \mathrm{Area}\,(E_0)\,,$$

cioè i due insiemi rispettano il vincolo di volume. Vediamo cosa succede ai perimetri; vale la disuguaglianza

$$\mathrm{Per}\,(E_0) \geq \frac{\mathrm{Per}\,(F_r) + \mathrm{Per}\,(F_l)}{2}.$$

Allora, almeno uno tra F_l ed F_r deve avere perimetro minore o uguale rispetto a quello di E_0. Ricordiamoci di questo fatto.

Il passaggio decisivo verso la soluzione è dovuto a Steiner, che ha introdotto l'omonima simmetrizzazione. Per ogni $k \in \mathbb{R}$, l'insieme E_0 incontra la retta $y = k$ in un certo numero di segmenti. Compattiamo questi segmenti e li mettiamo in posizione centrale rispetto all'asse delle y. In altri termini, se $\varphi(k)$ è la lunghezza dell'insieme ottenuto intersecando la retta $y = k$ con E_0, il simmetrizzato di Steiner di E_0 è l'insieme

$$\tilde{E}_0 = \left\{ (x, y) \in \mathbb{R}^2 \;\middle|\; \frac{-\varphi(y)}{2} \leq x \leq \frac{\varphi(y)}{2} \right\}.$$

Osserviamo che $\tilde{E}_0$ ha la stessa area di E_0: infatti, ha senso affermare che se due insiemi hanno la proprietà che tutte le sezioni rispetto a rette parallele hanno rispettivamente la stessa lunghezza, allora gli insiemi hanno la stessa area (principio di Cavalieri). In effetti, le cose stanno esattamente in questi termini.

Si può mostrare che, se $E_0 \neq \tilde{E}_0$, allora

$$\mathrm{Per}\left(\tilde{E}_0\right) < \mathrm{Per}\,(E_0)\,.$$

La disuguaglianza è stretta e l'insieme E_0 è isoperimetrico. Allora deve valere che $\tilde{E}_0 = E_0$. In altri termini, E_0 deve coincidere con il suo simmetrizzato di Steiner. A questo punto, si riesce a guadagnare la convessità dell'insieme E_0. Inoltre è facile mostrare che è possibile ridursi al caso in cui E_0 sia simmetrico rispetto all'origine.

A questo punto siamo in grado di mostrare che E_0 è una palla. Fissiamo un angolo θ e notiamo che l'intersezione tra E_0 e la retta individuata da θ (passante per l'origine) è un segmento della forma $[-l(\theta), l(\theta)]$. Se riusciamo a provare che la funzione l è costante, allora concludiamo che E_0 è una palla. Fissiamo due angoli θ, ν tali che l'angolo tra θ e ν (che chiamiamo α) sia minore di $\frac{\pi}{2}$. Sia P il punto sul segmento $[-l(\theta), l(\theta)]$ corrispondente all'estremo $l(\theta)$. Sia s la retta individuata da ν: notiamo che divide E_0 in due parti aventi lo stesso volume; sia E_r la parte contenente il punto P. Il simmetrizzato di E_r rispetto alla retta s è isoperimetrico; in particolare è convesso. Sia Q il punto simmetrico a P rispetto alla retta s e sia $R = \frac{P+Q}{2}$ il punto medio tra P e Q. Il segmento che unisce P e Q è tutto contenuto in E_0. Quindi R è in E_0 ed è situato sulla direzione individuata da ν. Il segmento che congiunge R all'origine è più corto di $l(\nu)$. Un semplice argomento di trigonometria mostra che la lunghezza di tale segmento è $l(\theta)\cos(\alpha)$. Abbiamo provato che

$$l(\theta)\cos(\alpha) \leq l(\nu).$$

Possiamo ripetere il procedimento suddividendo α in M parti uguali, cioè introducendo gli angoli

$$\theta = \theta_0 < \theta_1 < \cdots < \theta_{M-1} < \theta_M = \nu.$$

Ovviamente l'angolo tra θ_i e θ_{i+1} è $\frac{\alpha}{M}$. Applicando in maniera iterata la disuguaglianza precedente, otteniamo che

$$l(\nu) \geq \cos\left(\frac{\alpha}{M}\right) l(\theta_{M-1}) \geq \cos\left(\frac{\alpha}{M}\right)^2 l(\theta_{M-2}) \geq \cdots \geq \cos\left(\frac{\alpha}{M}\right)^M l(\theta_\theta).$$

Siccome il numero M è arbitrario, possiamo immaginare che sia molto grande. In tal caso, si vede che la quantità

$$\cos\left(\frac{\alpha}{M}\right)^M$$

è molto vicina ad 1. In conclusione, abbiamo mostrato che

$$l(\nu) \geq l(\theta).$$

Scambiando ν e θ, otteniamo la disuguaglianza opposta, ovvero

$$l(\theta) \geq l(\nu),$$

da cui segue che $l(\theta) = l(\nu)$.