Metodi di ottimizzazione su varietà differenziabili per la media di Fréchet su disco di Poincaré

Relatore: Bruno Iannazzo.
Candidato: Luca Moroni; Mat: 311279.
Anno accademico: 2020/2021.

Introduzione

Tratteremo metodi di ottimizzazione su varietà, ovvero tecniche di ottimizzazione di funzioni non lineari definite su spazi topologici chiamati varietà differenziabili, i quali rappresentano una generalizzazione dello spazio euclideo.

Parleremo nello specifico delle tecniche di ottimizzazione utilizzate per trovare il centroide di p punti in una varietà differenziabile chiamata **disco di Poincaré**.

… Introduzione

Nello specifico applicheremo vari algoritmi di ottimizzazione per la risoluzione del problema, confronteremo i risultati ottenuti, faremo vedere anche che è possibile trasferire il problema della ricerca del centroide dalla varietà disco alla varieta conforme iperboloide confrontando, anche in tal caso, il comportamento dei vari algoritmi implementati nelle due differenti varietà.

Teoria dell'ottimizzazione euclidea

Iniziamo parlando dell'ottimizzazione di una funzione $f : R^n \rightarrow R$, ovvero cercheremo di trovare un punto di minimo di *f*.

Le condizioni necessarie per far si che un punto *x* sia di minimo locale è che $\nabla f(x) = 0$ e che $\nabla^2 f(x)$ è una matrice definita positiva.

I metodi di ottimizzazione trattati sono del tipo *iterative descent*, ovvero partono da un vettore iniziale x^0 creano una sequenza $(x^1, x^2, \ldots)$, tale per cui $f(x^{k+1}) < f(x^k)$, la quale sperabilmente converga ad un punto di minimo.

… Teoria dell’ottimizzazione euclidea

I metodi iterative descent hanno bisogno di due informazioni per poter essere eseguiti:

- **direzione di discesa**: dato g = ∇ f(x) allora ogni vettore direzione d tale che g’d < 0 è una direzione di discesa
- **selezione del passo**, .

I metodi di ottimizzazione iterativi possono essere generalizzati nella seguente:

$$x^{k+1} = x^k - \alpha^k D^k \nabla f(x^k)$$

con D^k matrice definita positiva.

A seconda della scelta di D^k e di α^k si possono definire metodi differenti.

Daremo un cenno dei metodi trattati nelle nostre sperimentazioni.

Metodi di ottimizzazione

Metodo di discesa a passo fisso: fissato α ∈ (0,1] la formula di iterazione è la seguente:

$$x^{k+1} = x^k - \alpha \nabla f(x^k)$$

Non è garantita per ogni α la convergenza.

Metodo di Armijo: la scelta di α^k garantisce una decrescita necessaria per la convergenza globale del metodo sotto opportune ipotesi, siano fissati *s*, γ e σ, con 0 < γ < 1, 0 < σ < 1 allora α^k =$\sigma^m s$, dove *m* è il naturale più piccolo per cui vale:

$$f(x^k) - f(x^k + \sigma^m s d^k) \geq -\gamma \sigma^m s \nabla f(x^k)' d^k$$

… Metodi di ottimizzazione

Metodi di quasi newton: tali metodi agiscono nella scelta della matrice D^k la quale è selezionata in modo tale che $d^k = -D^k \nabla f(x^k)$ tende ad approssimare la direzione scelta dai metodi di Newton (nei quali D^k viene scelta come la matrice hessiana) e D^k approssima la matrice hessiana. Possiamo, tramite due iterazioni successive, avere accesso ad informazioni riguardo la matrice hessiana. Definiti $p^k = x^{k+1} - x^k$ e $q^k = \nabla f(x^{k+1}) - \nabla f(x^k)$ allora D^{k+1} deve soddisfare $D^{k+1}p^k = q^k$ la quale prende il nome di ***equazione delle secanti***.

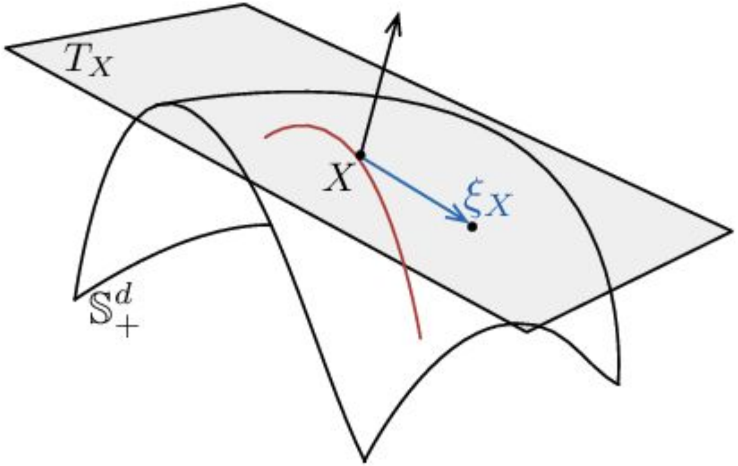
Teoria dell'ottimizzazione su varietà

In tale contesto la funzione da ottimizzare sarà $f : M \rightarrow R$, perciò il suo dominio è una varietà differenziabile *M*. Per poter effettuare ottimizzazione di una funzione definita su una varietà è necessario introdurre il concetto di retrazione.

Una **retrazione** *R* su x in M R_x è una mappa che va da T_xM in *M*, con una condizione di rigidità locale che preserva il gradiente in *x*. T_xM rappresentante lo spazio tangente ad *x* di *M*.

Un caso particolare di retrazione è la **mappa esponenziale** definita su *x* in *M* e *v* in T_xM che ha proprietà di retrazione e rappresenta la curva geodedica calcolata nel punto 1 originaria in *x* con derivata in 0 pari a *v*.

... Teoria dell'ottimizzazione su varietà



I metodi iterativi trattati nel caso euclideo hanno una controparte definita per il caso generale su varietà.

$$x^{k+1} = R_{x^k}(t^k \eta^k),$$

Dove η^k appartiene a $T_{x^k} M$ e t^k è uno scalare, i metodi analogamente al caso euclideo sono definiti in base alle scelte effettuate per la direzione η^k e la lunghezza del passo t^k, e sono analoghi al caso euclideo.

Disco di Poincaré

Andiamo ora a trattare una delle due varietà sulle quali effettueremo ottimizzazione. Il disco di Poincarè è un dei modelli più famosi per la descrizione dello spazio iperbolico ed è formalmente definito come segue.

$$\mathbb{D}^n = \{x \in \mathbb{R}^n \mid ||x|| < 1\}$$

Il disco di Poincaré è una varietà riemanniana

Lo spazio tangente è rappresentato da R^n in ogni punto.

In tale varietà sono ben definite le formule necessarie per effettuare ottimizzazione, ovvero la metrica, la distanza e la mappa esponenziale.

Iperboloide

La seconda varietà che trattiamo è l'iperboloide, è un altro modello per la rappresentazione dello spazio iperbolico.

$$\mathbb{H}^n = \{x \in \mathbb{R}^{n+1} | \langle x, x \rangle_{n:1} = -1; \quad x^{n+1} > 0\}$$

L'iperboloide è una varietà riemanniana.

Lo spazio tangente è rappresentato da,

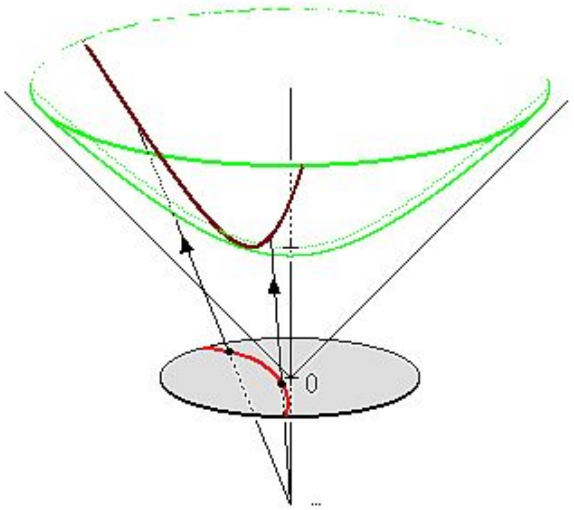
$$T_p\mathbb{H}^n \approx \{x \in \mathbb{R}^{n:1} \mid \langle p, x \rangle_{n:1} = 0\}$$

La metrica è definita dal ***prodotto di Minkowski***.

$$\langle u, v \rangle_{n:1} = \sum_{i=1}^{n} u_i v_i - u_{n+1} v_{n+1}; \quad u, v \in \mathbb{R}^{n+1}$$

Anche in questo caso tutti gli strumenti per effettuare ottimizzazione sono ben definiti, tra cui la distanza e la mappa esponenziale.

Modelli conformi



Le due varietà appena definite sono legate da un diffeomorfismo conforme (mantiene gli angoli ma non le distanze), il diffeomorfismo invertibile in questione è la mappa stereografica ρ, definita come segue.

$$\rho : \mathbb{H}^n \to \mathbb{D}^n \mid x \to \frac{1}{x_{n+1} + 1}(x_1, \dots, x_n)$$

$$\rho^{-1} : \mathbb{D}^n \to \mathbb{H}^n \mid y \to \frac{2}{1-r}(y_1, \dots, \frac{1+r}{2}) \ ; \text{con } r = ||y||^2$$

… Modelli conformi

In funzione della mappa conforme che lega le due varietà presentate verifichiamo che vale $d_{\mathbb{D}^n}(a,b) = d_{\mathbb{H}^n}(\rho^{-1}(a), \rho^{-1}(b))$, per *a* a *b* appartenenti al disco.

$$d_{\mathbb{D}^n}(a,b) = arccosh(1 + 2\frac{||a+b||^2}{(1-||a||^2)(1-||b||^2)})$$

$$d_{\mathbb{H}^n}(\rho^{-1}(a), \rho^{-1}(b)) = arccosh(-\langle \rho^{-1}(a), \rho^{-1}(b)\rangle_{n:1})$$

Facciamo vedere che vale l'uguaglianza per gli argomenti della funzione *arccosh* (arcocoseno iperbolico).

… Modelli conformi

$$-\langle \rho^{-1}(a), \rho^{-1}(b)\rangle_{n:1} = -\langle \frac{2}{1-\|a\|^2}\begin{bmatrix} a \\ \frac{1+\|a\|^2}{2}\end{bmatrix}, \frac{2}{1-\|b\|^2}\begin{bmatrix} b \\ \frac{1+\|b\|^2}{2}\end{bmatrix}\rangle_{n:1}$$

$$= \frac{4}{(1-\|a\|^2)(1-\|b\|^2)}\left(a'b - \frac{(1+\|a\|^2)(1+\|b\|^2)}{4}\right)$$

$$= \frac{-4a'b + 1 + \|a\|^2 + \|b\|^2 + \|a\|^2\|b\|^2}{(1-\|a\|^2)(1-\|b\|^2)}$$

$$= \frac{-4a'b + 1 + 2\|a\|^2 + 2\|b\|^2 - \|a\|^2 - \|b\|^2 + \|a\|^2\|b\|^2}{(1-\|a\|^2)(1-\|b\|^2)}$$

$$= 1 + 2\frac{\|a-b\|^2}{(1-\|a\|^2)(1-\|b\|^2)}$$

Media di Fréchet, problema del centroide

Dati p punti in una varietà la media di Fréchet dei punti è l'argomento che minimizza la funzione **somma delle distanze al quadrato**.

$$f(\Theta) = \frac{1}{p} \sum_{i=1}^{p} d^2(\Theta, x^{(i)})$$

In funzione dell'uguaglianza nella precedente slide, possiamo asserire che, definendo la funzione *f* su disco se esiste un valore di minimo allora esiste un valore di minimo anche per la medesima funzione $\dot{f}$ definita su iperboloide. Essendo inoltre $f(x) = \dot{f}(\rho^{-1}(x))$ allora cercare il minimo di f equivale a cercare il minimo di $\dot{f} \circ \rho^{-1}$.

In funzione di tale dimostrazione possiamo trasportare il problema del calcolo della media di Fréchet definito su disco di Poincaré nella varietà iperboloide e viceversa.

Algoritmi

Gli esperimenti che descriveremo si basano sull'utilizzo di tre differenti algoritmi, implementati nella versione riemanniana sia su disco che su iperboloide.

Gli algoritmi implementati sono:

- Algoritmo a passo fisso
- Algoritmo che si basa sul metodo di Armijo
- Algoritmo di Barzilai-Borwain: metodo di quasi-Newton.

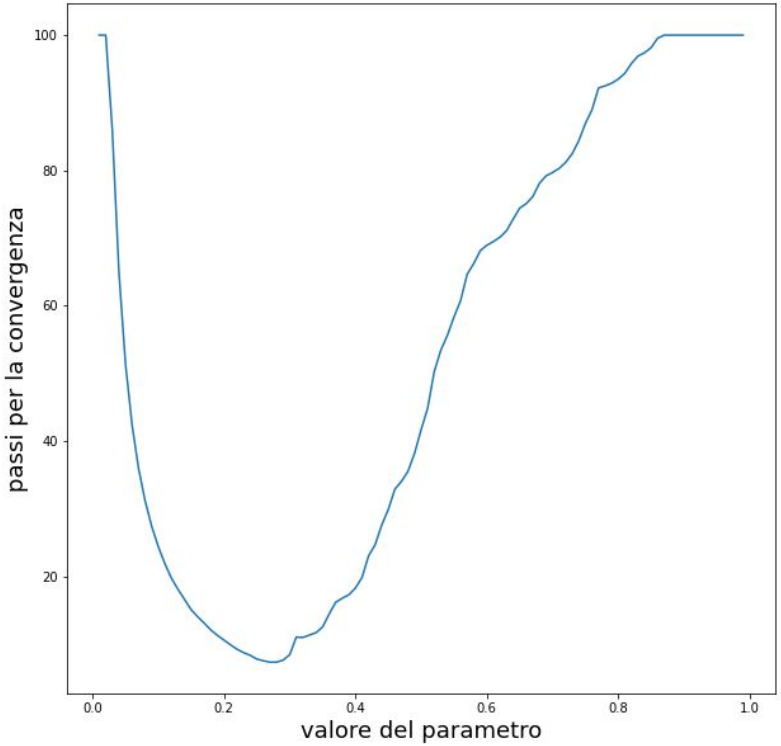
Esperimenti

Gli esperimenti effettuati hanno come scopo quello di confrontare gli algoritmi implementati, e di comprendere se le due varietà su cui sono stati implementati sono effettivamente equivalenti, non solo da un punto di vista geometrico ma anche da un punto di vista computazionale.

In una prima fase abbiamo creato un dataset contenente varie (duecento) istanze del problema della media di Fréchet su disco di Poincaré per una dimensione fissata ($n=2$) calcolandone il limite tramite l'algoritmo a passo fisso con passo molto piccolo e come guardia d'arresto l'annullamento del gradiente.

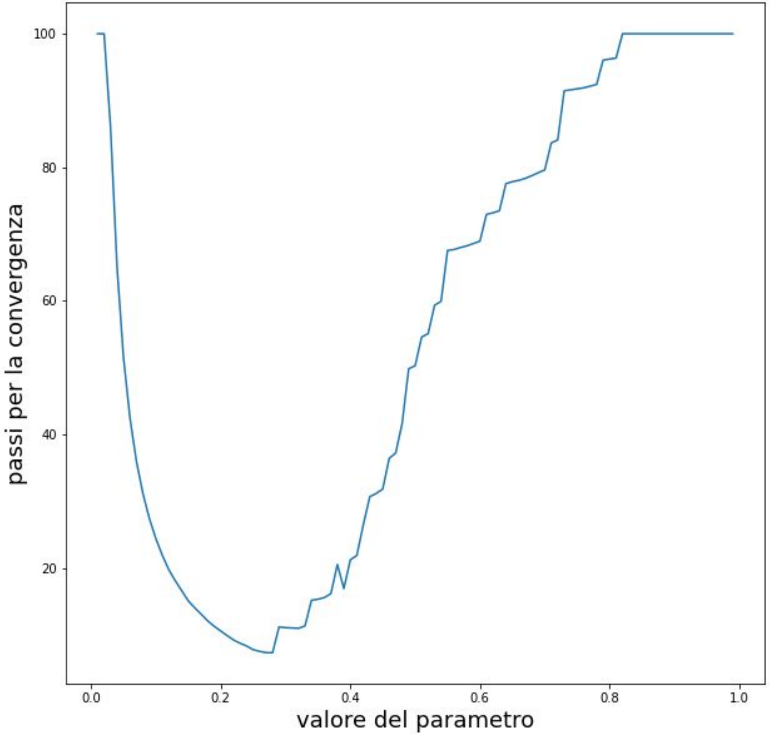
Dopo di che abbiamo utilizzato un sottoinsieme del dataset per definire i parametri "liberi" degli algoritmi, ovvero parliamo di α dell'algoritmo a passo fisso e di s per l'algoritmo di Armijo. Questo è stato fatto applicando gli algoritmi con un parametro incrementale su ogni istanza del sottodataset e per ogni esecuzione abbiamo calcolato il numero di passi per la convergenza generando per ogni istanza una sequenza ed alla fine ne abbiamo calcolato una media. Abbiamo selezionato i parametri che minimizzano le sequenze medie sia per l'implementazione su disco che su iperboloide.

… Esperimenti



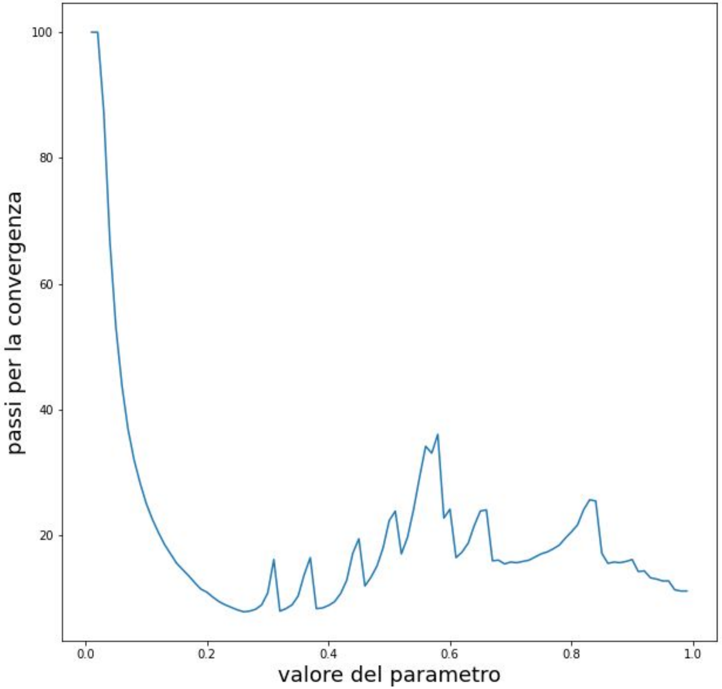
Sequenza relativa alla convergenza dell'algorimto a passo fisso implementato su disco. Punto di minimo: 0.27, valore di convergenza minimo: 7.35.

... Esperimenti



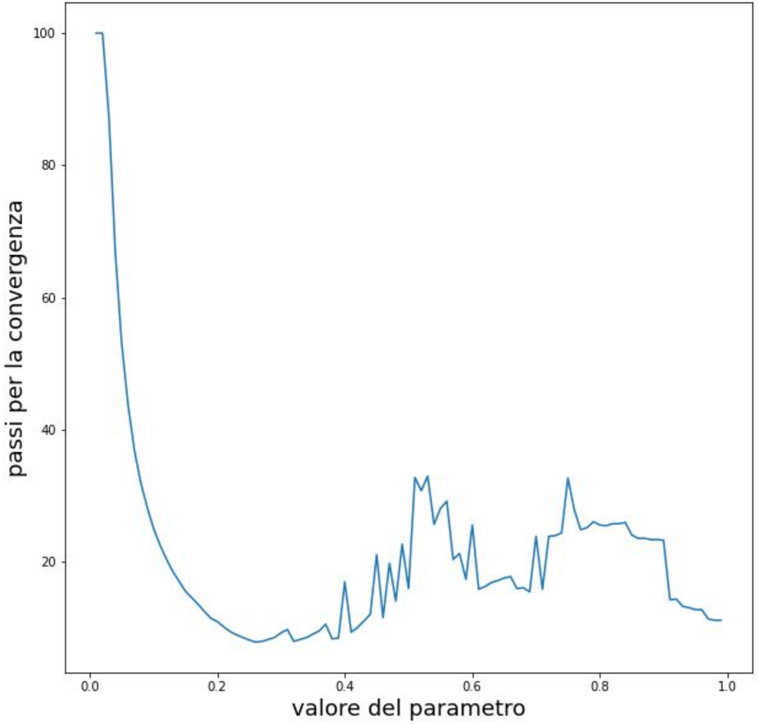
Sequenza relativa alla convergenza dell’algorimto a passo fisso implementato su iperboloide. Punto di minimo: 0.27, valore di convergenza minimo: 7.35.

… Esperimenti



Sequenza relativa alla convergenza dell’algoritmo di Armijo implementato su disco. Punto di minimo: 0.26, valore di convergenza minimo: 7.9.

... Esperimenti



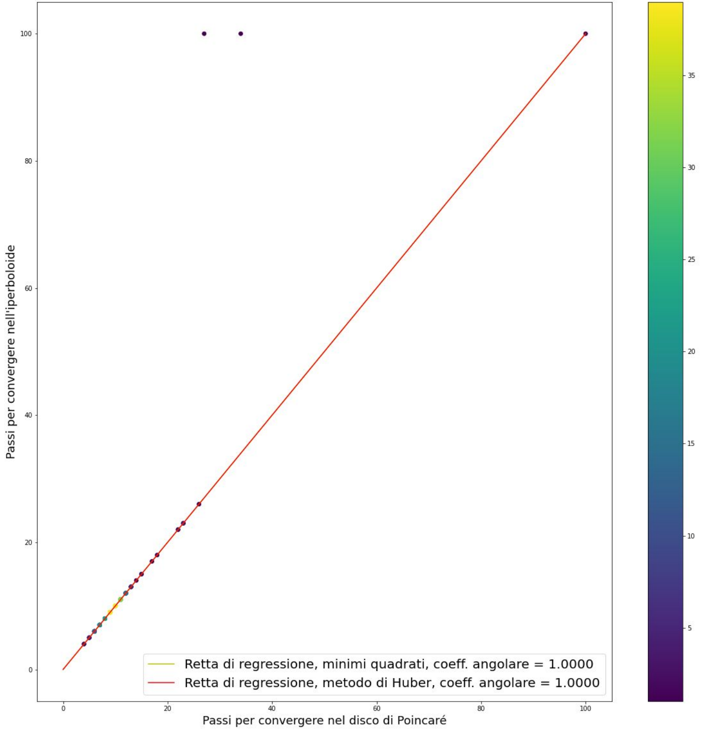
Sequenza relativa alla convergenza dell’algoritmo di Armijo implementato su iperboloide. Punto di minimo: 0.26, valore di convergenza minimo: 7.9.

… Esperimenti

Una volta selezionati tutti i parametri dei nostri algoritmi, abbiamo ripreso il dataset totale, per ogni algoritmo lo abbiamo eseguito su ogni istanza del dataset totale creando una serie di punti $(a, b) \in \mathrm{R}^2$, uno per ogni istanza creando così una nuvola di punti, con *a* indicante il numero di passi per la convergenza al limite dell'implementazione su disco e *b* il numero di passi per la convergenza al limite dell'implementazione su iperboloide.

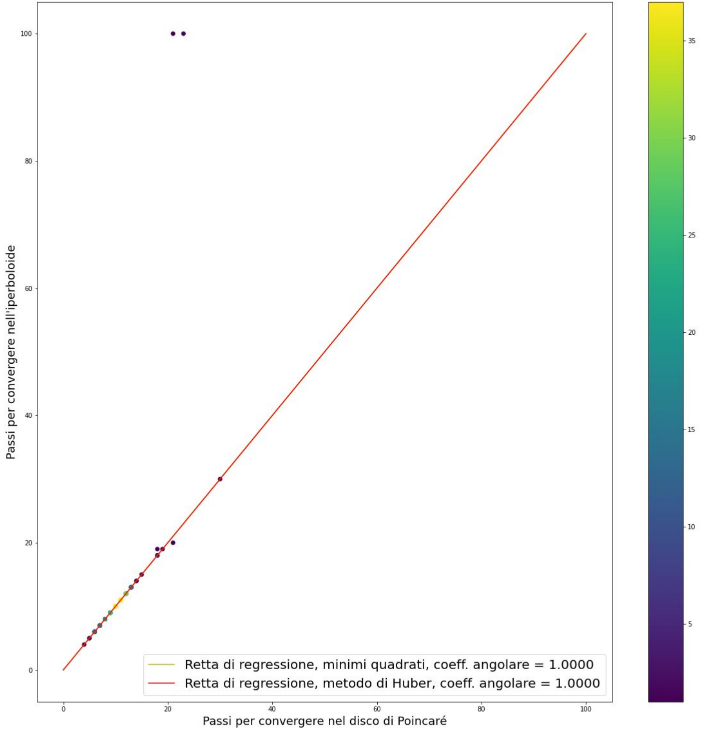
Dalla nuvola di punti si sono eliminati quei punti con una molteplicità bassa, e si sono calcolate due rette di regressione, la prima tramite il metodo dei minimi quadrati, la seconda tramite l'algoritmo di *Huber*.

… Esperimenti



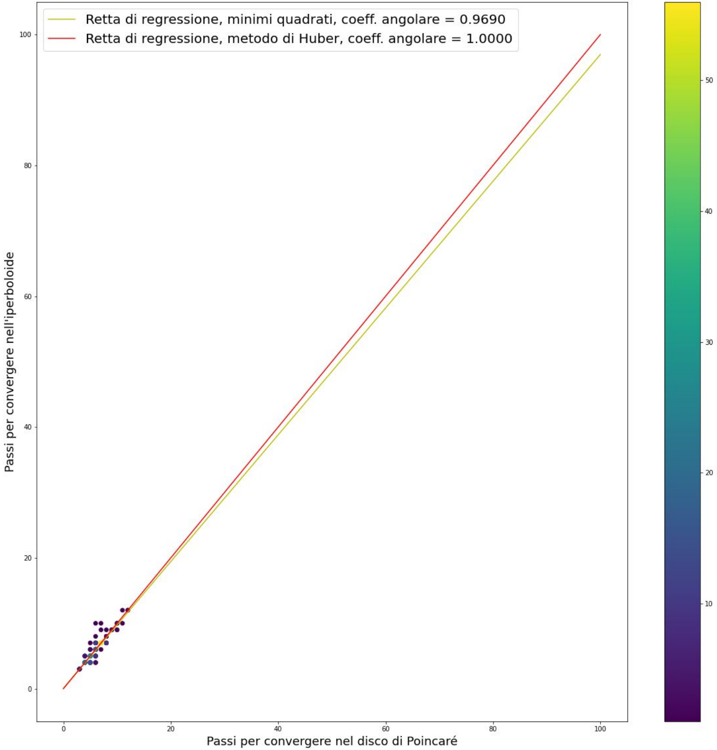
Nuvola di punti relativa all’algoritmo a passo fisso. Notiamo che i punti generati tendono a giacere sulla bisettrice e ciò indica un’equivalenza tra le due implementazioni. Valore medio di convergenza sul disco 10. 5, valore medio di convergenza su iperboloide 11.2.

… Esperimenti



Nuvola di punti relativa all'algoritmo di Armijo. Notiamo che i punti generati tendono a giacere sulla bisettrice e ciò indica un'equivalenza tra le due implementazioni. Valore medio di convergenza sul disco 10. 4, valore medio di convergenza su iperboloide 11.2.

… Esperimenti



Nuvola di punti relativa all'algoritmo di Barzilai-Borwein. Notiamo un'equivalenza tra le due implementazioni data dal coefficiente angolare tendente ad 1 delle rette di regressione. Valore medio di convergenza sul disco 6.13, valore medio di convergenza su iperboloide 6.8.

Conclusioni

I risultati di tali sperimentazioni ci portano a congetturare che dato un problema di ottimizzazione su varietà, è possibile trasferire il problema da tale varietà ad una varietà conforme tenendo presente che deve essere preservata l'esistenza di una soluzione ottima per la funzione costo.

Dai risultati è inoltre possibile notare come l'algoritmo di Barzilai-Borwain si comporta meglio rispetto agli algoritmi a passo fisso e di ricerca lineare (Armijo).

Grazie per l'attenzione.