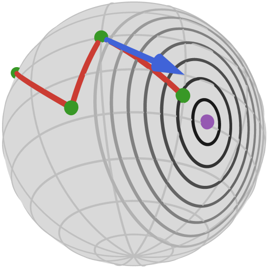
Metodi di ottimizzazione su varietà differenziabili per la media di Fréchet su disco di Poincaré

Candidato: Luca Moroni
Relatore: Bruno Iannazzo

Anno Accademico 2020/2021

Introduzione

Tratteremo metodi di ottimizzazione di funzioni definite su varietà differenziabili, oggetti geometrici che generalizzano le nozioni di curva e superficie a dimensioni maggiori.

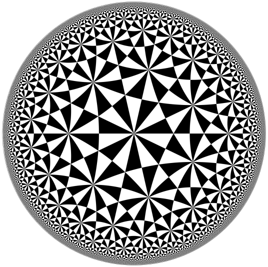


Introduzione

Problema

Calcolare il centroide di p punti nel **disco di Poincaré**.

- Radar clutter classification [Barbasesco et al., 2019]
- Hyperbolic deep neural networks [Peng et al., 2021]



Ottimizzazione euclidea

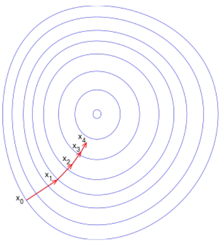
Data $f : \mathbb{R}^n \to \mathbb{R}$, con f due volte differenziabile, cercheremo di trovare un punto di minimo di f. Condizioni necessarie per dire se x è un punto di minimo:

- $\nabla f(x) = 0$
- $\nabla^2 f(x)$ definita positiva

Ottimizzazione euclidea

Metodi iterative descent.

- Si parte da $x^0 \in \mathbb{R}^n$
- Si genera una sequenza $(x^1, x^2, \ldots)$, che converga (sperabilmente) ad un punto di minimo



Ottimizzazione euclidea

I metodi iterative descent hanno bisogno di:

- **direzione di discesa**: dato $g = \nabla f(x)$ allora ogni vettore direzione d tale che $g'd < 0$ è una direzione di discesa
- **selezione del passo**: α

I metodi di ottimizzazione iterativi sono così generalizzati:

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

con D^k matrice definita positiva.

Metodo a passo fisso

Fissato $\alpha \in (0, 1]$ si ha che:

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

- Molto semplice
- Non richiede ulteriori valutazioni della funzione costo
- Non è facile determinare a priori per quali valori di α converge

Metodo di Armijo

Fissati s, γ e σ, con $0 < \gamma < 1$, $0 < \sigma < 1$ e $s > 0$, allora $\alpha^k = \sigma^m s$, dove m è il naturale più piccolo per cui si ha sufficiente decrescita:

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

- Richiede molteplici valutazioni della funzione costo
- Garantisce convergenza globale sotto blande ipotesi

Metodi quasi Newton

Famiglia di metodi che scelgono $D^k \approx \nabla^2 f(x)^{-1}$.

D^{k+1} deve soddisfare l' equazione delle secanti $D^{k+1} q^k = p^k$, dove:

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

Ottimizzazione su varietà

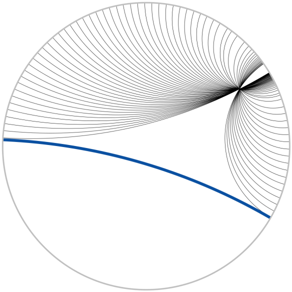
Data $f : M \to \mathbb{R}$. Dove M è una varietà differenziabile.

Una retrazione R_x su x in M è una mappa che va da T_xM in M.

T_xM rappresentante lo spazio tangente ad x di M.

Mappa esponenziale

Un caso particolare di retrazione è la mappa esponenziale.
Definita su x in M e v in T_xM rappresenta la curva geodetica calcolata nel punto 1 originaria in x con derivata in 0 pari a v.

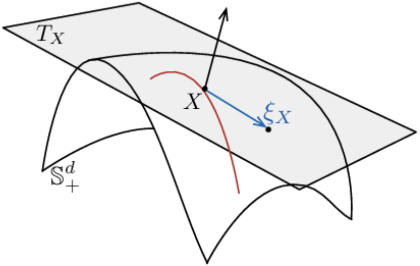


Ottimizzazione su varietà

La formulazione generale dei metodi iterative descent su varietà è,

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

Dove η^k appartiente a $T_{x^k}M$ e t^k è uno scalare.
I metodi sono analoghi al caso euclideo.



Disco di Poincaré

Definizione

Modello per la descrizione dello spazio iperbolico.

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

È una varietà riemanniana.
Lo spazio tangente è rappresentato da $\mathbb{R}^n$ in ogni punto.
La metrica nello spazio tangente rispetto a p nel disco:

$$g^{\mathbb{D}} = \lambda_p^2 g^E; \text{ con } \lambda_p = \frac{2}{1 - \|p\|^2} \text{ e con } g^E \text{ la metrica euclidea.}$$

Sono ben note le formule necessarie per effettuare ottimizzazione ovvero, la distanza e la mappa esponenziale.

Iperboloide

Definizione

È un altro modello per la rappresentazione dello spazio iperbolico.

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

È una varietà riemanniana.
Lo spazio tangente è rappresentato da,

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

La metrica è definita dal prodotto di Minkowski.

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

Sono ben note la distanza e la mappa esponenziale.

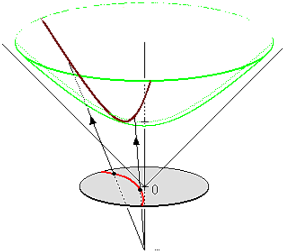
Modelli conformi

Mappa stereografica

Le due varietà appena definite sono legate da un diffeomorfismo conforme, la mappa stereografica ρ, definita come segue.

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

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



Modelli conformi

In funzione della mappa conforme che lega le due varietà presentate vale:

$$d_{\mathbb{D}^n}(a,b) = d_{\mathbb{H}^n}(\rho^{-1}(a), \rho^{-1}(b)) \quad a, b \in \mathbb{D}^n.$$

Con:

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

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

Media di Fréchet e problema del centroide

Media di Fréchet

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

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

Definendo la funzione f su disco se esiste un valore di minimo allora esiste un valore di minimo anche per la funzione $\hat{f}$ definita su iperboloide.

Media di Fréchet e problema del centroide

Essendo $f(x) = \hat{f}(\rho^{-1}(x))$ allora cercare il minimo di f equivale a cercare il minimo di $\hat{f} \circ \rho^{-1}$.

È possibile trasportare il problema del calcolo della media di Fréchet definito su disco di Poincaré nella varietà iperboloide e viceversa.

Algoritmi

Abbiamo implementato i seguenti tre algoritmi nella versione riemanniana sia su disco che su iperboloide.

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

Esperimenti

Obiettivo Sperimentale

Capiremo se le due varietà sono equivalenti, non solo da un punto di vista geometrico ma anche da un punto di vista computazionale.

Abbiamo creato un dataset contenente varie (duecento) istanze del problema della media di Fréchet su disco di Poincaré, per ogni istanza ne abbiamo calcolato il limite.

Esperimenti

Abbiamo utilizzato un sottoinsieme del dataset per definire i parametri “liberi” degli algoritmi.

- α: dell’algoritmo a passo fisso
- s: per l’algoritmo di Armijo

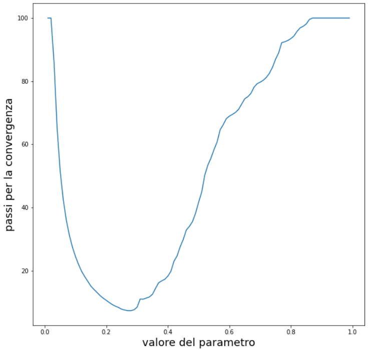
Esperimenti

Selezione parametri liberi

- Abbiamo applicato gli algoritmi con un parametro incrementale su ogni istanza del sotto-dataset
- 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

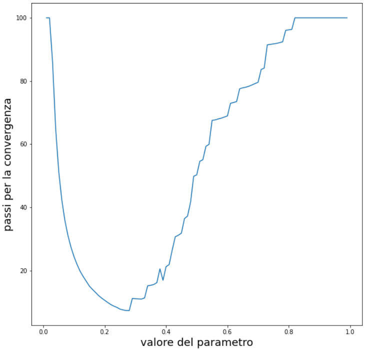
Scelta del parametro passo fisso

Disco di Poincaré



Punto di minimo: 0.27, valore di convergenza minimo: 7.35.

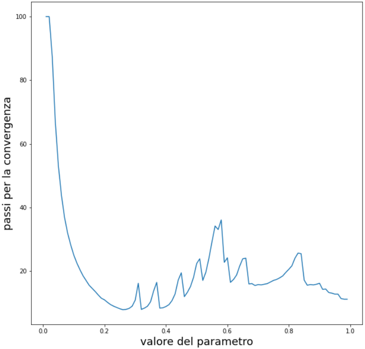
Iperboloide



Punto di minimo: 0.27, valore di convergenza minimo: 7.35.

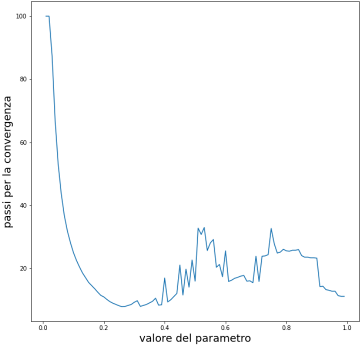
Scelta del parametro metodo di Armijo

Disco di Poincaré



Punto di minimo: 0.26, valore di convergenza minimo: 7.9.

Iperboloide



Punto di minimo: 0.26, valore di convergenza minimo: 7.9.

Esperimenti

Metodi di confronto tra le due implementazioni

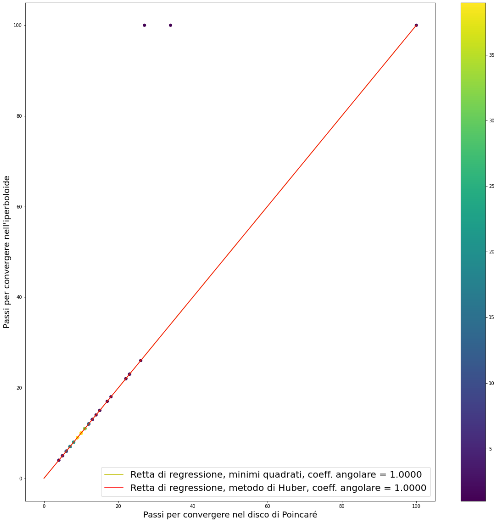
Per ogni algoritmo lo abbiamo eseguito su ogni istanza del dataset totale creando una serie di punti $(a, b) \in \mathbb{R}^2$.

- a: numero di passi per la convergenza su disco
- b: il numero di passi per la convergenza su iperboloide

Dalla nuvola di punti si sono eliminati quei punti con una molteplicità bassa.

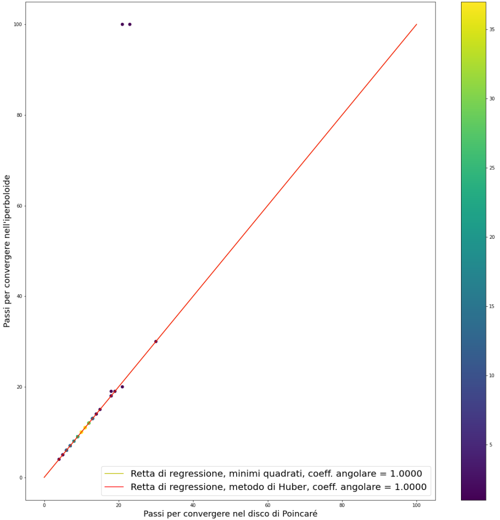
Si sono calcolate due rette di regressione, la prima tramite il metodo dei minimi quadrati, la seconda tramite l'algoritmo di Huber.

Esperimenti



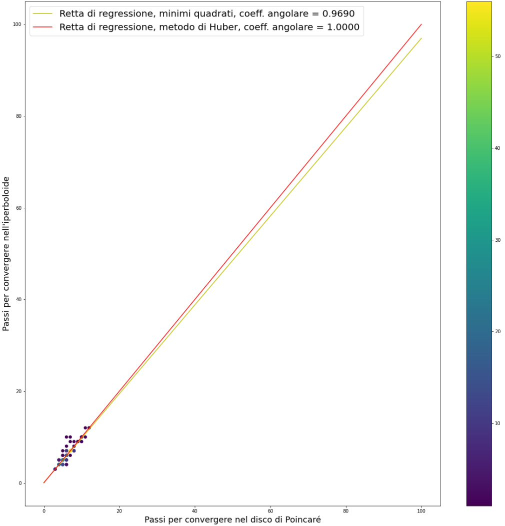
Nuvola di punti relativa all'algoritmo a passo fisso.

Esperimenti



Nuvola di punti relativa all'algoritmo di Armijo.

Esperimenti



Nuvola di punti relativa all'algoritmo di Barzilai-Borwein.

Convergenza media

Confronto passi di convergenza medi per gli algoritmi implementati.

	Passo fisso	Armijo	B.B.
Disco di Poincaré	10.5	10.4	6.13
Iperboloide	11.2	11.2	6.8

Conclusioni

Risultato

È possibile notare come l'algoritmo di Barzilai-Borwein si comporta meglio rispetto agli algoritmi a passo fisso e di ricerca lineare (Armijo).

Congettura

È possibile trasferire un problema tra due varietà conformi, deve comunque essere preservata l'esistenza di una soluzione ottima per la funzione costo.

Conclusioni

Contributi

Si sono implementate le classi per la gestione del disco di Poincaré e dell'iperboloide, che saranno proposte come aggiunta alla libreria PyManopt implementazione python della libreria Manopt

Grazie per l'attenzione.