STOCHASTIC ALGORITHMS AND CONSTRAINED OPTIMIZATION

Dario Comanducci, 22 aprile 2024

Esercizio 1

$$\min_{x,y} f(x,y) = x^2 - xy^2 + y^4$$

Carry out, if possible, an iteration of Newton's method starting at point $(7,1)$.

Svolgimento

Nel metodo di Newton l'iterazione al passo (k) avviene secondo

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left[\nabla^2 f(\mathbf{x}_k)\right]^{-1} \nabla f(\mathbf{x}_k)$$

Valutiamo quindi

$$\nabla f(7,1) = \begin{bmatrix} 2x - y^2 \\ 4y^3 - 2xy \end{bmatrix}_{(7,1)} = \begin{bmatrix} 13 \\ -10 \end{bmatrix}$$
$$\nabla^2 f(7,1) = \begin{bmatrix} 2 & -2y \\ -2y & 12y^2 - 2x \end{bmatrix}_{(7,1)} = \begin{bmatrix} 2 & -2 \\ -2 & -2 \end{bmatrix}$$

Notiamo che $\nabla^2 f(7,1)$ è invertibile ma non definita positiva: in particolare i suoi autovalori valgono[1] $\pm\sqrt{8}$, quindi l'ipotesi su cui il metodo di Newton si basa non è soddisfatta.

Considerazione personale A causa dell'autovalore negativo non è detto in generale che la direzione ottenuta sia di discesa; tuttavia nel caso in esame

$$\mathbf{d}_0 = -\left[\nabla^2 f(\mathbf{x}_k)\right]^{-1} \nabla f(\mathbf{x}_k) = -\frac{1}{4} \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 13 \\ -10 \end{bmatrix} = -\frac{1}{4} \begin{bmatrix} 23 \\ -3 \end{bmatrix}$$

è una direzione di discesa, in quanto $\mathbf{d}_0^\top \nabla f(7,1) = -329/4 < 0$.

Proviamo quindi a vedere se la condizione di Armijo può essere soddisfatta: pur non avendo fissato la costante c_1, verifichiamo intanto cosa abbiamo per $\alpha_0 = 1$ essendo $f(\mathbf{x}_0) = 43$ e

$$\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{d}_0 = \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 23 \\ -3 \end{bmatrix} = \frac{1}{4} \left(\begin{bmatrix} 28 \\ 4 \end{bmatrix} - \begin{bmatrix} 23 \\ -3 \end{bmatrix} \right) = \frac{1}{4} \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$

[1] $\det\left(\nabla^2 f(7,1) - \lambda \mathtt{I}\right) = \begin{vmatrix} 2-\lambda & -2 \\ -2 & -2-\lambda \end{vmatrix} = \lambda^2 - 8 = 0 \Rightarrow \lambda = \pm\sqrt{8}$

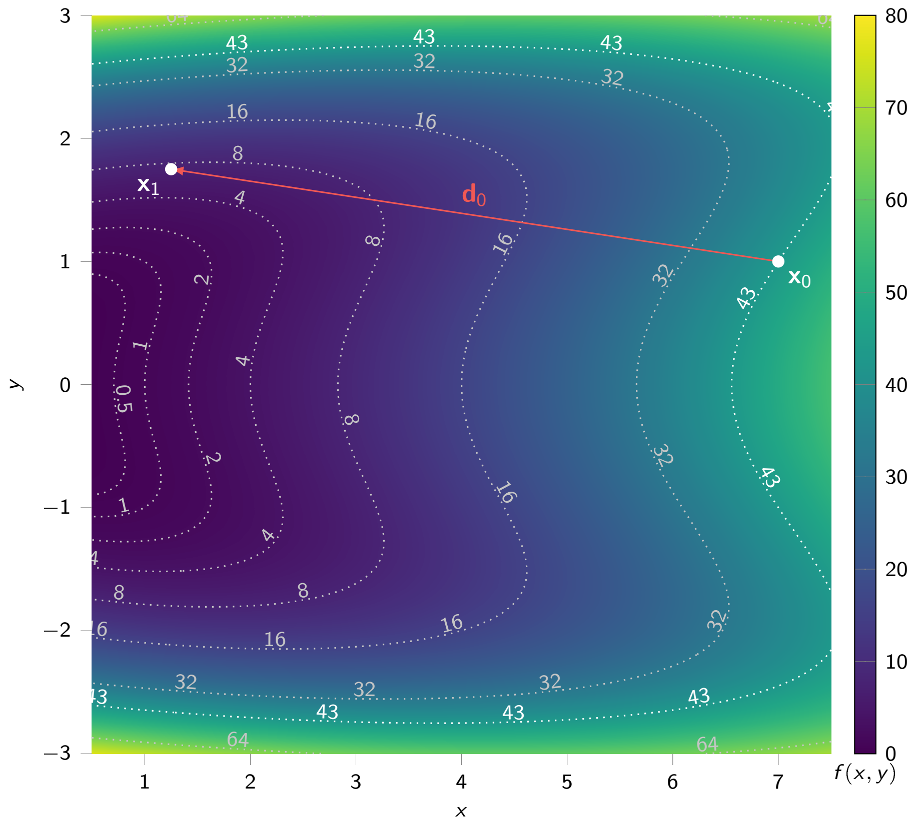

Fig. 1: Possibile applicazione della direzione $\mathbf{d}_0$ anche se la matrice Hessiana in $\mathbf{x}_0$ non è definita positiva.

Pertanto, per $\mathbf{x}_0$, $\mathbf{d}_0$, e $\alpha_0 = 1$, la condizione di Armijo in funzione di c_1 vale:

$$f(\mathbf{x}_1) - f(\mathbf{x}_0) \leqslant c_1 \mathbf{d}_0^\top \nabla f(\mathbf{x}_0)$$

ossia

$$f(\mathbf{x}_1) - f(\mathbf{x}_0) = \frac{1821}{256} - 43 = -\frac{9187}{256} \leqslant -\frac{329}{4} c_1 \iff c_1 \leqslant \frac{9187}{64 \cdot 329} \approx 0.436$$

Quindi i tipici valori di c_1 (es. $c_1 = 10^{-4}$) sarebbero pure coerenti con l'applicazione di un passo di line-search lungo $\mathbf{d}_0$ per $\alpha_0 = 1$.

In definitiva, anche se la matrice Hessiana in $\mathbf{x}_0$ non è definita positiva, perché in questo caso non potremmo comunque permetterci di andare in $\mathbf{x}_1$ (Fig. 1)?

Esercizio 2

$$\min_{a,b} \frac{1}{4} \sum_{n=1}^{4} f_n(a, b)$$

$$f_1(a, b) = a^2 + 2b^4 \quad f_2(a, b) = 3a^2 + b \quad f_3(a, b) = -a + b \quad f_4(a, b) = ab$$

Find the updating direction in SGD at $(1, 1)$ *with* $B_k = \{3, 4\}$ *as minibatch.*

Svolgimento

Nel metodo Minibatch SGD la direzione di aggiornamento al passo k è opposta al gradiente calcolato solo sul sottoinsieme B_k delle $f_n(\cdot)$:

$$\begin{aligned}\mathbf{d}_k &= -\frac{1}{\|B_k\|}\sum_{n\in B_k}\nabla f_n\big(a^{(k)},b^{(k)}\big) = \frac{1}{2}\Big(\nabla f_3\big(a^{(k)},b^{(k)}\big)+\nabla f_4\big(a^{(k)},b^{(k)}\big)\Big)\\ &= -\frac{1}{2}\left(\begin{bmatrix}-1\\1\end{bmatrix}+\begin{bmatrix}b\\a\end{bmatrix}_{(1,1)}\right)=\begin{bmatrix}0\\-1\end{bmatrix}\end{aligned}$$

Esercizio 3

$$\begin{aligned}\min_{x,y}\ & f(x,y) = 2x^2+y^2+2xy\\ \text{s.t.}\ & 0\leqslant x\leqslant 2\\ & 1\leqslant y\leqslant 3\end{aligned}$$

Carry out one iteration of projected gradient method at point $(1,2)$ *(Armijo:* $\alpha_0 = 1, c_1 = 0.1, \delta = 0.5$*), and then check whether the algorithm should go or if it reached a stationary point.*

Svolgimento

L'iterazione del metodo del gradiente proiettato è data da

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\underbrace{\big(P_{\mathsf{S}}\big(\mathbf{x}_k-\nabla f(\mathbf{x}_k)\big)-\mathbf{x}_k\big)}_{\mathbf{d}_k}$$

Determiniamo perciò il gradiente di f

$$\nabla f(x,y) = \begin{bmatrix}4x+2y\\2x+2y\end{bmatrix} = 2\begin{bmatrix}2x+y\\x+y\end{bmatrix}$$

ed il punto di appoggio $\mathbf{y}_0$ a partire dal punto $\mathbf{x}_0 = [1\ \ 2]^\top$

$$\mathbf{y}_0 = \mathbf{x}_0 - \nabla f(\mathbf{x}_0) = \begin{bmatrix}1\\2\end{bmatrix}-\begin{bmatrix}8\\6\end{bmatrix}=\begin{bmatrix}-7\\-4\end{bmatrix}$$

la cui proiezione sull'insieme box $S = [0,2]\times[1,3]$ è data da

$$\mathbf{p}_0 = P_{\mathsf{S}}(\mathbf{y}_0) = \begin{bmatrix}0\\1\end{bmatrix}$$

Quindi la direzione per il metodo del gradiente proiettato vale, al passo iniziale,

$$\mathbf{d}_0 = \mathbf{p}_0 - \mathbf{x}_0 = \begin{bmatrix}0\\1\end{bmatrix}-\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}-1\\-1\end{bmatrix}$$

portando l'algoritmo nel punto $\mathbf{x}_1$

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0\mathbf{d}_0 = \mathbf{p}_0$$

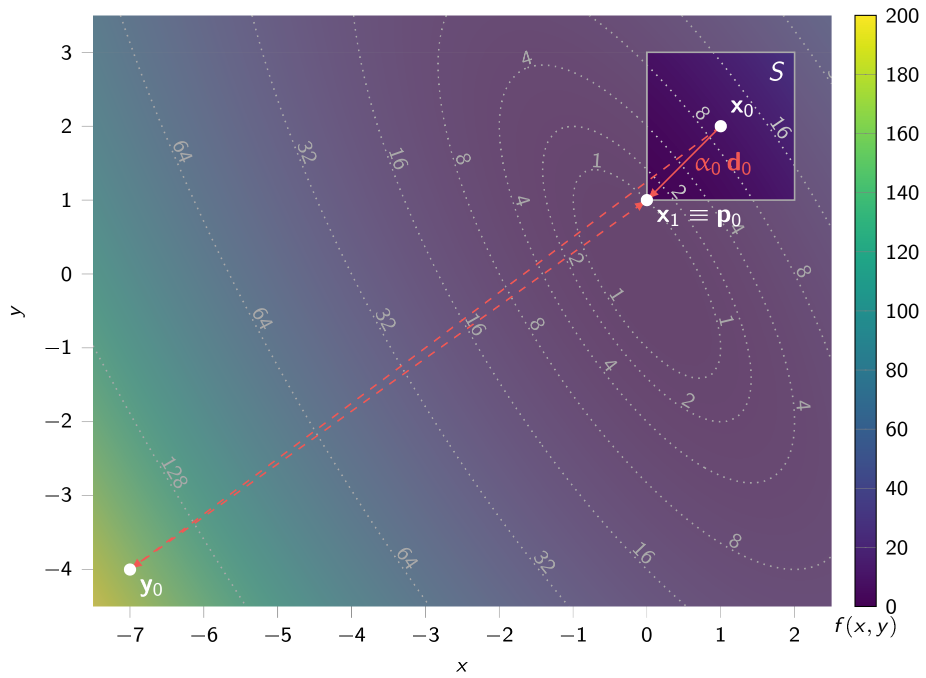

Fig. 2: Prima iterazione nel metodo del gradiente proiettato per l'esercizio 3: il punto di partenza si trova in $\mathbf{x}_0$; $\mathbf{x}_1$ è punto stazionario.

con $\alpha_0 = 1$ che soddisfa la condizione di Armijo per $c_1 = 0.1$.[2]

Notiamo che l'algoritmo è giunto in un punto stazionario: infatti $\mathbf{x}_1$ è un punto di frontiera di S che soddisfa le condizioni di punto stazionario per un insieme box; in particolare

$$\frac{\partial f(\mathbf{x}_1)}{\partial x} = 2 \geqslant 0$$
$$\frac{\partial f(\mathbf{x}_1)}{\partial y} = 2 \geqslant 0$$

Fig. 2 riassume la situazione.

[2] $f(\mathbf{x}_0 + \alpha_0 \mathbf{d}_0) = 1 \leqslant f(\mathbf{x}_0) + c_1 \alpha_0 \, \mathbf{d}_0^\top \nabla f(\mathbf{x}_0) = 10 - 0.1 \cdot 14 = 8.6$