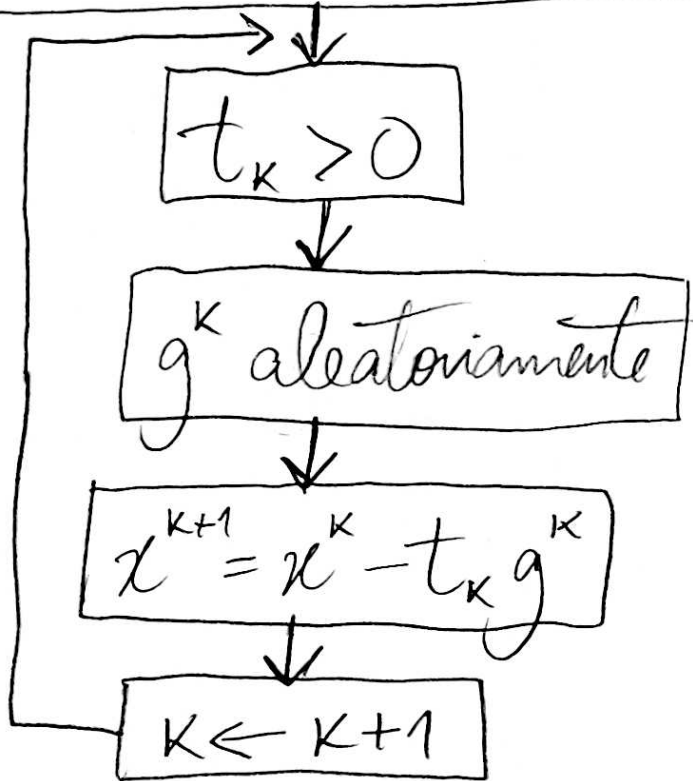


Convergência do método do gradiente estocástico (1)

Dado $x^0 \in \mathbb{R}^m$, $k \leftarrow 0$



Sequência gerada:
 $\{x^k\}$.

Problema: $\min_x f(x)$

HIPÓTESES.

$$H1) E(g^k | x^k) = \nabla f(x^k) \\ (\text{sem vies})$$

$$H2) \exists L > 0 \text{ tal que} \\ E(\|g^k\|^2 | x^k) \leq L^2, \forall k.$$

Passo decrescente:

$$(1) \ t_k \rightarrow 0^+, \quad \sum_{k=0}^{\infty} t_k = \infty, \quad \sum_{k=0}^{\infty} t_k^2 < \infty.$$

Teorema: Suponha f convexa, $H1, H2$ válidas e que f admita minimizador x^* . Seja $\{t_k\}$ como em (1). Então

$$\lim_{k \rightarrow \infty} E(f_k) = f^*$$

onde $f_k = \min_{0 \leq j \leq k} f(x^j)$ e $f^* = \min_x f(x) = f(x^*)$.

Prova: Para cada $k \geq 0$ temos (3)

$$\begin{aligned} E(\|x^{k+1} - x^*\|^2 | x^k) &= E(\|x^k - t_k g^k - x^*\|^2 | x^k) \\ &= E(\|x^k - x^*\|^2 - 2t_k (g^k)^t (x^k - x^*) + t_k^2 \|g^k\|^2 | x^k) \\ &= E(\|x^k - x^*\|^2 | x^k) - 2t_k E((g^k)^t (x^k - x^*) | x^k) \\ &\quad + t_k^2 E(\|g^k\|^2 | x^k) \end{aligned}$$

Note que $E(\|x^k - x^*\|^2 | x^k) = \|x^k - x^*\|^2$ visto que, dado x^k , $\|x^k - x^*\|^2$ fica determinado.

[4] Também, $E((g^k)^t (x^k - x^*) | x^k) = E(\sum g_i^k (x_i^k - x_i^*) | x^k)$
 $= \sum E(g_i^k (x_i^k - x_i^*) | x^k) = \sum E(g_i^k | x^k) (x_i^k - x_i^*)$
 cte em x^k

$$= E(g^k | x^k)^t (x^k - x^*)$$

Assim,

$$E(\|x^{k+1} - x^*\|^2 | x^k) = \|x^k - x^*\|^2 - 2t_k E(g^k | x^k)^t (x^k - x^*) + t_k^2 E(\|g^k\|^2 | x^k)$$

Relax hipóteses H1 e H2, obtemos

$$E(\|x^{k+1} - x^*\|^2 | x^k) \leq \|x^k - x^*\|^2 - 2t_k \left[\nabla f(x^k)^t (x^k - x^*) \right] + t_k^2 L^2, \quad (5)$$

f convexa

$$\leq \|x^k - x^*\|^2 - 2t_k (f(x^k) - f^*) + t_k^2 L^2.$$

Tomando a esperança em ambos os lados da desigualdade e lembrando que $E(E(w|z)) = E(w)$ e que $w \leq z \Rightarrow E(w) \leq E(z)$, obtemos

$$E(\|x^{k+1} - x^*\|^2) \leq E(\|x^k - x^*\|^2) - 2t_k (E(f(x^k)) -$$

$$-f^*) + t_k^2 L^2.$$

(6)

Aplicando essa desigualdade sucessivas vezes,

$$E(\|x^{k+1} - x^*\|^2) \leq E(\|x^0 - x^*\|^2) - 2 \sum_{j=0}^k t_j (E(f(x^j))$$

$$- f^*) + L^2 \sum_{j=0}^k t_j^2$$

$$\Rightarrow \sum_{j=0}^k t_j (E(f(x^j)) - f^*) \leq \frac{1}{2} \left[E(\|x^0 - x^*\|^2) + L^2 \sum_{j=0}^k t_j^2 \right].$$

Logo, usando o fato $E(\min\{w, z\}) \leq$


$\min\{E(w), E(z)\}$, temos

$$E(f_k) = E\left(\min_{0 \leq j \leq k} f(x^j)\right) \leq \min_{0 \leq j \leq k} E(f(x^j)),$$

e logo

$$0 \leq E(f_k) - f^* \leq \frac{\|x^0 - x^*\|^2 + L^2 \sum_{j=0}^k t_j^2}{2 \sum_{j=0}^k t_j}$$

(note que $E(\|x^0 - x^*\|^2) = \|x^0 - x^*\|^2$ pois x^0 e

x^* são determinísticos). Fazendo $k \rightarrow \infty$ (e obtemos $\lim_{k \rightarrow \infty} E(f_k) = f^*$ das hipóteses sobre $\{t_k\}$. 

Obs: compare com a prova de convergência do método do gradiente incremental.

Passo constante $t_k = t, \forall k$: assim como no gradiente incremental, é possível provar convergência do método do gradiente

estocásticos para funções convexas e passo $L\eta$ constante (o SG "básico" usa $t_k = \text{cte}$ na prática).

Vamos nos inspirar nas provas do gradiente estocástico para passo decrescente e do subgradiente para passo constante.

Seguindo a prova anterior, note que

$$E(\|x^{k+1} - x^*\|^2) \leq E(\|x^k - x^*\|^2) - 2t_k(E(f(x^k)) - f^*) + t_k^2 L^2 \quad (2)$$

permanece válido independentemente da escolha de t_k . Essa desigualdade é uma versão do lema do método do subgradiente com esperanças (compare!). (10)

Seguindo a prova de convergência do subgradiente para o caso $f^* > -\infty$ (estamos supondo aqui que f admite minimizadores), suponha

que

$$\liminf_{k \rightarrow \infty} E(f(x^k)) > f^* + \frac{tL^2}{2} + 2\epsilon, \quad (3),$$

para algum $\varepsilon > 0$, Assim, $t_k = t > 0, \forall k$. (11)

Também, para todo $k \gg 1$ temos

$$E(f(x^k)) \geq \liminf_{k \rightarrow \infty} E(f(x^k)) - \varepsilon. \quad (4)$$

Tomando (3) e (4) obtemos, para um certo k_0 ,

$$E(f(x^k)) - f^* \geq \frac{tL^2}{2} + \varepsilon, \quad \forall k \geq k_0.$$

Assim, da inequação (2) com $t_k = t$ temos

$$\begin{aligned}
 E(\|x^{k+1} - x^*\|^2) &\leq E(\|x^k - x^*\|^2) - 2t \left(\frac{tL^2}{2} + \varepsilon \right) + t^2 L^2 \\
 &= E(\|x^k - x^*\|^2) - 2t\varepsilon, \quad \forall k \geq k_0.
 \end{aligned}$$

Aplicando essa desigualdade sucessivas vezes
obtemos

$$E(\|x^{k+1} - x^*\|^2) \leq E(\|x^{k_0} - x^*\|^2) - 2(k+1-k_0)t\varepsilon.$$

Tomando $k \gg k_0$ obtemos uma contradição
pois $-2(k+1-k_0)t\varepsilon \rightarrow -\infty$ e $E(W) \geq 0$
quando $W \geq 0$.

A contradição vem da suposição (3), isto é, (13)
que $\exists \epsilon > 0$ tal que

$$\liminf_{k \rightarrow \infty} E(f(x^k)) > f^* + \frac{tL^2}{2} + 2\epsilon.$$

Em outras palavras, vale

$$\lim_{k \rightarrow \infty} E(f_k) - f^* \leq \frac{tL^2}{2}.$$

Compare com o resultado do método do subgradiente: só há garantia de chegar próximo do valor ótimo !!!

Isso demonstra o seguinte teorema: (14)

Teorema: Suponha f convexa, $H1, H2$ válidas e que f admita minimizador x^* .

Suponha ainda que $t_k = t > 0, \forall k$.

Então

$$\lim_{k \rightarrow \infty} E(f_k) \leq f^* + \frac{tL^2}{2},$$

onde $f_k = \min_{0 \leq j \leq k} f(x^j)$ e $f^* = \min_x f(x) = f(x^*)$.