UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Matemática

Título a definir

Axel Sirota

Director de Tesis: Dr. Pablo Amster

Lugar de Trabajo: Departamento de Matemática- FCEyN (UBA)

Buenos Aires, Septiembre?? de 2018.

Índice general

Capítulo 1

Intuición

Usemos un caso modelo para ejemplificar porque no es probable que los metodos de primer orden (entre ellos *gradient descent*) convergan a puntos silla. Sea $f : \mathbb{R}^n \to \mathbb{R}^n$ dada por $f(x) = \frac{1}{2}x^T H x$ con $H = \mathbf{diag}\,(\lambda_1, \ldots, \lambda_n)$; supongamos además que $\lambda_1, \ldots, \lambda_k > 0$ y $\lambda_{k+1}, \ldots, \lambda_n < 0$.

Si usamos en la base canónica de $\mathbb{R}^n$ $\{e^1, \ldots, e^n\}$ entonces:

$$f(x) = f(x^1, \ldots, x^n) = \frac{1}{2}\left(\lambda_1 {x_1}^2 + \cdots + \lambda_n {x_n}^2\right)$$

Por lo tanto:

$$\nabla f(x) \;=\; \lambda_i x_i e^i = 0 \iff x = x_1 e^1 = 0$$

Y tenemos que en el único punto crítico el Hessiano de f es $\nabla^2 f(0) = H$.

Recordemos que si $g(x) = x - \alpha \nabla f(x)$ entonces *gradient descent* está dado por la iteración $x_{t+1} = \; g(x_t) \; := g^t(x_0)$ con $t \in \mathbb{N}$ y $x_0 \in \mathbb{R}^n$, y en este caso esta representado por:

$$\begin{aligned} x_{t+1} &= g(x_t) \\ &= x_t - \alpha \nabla f(x_t) \\ &= (1 - \alpha\lambda_i)\, x_{it} e^i \\ &= (1 - \alpha\lambda_i)\left\langle x_t, e^i \right\rangle e^i \end{aligned}$$

Por lo tanto por inducción es fácil probar que:

$$x_{t+1} = (1 - \alpha\lambda_i)^t \left\langle x_o, e^i \right\rangle e^i$$

Sea $L = \max_i |\lambda_i|$ y supongamos que $\alpha < \frac{1}{L}$, luego:

$$\begin{aligned} 1 - \alpha\lambda_i < 1 \quad & \text{Si } i \leq k \\ 1 - \alpha\lambda_i > 1 \quad & \text{Si } i > k \end{aligned}$$

Con lo que concluímos que:

$$\lim_t x_t = \begin{cases} 0 & \text{Si } x \in E_s := \left\langle e^1, \ldots, e^k \right\rangle \\ \infty & \text{Si no} \end{cases}$$

Finalmente, si $k < n$ entonces concluímos que:

$$P_{\mathbb{R}^n}(\left\{ x \in \mathbb{R}^n \;/\; \lim_t g^t(x) = 0 \right\}) = |E_s| = 0$$

Para notar este fenómeno en un ejemplo no cuadrático consideremos $f(x,y) = \frac{1}{2}x^2 + \frac{1}{4}y^4 - \frac{1}{2}y^2$, reproduciendo los calculos anteriores:

$$\begin{aligned} \nabla f &= \big(x, y^3 - y\big) \\ g &= \big((1-\alpha)x, (1+\alpha)y - \alpha y^3\big) \\ \nabla^2 f &= \begin{pmatrix} 1 & 0 \\ 0 & 3y^2 - 1 \end{pmatrix} \end{aligned} \tag{1.1}$$

De lo que vemos que los puntos críticos son:

$$z_1 = (0,0) \qquad z_2 = (0,1) \qquad z_3 = (0,-1)$$

Y del critério del Hessiano concluímos que z_2, z_3 son mínimos locales mientras que z_1 es un punto silla. De la intuición previa, como en z_1 el autovector asociado al autovalor positivo es e^1 podemos intuir que:

Lema 1.0.0.1 *Para* $f(x,y) = \frac{1}{2}x^2 + \frac{1}{4}y^4 - \frac{1}{2}y^2$ *resulta que* $E_s = \langle t * e^1 \ / \ t \in \mathbb{R}\rangle := W_s$

Asumiendo el resultado por un momento, dado que $\dim_{\mathbb{R}^2}(E_s) = 1 < 2$ entonces $P_{\mathbb{R}^2}(E_s) = 0$ que es lo que queríamos verificar. Demostremos el lema ahora:

Demostración Del lema Sea $x_0 \in \mathbb{R}^n$ y g la iteración de *gradient descent* dada por 1, luego:

$$(x_t, y_t) = g^t(x,y) = \begin{pmatrix} (1-\alpha)^t x_0 \\ g_y^t(y_0) \end{pmatrix} \underset{(t\to\infty)}{\longrightarrow} \begin{pmatrix} 0 \\ \lim_t g_y^t(y_0) \end{pmatrix}$$

Por lo que todo depende de y_0. Analizando $\dfrac{dg_y}{dy} = 1 + \alpha - 3\alpha y^2$ notemos que:

$$\begin{aligned} \left|\frac{dg_y}{dy}\right| < 1 &\iff \left|1 + \alpha - 3\alpha y^2\right| < 1 \\ &\iff -1 < 1 + \alpha - 3\alpha y^2 < 1 \\ &\iff -2 - \alpha < -3\alpha y^2 < -\alpha \\ &\iff \sqrt{\frac{2+\alpha}{3\alpha}} > |y| > \sqrt{\frac{1}{3}} \\ &\iff \sqrt{\frac{1+\frac{2}{\alpha}}{3}} > |y| > \sqrt{\frac{1}{3}} \end{aligned}$$

Por lo que por el Teorema de Punto Fijo de Banach:

$$\lim_t g_y^t(y_0) = \begin{cases} 1 & \text{Si } \sqrt{\frac{1+\frac{2}{\alpha}}{3}} > y_0 > \sqrt{\frac{1}{3}} \\ -1 & \text{Si } \sqrt{\frac{1+\frac{2}{\alpha}}{3}} < -y_0 < \sqrt{\frac{1}{3}} \end{cases}$$

Si analizamos simplemente los signos de g y $\dfrac{dg_y}{dy}$ en los otros intervalos podemos conluir que:

$$\lim_t g_y^t(y_0) = \begin{cases} -\infty & \text{Si } y_0 > \sqrt{\dfrac{1+\frac{2}{\alpha}}{3}} \\ 1 & \text{Si } \sqrt{\dfrac{1+\frac{2}{\alpha}}{3}} > y_0 > 0 \\ -1 & \text{Si } -\sqrt{\dfrac{1+\frac{2}{\alpha}}{3}} < y_0 < 0 \\ \infty & \text{Si } y_0 < -\sqrt{\dfrac{1+\frac{2}{\alpha}}{3}} \end{cases}$$

Dedujimos entonces que $(x,y) \in E_s \iff (x,y) = (t,0)\ t \in \mathbb{R} \iff (x,y) \in W_s$. ∎