Apéndice A

Búsqueda por líneas en espacios de Banach

La única cosa realmente valiosa es la intuición.
Albert Einstein (1879-1955).

Los métodos de búsqueda POR líneas se introdujeron en la década de los ochenta teniendo un amplio uso en la Ingeniería y en la matemáticas, como se menciona en [101]. Sin embargo, estos métodos optimizan funciones de dominio en un subespacio de $\mathbb{R}^n$, como se puede ver en los artículos [102]-[107], donde proponen varios métodos de búsqueda relacionados con la forma, como se halla el tamaño del paso o el tipo de función. Para todos ellos, se han realizando pruebas de convergencia. En [108], [109], generalizan la teoría de optimización a espacios de Banach, pero no mencionan los métodos de búsqueda por línea. En este apéndice se generaliza la búsqueda por líneas a espacios de Banach; además se prueba la convergencia del método.

A.1. Generalización de la búsqueda por líneas

Problema de minimización

$$\min f(x); \qquad x \in \mathbb{E} \tag{A.1}$$

donde $\mathbb{E}$ espacio de Banach, $f : \mathbb{E} \to \mathbb{R}$ función diferenciable y continua en $\mathbb{E}$. El método de línea para resolver (1) esta dado por:

$$x_{k+1} = x_k + \alpha_k d_k \tag{A.2}$$

donde x_k es un punto iterativo, d_k la búsqueda direccional y α_k entero positivo llamado salto-paso. Denotamos $\nabla f(x_k)$ por g_k, $f(x_k)$ por f_k, $f(x*)$ por $f*$, respectivamente. Sea $x*$ el minimizador de (1), así $g(x*) = 0 \in \mathbb{E}$.

La búsqueda direccional d_k requiere que satisfaga la condición descendiente

$$\langle g_k, d_k \rangle < 0 \tag{A.3}$$

lo cual garantiza que d_k sea una dirección descendiente de $f(x)$ en x_k.

A.2. Reglas de búsqueda por línea

Primero suponemos que

(H1): La función f tiene cota inferior en $L_0 = \{x \in \mathbb{E} : f(x) \leq f(x_0)\}$ donde $x_0 \in \mathbb{E}$.

(H2): El gradiente g es Lipschitz continuo en un conjunto abierto convexo B, el cual contiene a L_0, i.e, existe $L > 0$ tal que

$$\|g(x) - g(y)\| \leq L \|x - y\|, \qquad \forall x, y \in \mathbb{E} \tag{A.4}$$

reglas, para escoger el paso α_k, son:

(a) *Minimización.* Para cada iteración α_k es seleccionado tal que

$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha_k d_k) \tag{A.5}$$

(b) *Minimización Aproximada.* Para cada iteración α_k, es seleccionado tal que

$$\alpha_k = \text{mín}\left\{\alpha : \left\langle g\left(x_k + \alpha d_k\right), d_k\right\rangle = 0, \alpha > 0\right\} \tag{A.6}$$

(c) *Armijo.* Sean escalares s_k, β y σ con $s_k = -\frac{\langle g_k, d_k\rangle}{\|d_k\|^2}$, $\beta \in (0,1)$ y $\sigma \in \left(0, \frac{1}{2}\right)$. Sea $\alpha_k = \beta^{m_k}.s_k$, donde m_k es el primer entero no negativo m para el cual

$$f_k - f\left(x_k + \beta^m s_k d_k\right) \geq -\sigma\beta^m s_k \left\langle g_k, d_k\right\rangle \tag{A.7}$$

i,e; $m = 0, 1, ...$ sucesivamente hasta que la desigualdad anterior se satisfaga para $m = m_k$.

(d) *Minimización Límitada.* Sea $s_k = -\frac{\langle g_k, d_k\rangle}{\|d_k\|^2}$, tal que

$$f\left(x_k + \alpha_k d_k\right) = \min_{\alpha\in[0,s_k]} f\left(x_k + \alpha_k d_k\right) \tag{A.8}$$

(e) *Goldstein.* Un escalar fijo $\sigma \in \left(0, \frac{1}{2}\right)$ es seleccionado, y α_k es escogido tal que

$$\sigma \leq \frac{f\left(x_k + \alpha_k d_k\right) - f_k}{\alpha_k \left\langle g_k, d_k\right\rangle} \leq 1 - \sigma \tag{A.9}$$

Es posible demostrar que si f es acotada, existe un intervalo de tamaño de paso α_k que cumpla la relación anterior, y son bastante sencillos de encontrar a través de un número finito de operaciones aritméticas.

(f) *Wolfe fuerte.* α_k es escogido para que satisfaga simultáneamente,

$$f_k - f\left(x_k + \alpha_k d_k\right) \geq -\sigma\alpha_k \left\langle g_k, d_k\right\rangle \tag{A.10}$$

y

$$\left|\left\langle g\left(x_k + \alpha_k d_k\right), d_k\right\rangle\right| \leq -\beta \left\langle g_k, d_k\right\rangle \tag{A.11}$$

donde σ y β son escalares tal que $\sigma \in \left(0, \frac{1}{2}\right)$ y $\beta \in (\sigma, 1)$.

(q) *Wolfe.* α_k es escogido para que satisfaga (10) y

$$\left\langle g\left(x_k + \alpha_k d_k\right), d_k\right\rangle \geq \beta \left\langle g_k, d_k\right\rangle \tag{A.12}$$

A.3. Método de búsqueda por línea

Algoritmo. Escoja un parámetro y seleccione un punto inicial x_1, $k := 0$

Paso 1. Si $\|g_k\| = 0$ entonces, deténgase de lo contrario, prosiga con el paso 2.

Paso 2. $x_{k+1} = x_k + \alpha_k d_k$ donde d_k es dirección descendente de $f(x)$ en x_k. α_k es definida por las reglas de búsqueda línea (a), (b), (c), (d), (e), (f) o (g).

Paso 3. $k = k + 1$ volver al paso 1.

Para las anteriores siete reglas de búsqueda línea, denotamos el correspondiente algoritmo por Algoritmo (a)-(g), respectivamente.

A.4. Propiedad de Convergencia

Si (H1), (H2) y (3) se cumplen, entonces para cualquier regla de búsqueda línea la siguiente propiedad de convergencia se obtiene:

$$\lim_{k\to\infty} \left(-\frac{\langle g_k, d_k \rangle}{\|d_k\|} \right)^2 = 0 \tag{A.13}$$

Teorema 7 *Si (H1), (H2) y (3) se cumple, entonces el Algoritmo (a) produce una sucesión infinita $\{x_k\}$, tal que (13) se satisface.*

Prueba 6 *Sea $\alpha = -\frac{\langle g_k, d_k \rangle}{L\|d_k\|^2}$, por la regla de línea (a), tenemos*

$$f(x_k + \alpha_k d_k) \le f(x_k + \alpha d_k)$$
$$f_{k+1} \le f(x_k + \alpha d_k)$$

$$f_k - f_{k+1} \ge f_k - f(x_k + \alpha d_k)$$

Por el teorema fundamental del Cálculo, tenemos:

$$\begin{aligned} f(x_k + \alpha d_k) - f_k &= \int_0^1 \langle g(x_k + t\alpha d_k), \alpha d_k \rangle \, dt \\ &= \alpha \int_0^1 \langle g(x_k + \alpha t d_k), d_k \rangle \, dt \end{aligned}$$

así

$$\begin{aligned} f_k - f_{k+1} \geq f_k - f\left(x_k + \alpha d_k\right) &= -\alpha \int_0^1 \left\langle g\left(x_k + \alpha t d_k\right), d_k \right\rangle dt \\ &= -\alpha \left\langle g_k, d_k \right\rangle - \alpha \int_0^1 \left\langle g\left(x_k + \alpha t d_k\right) - g_k, d_k \right\rangle dt \end{aligned}$$

Por la desigualdad de Cauchy-Schwartz, tenemos:

$$\left\langle g\left(x_k + \alpha t d_k\right) - g_k, d_k \right\rangle \leq \left\| g\left(x_k + \alpha t d_k\right) - g_k \right\| \left\| d_k \right\| \qquad \text{para } \; t \in [0,1]$$

$$\alpha \int_0^1 \left\langle g\left(x_k + \alpha t d_k\right), d_k \right\rangle dt \leq \alpha \int_0^1 \left\| g\left(x_k + \alpha t d_k\right) - g_k \right\| \left\| d_k \right\| dt$$

así obtenemos;

$$\begin{aligned} f_k - f_{k+1} &\geq -\alpha \left\langle g_k, d_k \right\rangle - \alpha \int_0^1 \left\langle g\left(x_k + \alpha t d_k\right) - g_k, d_k \right\rangle dt \\ &\geq -\alpha \left\langle g_k, d_k \right\rangle - \alpha \int_0^1 \left\| g\left(x_k + \alpha t d_k\right) - g_k \right\| \left\| d_k \right\| dt \end{aligned}$$

Además, por (H2) tenemos que g es Lipschitz continua, por tanto; existe $L > 0$ tal que

$$\begin{aligned} \left\| g\left(x_k + \alpha t d_k\right) - g_k \right\| &\leq L \left\| x_k + \alpha t d_k - x_k \right\| \\ &\leq L \left\| \alpha t d_k \right\| = \alpha L t \left\| d_k \right\| \end{aligned}$$

$$\alpha \int_0^1 \left\| g\left(x_k + \alpha t d_k\right) - g_k \right\| \left\| d_k \right\| dt \leq \int_0^1 \alpha^2 L t \left\| d_k \right\|^2 dt = \frac{\alpha^2}{2} L \left\| d_k \right\|^2$$

por lo tanto,

$$\begin{aligned} f_k - f_{k+1} &\geq -\alpha \left\langle g_k, d_k \right\rangle - \alpha \int_0^1 \left\| g\left(x_k + \alpha t d_k\right) - g_k \right\| \left\| d_k \right\| dt \\ &\geq -\alpha \left\langle g_k, d_k \right\rangle - \frac{\alpha^2}{2} L \left\| d_k \right\|^2 \\ &= \frac{\left\langle g_k, d_k \right\rangle^2}{L \left\| d_k \right\|^2} - \frac{\left\langle g_k, d_k \right\rangle^2}{2L \left\| d_k \right\|^2} \\ &= \frac{1}{2L} \left(-\frac{\left\langle g_k, d_k \right\rangle}{\left\| d_k \right\|} \right)^2 \end{aligned}$$

o bien,

$$f_k - f_{k+1} \geq \frac{1}{2L}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2$$
$$f_k \geq f_{k+1}$$

Luego, $\{f_k\}$ *es una sucesión monotona decreciente, y por (H1) obtenemos que* $\{f_k\}$ *tiene límite; así*

$$\lim_{k\to\infty}\{f_k - f_{k+1}\} \geq \frac{1}{2L}\lim_{k\to\infty}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2 \geq 0$$

$$\lim_{k\to\infty}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2 = 0$$

Teorema 8 *Si (H1), (H2) y (3) se cumple, entonces el Algoritmo (b) produce una sucesión infinita* $\{x_k\}$ *tal que (13) se satisface.*

Prueba 7 *Por la regla de tamaño de paso (b), tenemos*

$$\begin{aligned}
&\langle g(x_k + \alpha_k d_k), d_k\rangle = 0\\
&\langle g(x_k), d_k\rangle + \alpha_k \langle g(d_k), d_k\rangle = 0\\
&\langle g(x_k), d_k\rangle = -\alpha_k \langle g(d_k), d_k\rangle\\
&\langle g(d_k), d_k\rangle < 0
\end{aligned}$$

Consideremos $t \in (0,1)$*, así*

$$t(\alpha_k - \alpha)(\alpha - \alpha_k)\langle g(d_k), d_k\rangle \leq 0$$

$$\begin{aligned}
&(\alpha_k - \alpha)\left[\langle g(t\alpha d_k), d_k\rangle - \langle g(t\alpha_k d_k), d_k\rangle\right] \leq 0\\
&(\alpha_k - \alpha)\left[\langle g(x_k + \alpha_k d_k), d_k\rangle + \langle g(t\alpha d_k), d_k\rangle - \langle g(t\alpha_k d_k), d_k\rangle\right] \leq 0\\
&(\alpha_k - \alpha)\left[\langle g(x_k + \alpha_k d_k - t\alpha_k d_k), d_k\rangle + \langle g(t\alpha d_k), d_k\rangle\right] \leq 0\\
&(\alpha_k - \alpha)\left[\langle g(x_k + (1-t)\alpha_k d_k), d_k\rangle + \langle g(t\alpha d_k), d_k\rangle\right] \leq 0\\
&(\alpha_k - \alpha)\langle g(x_k + t\alpha d_k + (1-t)\alpha_k d_k), d_k\rangle \leq 0\\
&(\alpha_k - \alpha)\langle g(x_k + (t\alpha + (1-t)\alpha_k)d_k), d_k\rangle \leq 0
\end{aligned}$$

Sea $c = x_k + (t\alpha + (1-t)\alpha_k)d_k$ *para algún* $t \in (0,1)$*, luego;*

$$c = x_k + (t\alpha + (1-t)\alpha_k)d_k + tx_k - tx_k$$
$$c = t(x_k + \alpha d_k) + (1-t)(x_k + \alpha_k d_k)$$

así, $c \in (x_k + \alpha d_k, x_k + \alpha_k d_k)$*; donde*

$$(x_k + \alpha d_k, x_k + \alpha_k d_k) = \{x \in E : x = t(x_k + \alpha d_k) + (1-t)(x_k + \alpha_k d_k), t \in (0,1)\}$$

por lo tanto,

$$\begin{aligned}(\alpha_k - \alpha)\langle g(c), d_k\rangle \leq 0 &\Leftrightarrow \langle g(c), (\alpha_k - \alpha) d_k\rangle \leq 0 \\ &\Leftrightarrow f'(c)((\alpha_k - \alpha) d_k) \leq 0 \\ &\Leftrightarrow f'(c)((x_k + \alpha_k d_k) + (x_k - \alpha d_k)) \leq 0\end{aligned}$$

Aplicando el teorema del valor medio, obtenemos

$$f(x_k + \alpha_k d_k) - f(x_k - \alpha d_k) = f'(c)((x_k + \alpha_k d_k) + (x_k - \alpha d_k)) \leq 0$$

$$f_{k+1} \leq f(x_k - \alpha d_k)$$

Actuando de manera análoga al teorema anterior y siempre que $\alpha \leq \alpha_k$*, obtenemos*

$$\begin{aligned}f_k - f_{k+1} &\geq f_k - f(x_k + \alpha d_k) \\ &= -\alpha \int_0^1 \langle g(x_k + \alpha t d_k), d_k\rangle\, dt \\ &= -\alpha\langle g_k, d_k\rangle - \alpha \int_0^1 \langle g(x_k + \alpha t d_k) - g_k, d_k\rangle\, dt \\ &\geq -\alpha\langle g_k, d_k\rangle - \alpha \int_0^1 \|g(x_k + \alpha t d_k) - g_k\|\,\|d_k\|\, dt \\ &\geq -\alpha\langle g_k, d_k\rangle - \frac{\alpha^2}{2} L \|d_k\|^2\end{aligned}$$

$$f_k - f_{k+1} \geq -\alpha\langle g_k, d_k\rangle - \frac{\alpha^2}{2} L \|d_k\|^2, \quad \alpha \leq \alpha_k$$

Por la desigualdad de Cauchy-Schwartz y la regla de búsqueda por línea (b), tenemos

$$\langle g_{k+1} - g_k, d_k\rangle \leq \|g_{k+1} - g_k\|\,\|d_k\|, \quad \langle g_{k+1} - g_k, d_k\rangle = \langle g_{k+1}, d_k\rangle - \langle g_k, d_k\rangle = -\langle g_k, d_k\rangle$$

dado que g *es Lipschitz continua;*

$$\|g_{k+1} - g_k\|\,\|d_k\| \leq L\|x_k + \alpha_k d_k - x_k\|\,\|d_k\| = \alpha_k L \|d_k\|^2$$

por lo tanto;

$$\alpha_k L \|d_k\|^2 \geq -\langle g_k, d_k\rangle$$

$$\alpha_k \geq -\frac{\langle g_k, d_k\rangle}{L\|d_k\|^2}$$

Considerando que $\alpha \leq \alpha_k$, $\alpha = -\frac{\langle g_k, d_k\rangle}{L\|d_k\|^2}$, *así,*

$$f_k - f_{k+1} \geq \frac{1}{2L}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2$$

luego, $\{f_k\}$ *es una sucesión monótona decreciente, y por (H1) obtenemos que* $\{f_k\}$ *tiene límite; así*

$$\lim_{k\to\infty}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2 = 0$$

Teorema 9 *Si (H1), (H2) y (3) se cumple, entonces el Algoritmo (c) o (d) producen una sucesión infinita* $\{x_k\}$, *tal que (13) se satisface.*

Prueba 8 *Por el algoritmo (c), consideremos los conjuntos* $K_1 = \{k : \alpha_k = s_k\}$, $K_2 = \{k : \alpha_k < s_k\}$; *por (7), tenemos*

$$f_k - f_{k+1} \geq -s_k\sigma\langle g_k, d_k\rangle; \quad \forall k \in K_1, \tag{A.14}$$

$$f_k - f_{k+1} \geq -\alpha_k\sigma\langle g_k, d_k\rangle; \quad \forall k \in K_2. \tag{A.15}$$

Aplicando el regla de búsqueda en línea (c) y como $\alpha_k/\beta \leq s_k$, $\forall k \in K_2$, *tenemos*

$$f_k - f(x_k + \alpha_k d_k/\beta) \leq -\alpha_k\sigma\langle g_k, d_k\rangle/\beta, \quad \forall k \in K_2$$

Utilizando el teorema del valor medio para la parte izquierda de la anterior desigualdad, existe $\theta_k \in [0,1]$ *tal que:*

$$\begin{aligned} f_k - f(x_k + \alpha_k d_k/\beta) &= -D_{\alpha_k d_k/\beta} f(x_k + \alpha_k\theta_k d_k/\beta) \\ &= -Df(x_k + \alpha_k\theta_k d_k/\beta)(\alpha_k d_k/\beta) \\ &= -\langle g(x_k + \alpha_k\theta_k d_k/\beta), \alpha_k d_k/\beta\rangle \\ &= -\alpha_k\langle g(x_k + \alpha_k\theta_k d_k/\beta), d_k\rangle/\beta \end{aligned}$$

$$-\alpha_k \left\langle g\left(x_k + \alpha_k\theta_k d_k/\beta\right), d_k\right\rangle/\beta \leq -\alpha_k\sigma \left\langle g_k, d_k\right\rangle/\beta, \quad \forall k \in K_2$$

Luego,

$$\left\langle g\left(x_k + \alpha_k\theta_k d_k/\beta\right), d_k\right\rangle \geq \sigma \left\langle g_k, d_k\right\rangle, \quad \forall k \in K_2 \tag{A.16}$$

Por (16) y utilizando la desigualdad de Cauchy- Schwartz, tenemos

$$\left\|g\left(x_k + \alpha_k\theta_k d_k/\beta\right) - g_k\right\| \left\|d_k\right\| \geq \left\langle g\left(x_k + \alpha_k\theta_k d_k/\beta\right) - g_k, d_k\right\rangle \geq -\left(1-\sigma\right)\left\langle g_k, d_k\right\rangle, \forall k \in K_2$$

Dado que g es Lipschitz continua, tenemos

$$\alpha_k L \left\|d_k\right\|^2/\beta \geq \alpha_k\theta_k L \left\|d_k\right\|^2/\beta \geq \left\|g\left(x_k + \alpha_k\theta_k d_k/\beta\right) - g_k\right\| \left\|d_k\right\|$$

así

$$\alpha_k L \left\|d_k\right\|^2/\beta \geq -\left(1-\sigma\right)\left\langle g_k, d_k\right\rangle$$

$$\alpha_k \geq -\frac{\beta\left(1-\sigma\right)\left\langle g_k, d_k\right\rangle}{L\left\|d_k\right\|^2}, \qquad \forall k \in K_2$$

$$\alpha_k\sigma \left\langle g_k, d_k\right\rangle \leq -\frac{\beta\sigma\left(1-\sigma\right)\left\langle g_k, d_k\right\rangle^2}{L\left\|d_k\right\|^2}$$

$$-\alpha_k\sigma \left\langle g_k, d_k\right\rangle \geq \frac{\beta\sigma\left(1-\sigma\right)}{L}\left(-\frac{\left\langle g_k, d_k\right\rangle}{\left\|d_k\right\|}\right)^2 \tag{A.17}$$

De (15) y (17) tenemos;

$$f_k - f_{k+1} \geq \frac{\beta\sigma\left(1-\sigma\right)}{L}\left(-\frac{\left\langle g_k, d_k\right\rangle}{\left\|d_k\right\|}\right)^2, \quad k \in K_2 \tag{A.18}$$

Por (14) y la definición de s_k, obtenemos

$$f_k - f_{k+1} \geq \sigma\left(-\frac{\left\langle g_k, d_k\right\rangle}{\left\|d_k\right\|}\right)^2, \quad k \in K_1 \tag{A.19}$$

Consideremos $\eta = \text{mín}\left\{\sigma, \frac{\beta\sigma(1-\sigma)}{L}\right\}$*, por (18) y (19); se tiene*

$$f_k - f_{k+1} \geq \eta \left(-\frac{\langle g_k, d_k \rangle}{\|d_k\|} \right)^2 \tag{A.20}$$

luego, $\{f_k\}$ *es una sucesión monótona decreciente, y por (H1) obtenemos que* $\{f_k\}$ *tiene límite; así*

$$\lim_{k\to\infty} \left(-\frac{\langle g_k, d_k \rangle}{\|d_k\|} \right)^2 = 0$$

Por otra parte, para el Algoritmo (d), considerando la regla de búsqueda de línea (d), tenemos

$$\begin{gathered} f\left(x_k + \alpha_k d_k\right) \leq f\left(x_k + \alpha d_k\right) \\ f_k - f_{k+1} \geq f_k - f\left(x_k + \alpha d_k\right), \quad \forall \alpha \in [0, s_k] \end{gathered}$$

Suponemos que α_k' *es el tamaño del paso en el Algoritmo (c), por tanto* $\alpha_k' \in [0, s_k]$*, luego por (20), tenemos;*

$$f_k - f_{k+1} \geq f_k - f\left(x_k + \alpha_k' d_k\right) \geq \eta \left(-\frac{\langle g_k, d_k \rangle}{\|d_k\|} \right)^2$$

actuando de manera análoga, concluimos

$$\lim_{k\to\infty} \left(-\frac{\langle g_k, d_k \rangle}{\|d_k\|} \right)^2 = 0$$

Teorema 10 *Si (H1), (H2) y (3) se cumple, entonces el Algoritmo (e) o (f) o (g) produce una sucesión infinita* $\{x_k\}$*, tal que (13) se satisface.*

Prueba 9 *Es suficiente probar que el Algoritmo (e) y (g) tienen la propiedad (13). Para el Algoritmo (e), tenemos*

$$\frac{f_{k+1} - f_k}{\alpha_k \langle g_k, d_k \rangle} \leq 1 - \sigma, \ , \ \sigma \in \left(0, \frac{1}{2}\right)$$

$$f_{k+1} - f_k \geq \alpha_k (1 - \sigma) \langle g_k, d_k \rangle \tag{A.21}$$

Aplicando el teorema del valor medio para la parte izquierda de la anterior desigualdad, existe $\theta_k \in [0,1]$*, tal que*

$$\begin{aligned} f(x_k + \alpha_k d_k) - f(x_k) &= Df(x_k + \alpha_k\theta_k d_k)(\alpha_k d_k) \\ &= \langle g(x_k + \alpha_k\theta_k d_k), \alpha_k d_k\rangle \\ &= \alpha_k \langle g(x_k + \alpha_k\theta_k d_k), d_k\rangle \end{aligned}$$

$$\alpha_k \langle g(x_k + \alpha_k\theta_k d_k), d_k\rangle \geq \alpha_k(1-\sigma)\langle g_k, d_k\rangle$$

luego,

$$\langle g(x_k + \alpha_k\theta_k d_k), d_k\rangle \geq (1-\sigma)\langle g_k, d_k\rangle$$

Por la desigualdad anterior y por la desigualdad de Cauchy-Schwartz, tenemos

$$\|g(x_k + \alpha_k\theta_k d_k) - g_k\| \|d_k\| \geq \langle g(x_k + \alpha_k d_k) - g_k, d_k\rangle \geq -\sigma\langle g_k, d_k\rangle$$

Dado que g *es Lipschitz continua, tenemos*

$$\alpha_k L\|d_k\|^2 \geq \alpha_k\theta_k L\|d_k\|^2 \geq \|g(x_k + \alpha_k\theta_k d_k) - g_k\| \|d_k\|$$

así,

$$\alpha_k L\|d_k\|^2 \geq -\sigma\langle g_k, d_k\rangle$$

$$\alpha_k \geq -\frac{\sigma\langle g_k, d_k\rangle}{L\|d_k\|^2} \tag{A.22}$$

Ahora, por Algoritmo (e) y (22), tenemos

$$\begin{aligned} \frac{f_{k+1} - f_k}{\alpha_k\langle g_k, d_k\rangle} \geq \sigma &\Longleftrightarrow f_{k+1} - f_k \leq \alpha_k\sigma\langle g_k, d_k\rangle \\ &\Longleftrightarrow f_k - f_{k+1} \geq -\alpha_k\sigma\langle g_k, d_k\rangle \end{aligned}$$

$$\begin{aligned} \alpha_k \geq -\frac{\sigma\langle g_k, d_k\rangle}{L\|d_k\|^2} &\Longleftrightarrow \alpha_k\sigma\langle g_k, d_k\rangle \leq -\frac{\sigma^2}{L}\left(\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2 \\ &\Longleftrightarrow -\alpha_k\sigma\langle g_k, d_k\rangle \geq \frac{\sigma^2}{L}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2 \end{aligned}$$

luego,

$$f_k - f_{k+1} \geq \frac{\sigma^2}{L}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2$$

luego, $\{f_k\}$ es una sucesión monótona decreciente, y por (H1) obtenemos que $\{f_k\}$ tiene límite; así

$$\lim_{k\to\infty}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2 = 0$$

Para el Algoritmo (d),tenemos

$$\begin{aligned}\langle g_{k+1}, d_k\rangle \geq \beta\langle g_k, d_k\rangle &\iff \langle g_{k+1}, d_k\rangle - \langle g_k, d_k\rangle \geq \beta\langle g_k, d_k\rangle - \langle g_k, d_k\rangle\\ &\iff \langle g_{k+1} - g_k, d_k\rangle \geq -(1-\beta)\langle g_k, d_k\rangle\end{aligned}$$

Por (H2), la desigualdad de Cauchy-Schwartz, se cumple, entonces:

$$\alpha_k L\|d_k\|^2 \geq \|g_{k+1} - g_k\|\|d_k\| \geq \langle g_{k+1} - g_k, d_k\rangle$$

luego,

$$\alpha_k L\|d_k\|^2 \geq -(1-\beta)\langle g_k, d_k\rangle$$

$$\alpha_k \geq -\frac{(1-\beta)\langle g_k, d_k\rangle}{L\|d_k\|^2}$$

$$\alpha_k\sigma\langle g_k, d_k\rangle \leq -\frac{\sigma(1-\beta)}{L}\left(\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2$$

$$-\alpha_k\sigma\langle g_k, d_k\rangle \geq \frac{\sigma(1-\beta)}{L}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2 \tag{A.23}$$

Luego por (10) y (23), se tiene

$$f_k - f_{k+1} \geq \frac{\sigma(1-\beta)}{L}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2$$

luego actuando de manera análoga, concluimos

$$\lim_{k\to\infty}\left(-\frac{\langle g_k, d_k\rangle}{\|d_k\|}\right)^2 = 0$$

A.5. Tasa de Convergencia

En la sección 3, hemos probado que si (H1), (H2) y (3) se cumplen, entonces (13) se satisface para todos los siete algoritmos y existe un entero positivo η_0 tal que,

$$f_k - f_{k+1} \geq \eta_0 \left(-\frac{\langle g_k, d_k \rangle}{\|d_k\|} \right)^2, \qquad \forall k$$

En este orden de ideas analizamos la Tasa de Convergencia. Restringimos nuestra discusión al caso de las funciones objetivo uniformemente convexas.
Supondremos

(H3) f es dos veces diferenciable continua y uniformente convexa en $\mathbb{E}$

Proposición 1 *Si f es una función que posee segunda derivada definida positiva, entonces f es uniformente convexa.*

Prueba 10 *Sea $X \in \mathbb{B}$. Por hipótesis se tiene que f'' es definida positiva entonces para cada $Z \in \mathbb{B}$ $\langle Z, f''(X)Z \rangle > 0$ por la propiedad arquimediana entonces existen M, N números naturales tal que $\|Z\| < M \langle Z, f''(X)Z \rangle$ y $\langle Z, f''(X)Z \rangle < N \|Z\|$ organizando lo anterior se obtiene*

$$\frac{1}{M} \|Z\| < \langle Z, f''(X)Z \rangle < N \|Z\|$$

A.6. Resultados y discusión

A continuación se dan algunos ejemplos de la teoría previamente mencionada.

Ejemplo 2 *Sean $A_{n\times n}, B_{n\times n}$ matrices definidas positivamente y simétricas tal que $\Sigma = A + B$. Sea $\mathbb{X} = \{X : X_{n\times m}\}$ el espacio de matrices rectangulares dotado de la norma $\|X\|_B = traza\left(X^T BX\right)$. Se construye la función $f : \mathbb{X} \to \mathbb{R}$ definida por $f(X) = traza\left(X^T AX\right)$*

Proposición 2 *La función f definida previamente es convexa*

Prueba 11 *Sean $X_1, X_2 \in \mathbb{X}$ y sea α tal que $0 \leq \alpha \leq 1$ consideremos*

$$f(\alpha X_1 + (1-\alpha)X_2) = traza((\alpha X_1 + (1-\alpha)X_2)^T A(\alpha X_1 + (1-\alpha)X_2))$$

$$\leq \alpha^2 f(X_1) + (1-\alpha)X_2 - 2\alpha(1-\alpha)traza(X_1^T A X_2)$$

propiedades de la traza y distributividad del producto matricial.

$$\leq \alpha^2 f(X_1) + (1-\alpha)^2 f(X_2) - 2\alpha(1-\alpha)traza(X_1^T(\Sigma - B)X_2)$$

Pero $traza(X_1^T \Sigma X_2) > traza(X_1^T B X_2)$ *entonces* $-2\alpha(1-\alpha)traza(X_1^T(\Sigma - B)X_2) < 0$ *luego se obtiene*

$$f(\alpha X_1 + (1-\alpha)X_2) \leq \alpha^2 f(X_1) + (1-\alpha)^2 f(X_2)$$

por hipótesis $0 \leq \alpha 1$ *entonces* $\alpha^2 \leq \alpha$ *y* $(1-\alpha)^2 \leq (1-\alpha)$ *lo cual garantiza finalmente que* f *sea convexa.*

Proposición 3 *La funcion* f *definida en el ejemplo es uniformemente convexa*

Prueba 12 *Basta demostrar que la función posee Hessiano definido positivamente, note que* $f'(X) = AX + A^T X$ *además* $f''(X) = A + A^T$ *lo cual es definido positivamente por hipótesis.*

Apéndice B

Bases de datos usadas

Si no chocamos contra la razón nunca llegaremos a nada.
Albert Einstein (1879-1955).

B.1. Bases Benchmark

La siguiente tabla presenta una descripción del desempeño de las bases de datos utilizadas para evaluar cual de los métodos de reducción de dimensionalidad ponderados de los mejores resultados. En la primera columna se encuentre el número correspondiente de la base que se encuentra en [86], y en las columnas dos a siete se ubica el desempeño y error de clasificación del método de reducción. Se observa que el método LPCA1 fue el que obtuvo en muchos casos el menor error y en varios casos se obtiene 1 mientras que el método WRDA ocupó el segundo lugar.