



MÉTODOS DE PRIMER ORDEN?

ANÁLISIS DE CONVERGENCIA??

Universidad de Buenos Aires

Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura
Director de Tesis: Dr. Pablo Amster
Septiembre 2018 – version 0.1

ABSTRACT

Aca va a ir el abstract cuando lo tengamos

*We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.*

— **knuth:1974** [knuth:1974]

AGRADECIMIENTOS

Agradecimientos para todos

CONTENTS

I	Introducción	1
1	INTUICIÓN	3
II	El teorema y aplicaciones	7
2	TEOREMA DE LA VARIEDAD ESTABLE Y LOS PUNTOS FI- JOS INESTABLES	9
2.1	Resultados previos	9
2.2	Puntos fijos inestables	9
3	APLICACIONES	13
3.1	Gradient Descent	13
3.2	Punto Próximo	13
3.3	Descenso por coordenadas	14
III	Apéndice	17
A	APÉNDICE	19

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRÓNIMOS

Part I

Introducción

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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $f(x) = \frac{1}{2}x^T H x$ con $H = \mathbf{diag}(\lambda_1, \dots, \lambda_n)$; supongamos además que $\lambda_1, \dots, \lambda_k > 0$ y $\lambda_{k+1}, \dots, \lambda_n < 0$.

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

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

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:

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

Por lo tanto por inducción es fácil probar que:

$$x_{t+1} = (1 - \alpha \lambda_i)^t \langle x_0, e^i \rangle e^i$$

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 concluimos que:

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

Finalmente, si $k < n$ entonces concluimos 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:

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

De lo que vemos que los puntos críticos son:

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

Y del criterio del Hessiano concluimos 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 1.0.1 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:

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

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

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

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 $\frac{dg_y}{dy}$ en los otros intervalos podemos concluir que:

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

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

Part II

El teorema y aplicaciones

En esta parte vamos a demostrar el resultado principal referido a la convergencia a mínimos de los diferentes algoritmos de primer orden usados en Machine Learning

TEOREMA DE LA VARIEDAD ESTABLE Y LOS PUNTOS FIJOS INESTABLES

RESULTADOS PREVIOS

Por el resto del documento, $g : \chi \rightarrow \chi$ y χ es una d -variedad sin borde.

Esto quizás debería ir en prerequisites cuando lo tengamos

Definición Dada una variedad de dimensión d χ y el espacio de medida $(\mathbb{R}^d, \mathcal{B}, \mu)$, decimos que $E \subset \chi$ tiene *medida cero* si existe un atlas $\mathcal{A} = \{U_i, \phi^i\}_{i \in \mathbb{N}}$ tal que $\mu(\phi^i(E \cap U_i)) = 0$. En este caso usamos el abuso de notación $\mu(E) = 0$.

Lema 2.1.1 Sea $E \subset \chi$ tal que $\mu(E) = 0$; si $\det(Dg(x)) \neq 0$ para todo $x \in \chi$, luego $\mu(g^{-1}(E)) = 0$

Demostración Sea $h = g^{-1}$ y (V_i, ψ^i) una colección de cartas en el dominio de g , si verificamos que $\mu(h(E) \cap V_i) = 0$ para todo $i \in \mathbb{N}$ entonces:

$$\mu(h(E)) = \mu\left(\bigcup_{i \in \mathbb{N}} h(E) \cap V_i\right) \leq \sum_{i \in \mathbb{N}} \mu(h(E) \cap V_i) = 0$$

Sin pérdida de generalidad podemos asumir que $h(E) \subseteq V$ con $(V, \psi) \in \{(V_i, \psi^i)\}$ una carta determinada. Sea $\mathcal{A} := \{(U_i, \phi^i)\}$ un atlas de χ y notemos $E_i = E \cap U_i$; luego $E = \bigcup_{i \in \mathbb{N}} E_i = \bigcup_{i \in \mathbb{N}} \phi^{i-1} \circ \phi^i(E_i)$ por lo que:

$$\begin{aligned} \mu(\psi \circ h(E)) &= \mu\left(\psi \circ h\left(\bigcup_{i \in \mathbb{N}} \phi^{i-1} \circ \phi^i(E_i)\right)\right) \\ &\leq \sum_{i \in \mathbb{N}} \mu\left(\psi \circ h \circ \phi^{i-1}\left(\phi^i(E_i)\right)\right) \end{aligned}$$

Por hipótesis $\phi^i(E_i)$ es de medida cero, luego como g es difeomorfismo local por ?? entonces $\psi \circ h \circ \phi^{i-1} \in C^1$. Como si $f \in C^1(\mathbb{R}^d)$ entonces es localmente Lipschitz, ergo f preserva la medida, concluimos que $\mu(\psi \circ h \circ \phi^{i-1}(\phi^i(E_i))) = 0$ para todo $i \in \mathbb{N}$. ■

Uso Teorema de la funcion inversa en variedades y que localmente Lipschitz preserva medida

PUNTOS FIJOS INESTABLES

Definición Sea:

$$\mathcal{A}_g^* := \left\{ x : g(x) = x \quad \max_i |\lambda_i(Dg(x))| > 1 \right\}$$

El conjunto de puntos fijos de g cuyo diferencial en ese punto tiene algún autovalor mayor que 1. A este conjunto lo llamaremos el conjunto de *puntos fijos inestables*

Este teorema debería ir en prerequisites

Teorema 2.2.1 Sea x^* un punto fijo de $g \in C^r(\chi)$ un difeomorfismo local. Supongamos que $E = E_s \oplus E_u$ donde

$$\begin{aligned} E_s &= \langle \{v_i / Dg(x^*)v_i = \lambda_i v_i, \lambda_i \leq 1\} \rangle \\ E_u &= \langle \{v_i / Dg(x^*)v_i = \lambda_i v_i, \lambda_i > 1\} \rangle \end{aligned}$$

Entonces existe $W_{loc}^{cs} \hookrightarrow \chi$ un embedding C^r local tangente a E_s en x^* llamado la variedad local estable central que cumple que existe $B \ni x^*$ entorno tal que $g(W_{loc}^{cs}) \cap B \subseteq W_{loc}^{cs}$ y $\bigcap_{k \in \mathbb{N}} g^{-k}(B) \subseteq W_{loc}^{cs}$

Con todos estos resultados demostremos el teorema principal:

Teorema 2.2.2 Sea $g \in C^1(\chi)$ tal que $\det(Dg(x)) \neq 0$ para todo $x \in \chi$, luego el conjunto de puntos iniciales que convergen por g a un punto fijo inestable tiene medida cero, i. e.:

$$\mu \left(\left\{ x_0 : \lim_k g^k(x_0) \in \mathcal{A}_g^* \right\} \right) = 0$$

Demostración Para cada $x^* \in \mathcal{A}_g^*$ por 2.2.2 existe B_{x^*} un entorno abierto; es más, $\bigcup_{x^* \in \mathcal{A}_g^*} B_{x^*}$ forma un cubrimiento abierto del cual existe un subcubrimiento numerable pues X es variedad, i. e.

$$\bigcup_{x^* \in \mathcal{A}_g^*} B_{x^*} = \bigcup_{i \in \mathbb{N}} B_{x_i^*}$$

Usamos que en una variedad se cumple la propiedad de Lindeloff

Primero si $x_0 \in \chi$ sea:

$$\begin{aligned} x_k &= g^k(x_0) \\ &= \underbrace{g \circ \dots \circ g}_{k \text{ veces}}(x_0) \end{aligned}$$

la sucesión del flujo de g evaluado en x_0 , entonces si $W := \left\{ x_0 : \lim_k x_k \in \mathcal{A}_g^* \right\}$ queremos ver que $\mu(W) = 0$.

Sea $x_0 \in W$, luego como $x_k \rightarrow x^* \in \mathcal{A}_g^*$ entonces existe $T \in \mathbb{N}$ tal que para todo $t \geq T$, $x_t \in \bigcup_{i \in \mathbb{N}} B_{x_i^*}$ por lo que $x_t \in B_{x_i^*}$ para algún

$x_i^* \in \mathcal{A}_g^*$ y $t \geq T$. Afirмо que:

Lema 2.2.3 $x_t \in \bigcap_{k \in \mathbb{N}} g^{-k}(B_{x_i^*})$ para todo $t \geq T$

Pablo: Hace falta demostrar esto??

Si notamos $S_i \triangleq \bigcap_{k \in \mathbb{N}} g^{-k}(B_{x_i^*})$, entonces por 2.2.1 sabemos por un lado que es una subvariedad de W_{loc}^{cs} y por el otro que $\dim(S_i) \leq \dim(W_{loc}^{cs}) = \dim(E_s) < d - 1$ ¹; por lo que $\mu(S_i) = 0$.

Finalmente como $x_T \in S_i$ para algún T entonces $x_0 \in \bigcup_{k \in \mathbb{N}} g^{-k}(S_i)$ por lo que $W \subseteq \bigcup_{i \in \mathbb{N}} \bigcup_{k \in \mathbb{N}} g^{-k}(S_i)$. Concluimos:

$$\begin{aligned} \mu(W) &\leq \mu\left(\bigcup_{i \in \mathbb{N}} \bigcup_{k \in \mathbb{N}} g^{-k}(S_i)\right) \\ &\leq \sum_{i \in \mathbb{N}} \sum_{k \in \mathbb{N}} \mu(g^{-k}(S_i)) \\ &\stackrel{2.1.1}{=} 0 \end{aligned}$$

■

Para finalizar veamos un caso simple que nos encontraremos seguido:

Corolario 2.2.4 *Bajo las mismas hipótesis que en 2.2.2 si agregamos que $\chi^* \subseteq \mathcal{A}_g^*$ entonces $\mu(W_g) = 0$*

Demostración Como $\chi^* \subseteq \mathcal{A}_g^*$ entonces $W_g \subseteq W$, luego $\mu(W_g) \leq \mu(W) \stackrel{2.2.2}{=} 0$. ■

Usamos que la dimension de la variedad es la de su tangente

Usamos que una subvariedad de dimension menor tiene medida 0

¹ Por que???

APLICACIONES

GRADIENT DESCENT

Como una aplicación del teorema en 2.2.2 demostremos que *gradient descent* tiene probabilidad cero de converger a puntos silla. Consideremos *gradient descent* con *learning rate* α :

$$x_{k+1} = g(x_k) \triangleq x_k - \alpha \nabla f(x_k) \quad (2)$$

Hipótesis 1 Asumamos que $f \in \mathcal{C}^2$ y $\|\nabla^2 f(x)\|_2 \leq L$

Proposición 3.1.1 *Todo punto silla estricto de f es un punto fijo inestable de g , i. e. $\chi^* \subseteq \mathcal{A}_g^*$.*

Demostración Es claro que un punto crítico de f es punto fijo de g ; si $x^* \in \chi^*$ entonces $Dg(x^*) = Id - \alpha \nabla^2 f(x^*)$ y entonces los autovalores de Dg son $\{1 - \alpha \lambda_i : \lambda_i \in \{\mu : \nabla^2 f(x^*)v = \mu v \text{ para algún } v \neq 0\}\}$. Como $x^* \in \chi^*$ existe $\lambda_{j^*} < 0$ por lo que $1 - \alpha \lambda_{j^*} > 1$; concluimos que $x^* \in \mathcal{A}_g^*$. ■

Usamos que $f(A)$
tiene autovalores
 $f(\{\lambda_i\})$

Proposición 3.1.2 *Bajo 3.1 y $\alpha < \frac{1}{L}$ entonces $\det(Dg(x)) \neq 0$.*

Demostración Como ya sabemos $Dg(x) = Id - \alpha \nabla^2 f(x)$ por lo que:

$$\det(Dg(x)) = \prod_{i \in \{1, \dots, d\}} (1 - \alpha \lambda_i)$$

Luego por 3.1 tenemos que $\alpha < \frac{1}{|\lambda_i|}$ y entonces $1 - \alpha \lambda_i > 0$ para todo $i \in \{1, \dots, d\}$; concluimos que $\det(Dg(x)) > 0$. ■

Corolario 3.1.3 *Gradient descent converge a mínimos Sea g dada por Gradient descent en 2, bajo 3.1 y $\alpha < \frac{1}{L}$ se tiene que $\mu(W_g) = 0$.*

Demostración Por 3.1.1 y 3.1.2 tenemos que vale 2.2.4 y concluimos que $\mu(W_g) = 0$. ■

PUNTO PRÓXIMO

El algoritmo de punto próximo esta dado por la iteración:

$$x_{k+1} = g(x_k) \triangleq \arg \min_{z \in \mathcal{X}} f(z) + \frac{1}{2\alpha} \|x_k - z\|_2^2 \quad (3)$$

Proposición 3.2.1 *Bajo 3.1 y $\alpha < \frac{1}{L}$ entonces vale:*

$$1. \det(Dg(x)) \neq 0$$

$$2. \chi^* \subseteq \mathcal{A}_g^*$$

Probamos esto? Me parece un poco claro

Demostración Veamos primero el siguiente lema:

Lema 3.2.2 *Bajo 3.1, $\alpha < \frac{1}{L}$ y $x \in \chi$ entonces $f(z) + \frac{1}{2\alpha} \|x - z\|_2^2$ es estrictamente convexa, por lo que $g \in \mathcal{C}^1(\chi)$*

Por lo tanto por 3.2.2 podemos tomar límite, i. e.

$$\begin{aligned} x_{k+1} &= g(x_k) = \arg \min_{z \in \chi} f(z) + \frac{1}{2\alpha} \|x_k - z\|_2^2 \\ \downarrow \quad \quad \downarrow \quad \quad \quad \downarrow \\ x &= g(x) = \arg \min_{z \in \chi} f(z) + \frac{1}{2\alpha} \|x - z\|_2^2 \\ \iff \nabla_z \left(f(z) + \frac{1}{2\alpha} \|x - z\|_2^2 \right) (g(x)) &= 0 \\ \iff \nabla f(g(x)) - \frac{1}{\alpha} (x - g(x)) &= 0 \\ \iff g(x) + \alpha \nabla f(g(x)) &= x \end{aligned}$$

Finalmente por diferenciación implícita obtenemos:

$$\begin{aligned} Dg(x) + \alpha \nabla^2 f(g(x)) Dg(x) &= Id \\ \implies Dg(x) &= (Id + \alpha \nabla^2 f(g(x)))^{-1} \end{aligned}$$

Luego si $x^* \in \chi^*$ entonces $Dg(x^*) = (Id + \alpha \nabla^2 f(x^*))^{-1}$ y tiene autovalores $\left\{ \frac{1}{1 + \alpha \lambda_i} \right\}$ con λ_i autovalores de $\nabla^2 f(x^*)$. Por lo tanto $x^* \in \mathcal{A}_g^*$ y para $\alpha < \frac{1}{L}$ se tiene que $\det(Dg(x)) \neq 0$. ■

Corolario 3.2.3 *Sea g dado por el algoritmo de punto próximo con ecuación 3, bajo 3.1 y $\alpha < \frac{1}{L}$ se tiene que $\mu(W_g) = 0$.*

Demostración Por 3.2.1 tenemos que vale 2.2.4 y concluimos que $\mu(W_g) = 0$. ■

DESCENSO POR COORDENADAS

Consideremos el algoritmo 1, luego si definimos $g_i(x) = x - \alpha \left(0, \dots, 0, \frac{\partial f}{\partial x_i}(x), 0, \dots, 0 \right)$ entonces:

Lema 3.3.1 *La iteración de Descenso por coordenadas esta dada por:*

$$x_{k+1} = g(x_k) \triangleq g_d \circ g_{d-1} \circ \dots \circ g_1(x) \quad (4)$$

Algoritmo 1 : Descenso por coordenadas

```

1 Input:  $f \in C^1$ ,  $\alpha > 0$ ,  $x_0 \in \chi$ 
2 for  $k \in \mathbb{N}$  do
3   for index  $i = 1, \dots, d$  do
4      $y_k^0 = x_k$  y  $y_k^i = (x_{k+1}^1, \dots, x_{k+1}^i, x_k^{i+1}, \dots, x_k^d)$ 
5      $x_{k+1}^i \leftarrow x_k^i - \alpha \frac{\partial f}{\partial x_i}(y_k^{i-1})$ 
6   end
7 end

```

Observación Sea $f \in C^2$ y supongamos que $\max_{i \in \{1, \dots, d\}} |e_i^T \nabla^2 f(x) e_i| \leq L_{\max}$

Lema 3.3.2 Si g está dada por 4 entonces:

$$Dg(x_k) = \prod_{j \in \{1, \dots, d\}} \left(Id - \alpha e_{d-j+1} e_{n-j+1}^T \nabla^2 f(y_k^{n-j}) \right) \quad (5)$$

Demostración Notemos primero que:

$$Dg_i(x) = Id - \alpha \begin{pmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ \vdots & & \frac{\partial^2 f}{\partial x_i^2}(x) & & \vdots \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{pmatrix} = Id - \alpha e_i e_i^T \nabla^2 f(x)$$

Por lo tanto:

$$\begin{aligned}
Dg(x_k) &= D(g_d \circ \dots \circ g_1)(x_k) \\
&= \left(Id - \alpha e_d e_d^T \nabla^2 f \right) \left(\underbrace{g_{d-1} \circ \dots \circ g_1(x_k)}_{y_k^{d-1}} \right) D(g_{d-1} \circ \dots \circ g_1)(x_k) \\
&\vdots \\
&= \prod_{j \in \{1, \dots, d\}} \left(Id - \alpha e_{d-j+1} e_{n-j+1}^T \nabla^2 f(y_k^{n-j}) \right)
\end{aligned}$$

■

Proposición 3.3.3 Bajo 7 y $\alpha < \frac{1}{L_{\max}}$ se tiene que $\det(Dg(x)) \neq 0$

Demostración Basta probar que cada término de 5 es invertible, para eso:

$$\begin{aligned}
 \chi_{Dg_i(x)}(\lambda) &= \det \left(\lambda Id_d - Id_d - \alpha e_i e_i^T \nabla^2 f(x) \right) \\
 &= \det \begin{pmatrix} \lambda - 1 & 0 & \dots & 0 & \dots & 0 \\ \vdots & & & \vdots & & \vdots \\ 0 & \dots & (\lambda - 1) + \alpha \frac{\partial^2 f}{\partial x_i^2}(x) & & & \vdots \\ 0 & \dots & & & & 0 \end{pmatrix} \\
 &= (\lambda - 1)^{d-1} \left(\lambda - 1 + \alpha \frac{\partial^2 f}{\partial x_i^2}(x) \right)
 \end{aligned}$$

Luego si $\alpha < \frac{1}{L_{max}}$ entonces $\lambda - 1 + \alpha \frac{\partial^2 f}{\partial x_i^2}(x) > 0$ para todo $i \in \{1, \dots, d\}$ por lo que todos los autovalores son positivos y $Dg_i(x)$ es invertible para todo i . ■

Part III

Apéndice



APÉNDICE
