UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

GEOMETRÍA DE SISTEMAS DE DESCENSO: ESTUDIO ASINTÓTICO MEDIANTE DESINGULARIZACIÓN

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN MATEMÁTICAS APLICADAS
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

ROBERTO JAVIER BOBADILLA SOLARI

PROFESOR GUÍA:
ARIS DANIILIDIS

MIEMBROS DE LA COMISIÓN:

ABDERRAHIM HANTOUTE
FETHI MAHMOUDI

Este trabajo ha sido parcialmente financiado por FONDECYT 1130176

SANTIAGO DE CHILE
2016

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO
POR: ROBERTO JAVIER BOBADILLA SOLARI
FECHA: 2016
PROF. GUÍA: SR. ARIS DANIILIDIS

GEOMETRÍA DE SISTEMAS DE DESCENSO: ESTUDIO ASINTÓTICO MEDIANTE DESINGULARIZACIÓN

Los sistemas de tipo gradiente son relevantes como sistemas dinámicos en sí y además sirven como marco teórico para estudiar algoritmos de optimización, en particular algoritmos de descenso. Relacionado con este último aspecto, es natural preguntarse si las órbitas tienen longitud finita y convergen, cuando están en un conjunto acotado.
El presente trabajo presenta respuestas a tales preguntas, bajo suposiciones especiales pero no restringidas en la práctica: se adoptará el marco de la geometría o-minimal que permite establecer resultados pertinentes sobre el comportamiento de las órbitas en torno a los puntos críticos. Como se verá a coninuación, una función suave f definible en una estructura o-minimal satisface la llamada desigualdad de Kurdyka-Łojasiewicz: en torno a cualquier valor crítico se acotan los gradientes de f inferiormente por una constante. Dicho resultado se adapta en el caso no suave (siempre gracias a las herramientas de la geometría o-minimal) y se obtiene una cota análoga válida uniformemente para la norma de los subgradientes de f.

A grandes rasgos el resultado de Kurdyka-Łojasiewicz consiste en encontrar una función auxiliar (la función desingularizante) estrictamente monótona y suave, de forma que por una parte el sistema gradiente (o bien subgradiente) inducido por la composición de dicha función con f tiene las mismas órbitas, y por otra parte los gradientes (o subgradientes) de dicha composición están acotados inferiormente por una constante. Este proceso es llamado desingularización de la función f, cuya potencia se aprecia explícitamente mediante la parametrización de las trayectorias a través de los niveles de la función f.

Por último existe un resultado similar para multiaplicaciones definibles, donde se desingulariza la coderivada, en un sentido que se determinará más adelante. En este caso el sistema dinámico de estudio ya no es un sistema de tipo gradiente o subgradiente, sino que es un *sweeping process* (véase [7]). Se muestra que si dicho *sweeping process* proviene de una función definible y continua, entonces mediante la desingularización de su coderivada se recuperan los resultados anteriores. En particular se pondrá en evidencia la relación entre la desingularización del *sweeping process* y la desingularización de la función f que lo define.

Tabla de Contenido

Introducción

El presente trabajo de tesis tiene como objetivo, a grandes rasgos, el estudio de sistemas de tipo gradiente:

$$\dot{x}(t) = -\nabla f(x(t)).$$

El sistema anterior se verá generalizado, para poder abordar el caso en que f no es una función diferenciable. Lo que dará lugar al sistema subgradiente siguiente:

$$\dot{x}(t) \in -\partial f(x(t)).$$

Un paso más allá será transformar el sistema de arriba en un *sweeping process*.

$$\dot{x}(r) \in -N_{S(r)}(x(r)),$$

y estudiar este último de forma que se recuperen las conclusiones de los casos anteriores. Para dicho estudio se presentan elementos de la teoría de geometría o-minimal que será, mayoritariamente, el marco de referencia para los resultados de esta memoria. Más precisamente el interés será en torno a clases de funciones, que son relativas a las estructuras de la teoría o-minimal y sobre estas se obtendrán potentes resultados.
Por otro lado se expondrán conceptos básicos de análisis variacional, los que son indispensables para la mayoría de los resultados a tratar a lo largo de este trabajo.

Seguido de lo anterior se mostrarán resultados de Kurdyka [6] que responden a preguntas relativas al sistema gradiente clásico $\dot{x}(t) = -\nabla f(x(t))$, donde f es una función de clase C^1 perteneciendo a una estructura o-minimal genérica. En dicho marco, el estudio asintótico se resuelve satisfactoriamente.
Luego, resultados de Daniilidis, Bolte y Lewis [1] extienden lo anterior al sistema no diferenciable $\dot{x}(t) \in -\partial f(x(t))$, manteniendo siempre la hipótesis que f pertenece a una estructura o-minimal. En [1] la estructura o-minimal fue una en particular, conocida como estructura de los conjuntos globalmente subanalíticos. En el presente documento se darán las demostraciones en el caso de una estructura o-minimal genérica, cabe mencionar que en algunos casos fue importante basarse en demostraciones de Kurdyka del caso diferenciable. Todo dentro del marco del análisis variacional.

Más aún se expondrán resultados de Daniilidis y Drusvyatskiy [5] los que llevan aún más allá las generalizaciones al sistema gradiente, para tratar el llamado *sweeping process*. Finalmente se resuelve un planteamiento dejado de estas generalizaciones del sistema gradiente.

Capítulo 1

Elementos de geometría o-minimal

El propósito del presente capítulo es dar una introducción a los conceptos básicos de geometría o-minimal.
Es digno de mencionar que esta teoría tiene variadas interacciones y/o aplicaciones en distintos campos, tales como sistemas dinámicos, análisis no diferenciable, lógica, ecuaciones en derivadas parciales, etc. Esto es debido, en parte, a que las estructuras o-minimales son sencillas para trabajar, puesto que son estables para muchas operaciones y proporcionan herramientas muy potentes. Se definirán pronto las llamadas funciones o-minimales, sobre las que se obtendrán varios resultados. Lo relevante es que, de cierta manera, dichas funciones presentan una clase suficientemente grande, lo que a su vez le da valor a los resultados obtenidos.

Antes de empezar cabe mencionar que la teoría o-minimal se puede desarrollar sobre un contexto algebraico más general, substituyendo $\mathbb{R}$ por otros cuerpos algebraicos más generales. Esto no aportará en este contexto, al momento de estudiar sistemas gradientes y generalizaciones, se hará en $\mathbb{R}^n$, por lo que bastará mantenerse en el marco del cuerpo real.

Definición 1.1 (Estructura o-minimal. Ver [3], [10] e.g.) *Una estructura o-minimal sobre* $\mathbb{R}$ *es una colección de familias de conjuntos* $(\mathcal{S}_n)_{n\in\mathbb{N}}$ *donde cada* $\mathcal{S}_n$ *es una colección de conjuntos de* $\mathbb{R}^n$*, que satisface los siguientes* 5 *axiomas:*

(i) *Los conjuntos algebraicos de* $\mathbb{R}^n$*, es decir, los ceros de un polinomio real, están en* $\mathcal{S}_n$*.*

(ii) *Para todo* $n \in \mathbb{N}$*, la familia* $\mathcal{S}_n$ *es cerrada para intersecciones finitas, uniones finitas y complemento.*

(iii) *Si* $A_m \in \mathcal{S}_m$ *y* $B \in \mathcal{S}_n$ *entonces* $A \times B \in \mathcal{S}_{n+m}$*.*

(iv) *Denotando la proyección* $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$*, se tiene que si* $A \in \mathcal{S}_{n+1}$ *entonces* $\pi(A) \in \mathcal{S}_n$*.*

(v) *Los conjuntos de la colección* $\mathcal{S}_1$ *son exactamente las uniones finitas de puntos e intervalos.*

A los conjuntos de la familia $\mathcal{S}_n$ se les dirá *definibles* en la estructura o-minimal, o simplemente o-minimales o incluso definibles.

Observación Cabe notar que la definición anterior, véase los puntos, (iii) y (iv), establece una relación entre conjuntos definibles de distintas dimensiones. En este sentido la estructura no es valiosa vista sobre $\mathbb{R}^n$, con n fijo, sino que debe ser coherente en toda la cadena $(\mathbb{R}^n)_n$. Esto tendrá un sentido más preciso, cuando se vea más adelante lo que se llama *descomposición cilíndrica definible en celdas* de un conjunto o-minimal.

Definición 1.2 (Funciones o-minimales. Ver [3]) *Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, donde A es un conjunto definible. La función f se dirá o-minimal (o definible) si su grafo, el conjunto:*

$$\{(x, f(x)) \,:\, x \in A\}$$

es definible en dicha estructura.

Como primer resultado de estabilidad de la estructura o-minimal, se tiene lo siguiente:

Teorema 1.3 (Ver [3]) *Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ definible y $B \subset A$ un definible de $\mathbb{R}^n$, entonces $f(B) \in \mathcal{S}_m$.*

Demostración. La idea detrás de la proposición anterior es utilizar la propiedad (iv) de la definición **1.1**, considerando el conjunto $C_1 = \{(x, y, z) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \,|\, x = z\}$, que es evidentemente definible, así al definir el conjunto $C_2 = C_1 \cap (\mathbb{R}^m \times \mathrm{gph}(f) \cap (\mathbb{R}^m \times B \times \mathbb{R}^m)$ también queda definible. Lo que hace C_1 es mantener iguales las primeras y últimas m coordenadas. De modo que al intersectarlo con C_2, tiene el efecto de copiar al principio las coordenadas finales.
Por otro lado, lo que hace la última intersección es restringir el grafo a los puntos con coordenadas en B. Por lo tanto las coordenadas finales son las imágenes de los puntos de B bajo f, y son replicadas en las primeras coordenadas del conjunto, de acuerdo al razonamiento anterior. Sigue que $\pi_{m+n+m,m}(C_2) = f(B)$. Cada uno de los conjuntos construidos aquí es definible por lo tanto $f(B)$ también lo es. $\square$

Mirando más atentamente la demostración anterior, lo que se hizo se reduce a mover de forma inteligente las coordenadas de un conjunto y ocupar los axiomas de las estructuras o-minimales. Utilizando esta idea se puede probar que pre-imágenes de conjuntos definibles bajo funciones definibles, o conjuntos que se obtienen de permutar sus coordenadas a partir de un conjunto definible, dan lugar a conjuntos definibles. La idea de que el derecho a clonar, permutar y borrar coordenadas se condensará próximamente en un teorema, y simplificará notablemente las demostraciones.

Antes de seguir con más definiciones y proposiciones cabe mencionar ejemplos de algunas estructuras o-minimales sobre $\mathbb{R}$, aunque de todos modos el desarrollo en los capítulos siguientes se hará sobre una estructura general. Dicho lo anterior, los conjuntos semialgebraicos, vistos a continuación, forman uno de los principales ejemplos sobre el cual basarse.
Las siguientes son estructuras o-minimales recurrentes en la literatura del tema (Ver [10]):

(i) Conjuntos semialgebraicos.
(ii) Conjuntos globalmente subanalíticos.
(iii) Conjuntos $(\mathbb{R}, exp)$-definibles.
(iv) Conjuntos $(\mathbb{R}_{an}, exp)$-definibles.
(v) Conjuntos $(\mathbb{R}_{an})$-definibles.

En lo que sigue se desarrollará brevemente la teoría de geometría semialgebraica como ejemplo particular.

Definición 1.4 (Conjunto semialgebraico. Ver [4]) *Se definen los conjuntos semialgebraicos de $\mathbb{R}^n$, SA_n, como la clase de conjuntos más pequeña que satisface los siguientes dos axiomas:*

(i) *Si p es un polinomio a n variables reales, entonces $\{x \in \mathbb{R}^n : p(x) = 0\}$ es un conjunto semialgebraico.*
(ii) *Si $A \in SA_n$ y $B \in SA_n$ entonces $A \cup B$, $A \cap B$ y A^c están en SA_n. Dicho de otro modo, la colección de los conjuntos semialgebraicos es cerrada para operaciones booleanas finitas de conjuntos.*

Proposición 1.5 (Ver [4]) *Todo subconjunto semialgebraico de $\mathbb{R}^n$ es unión finita de conjuntos semialgebraicos básicos de la forma:*

$$\{x \in \mathbb{R}^n \,|\, p(x) = 0,\, q_1(x) > 0,\, \dots,\, q_r(x) > 0\}$$

para $r \in \mathbb{N}$, q_i y p polinomios.

Demostración. Evidentemente tales conjuntos son semialgebraicos. Por otro lado la clase definida por uniones finitas de estos conjuntos contiene a los conjuntos algebraicos y es cerrada bajo operaciones finitas de conjuntos. El hecho de que dicha clase es cerrada para uniones finitas es evidente. Por otra parte las intersecciones agregan más polinomios como restricciones. Mientras que la operación de tomar complemento cambia restricciones del estilo $q(x) > 0$ por $q(x) < 0$ en unión a $q(x) = 0$. □

A continuación se comprueba que los conjuntos semialgebraicos forman una estructura o-minimal.

Observación Los conjuntos semialgebraicos de $\mathbb{R}$ son uniones finitas de puntos e intervalos abiertos. Esto se consecuencia directa de que son las únicas posibilidades para ceros y conjuntos obtenidos a partir de desigualdades con polinomios.

Proposición 1.6 *Si A y B están en SA_n y SA_m respectivamente entonces $A \times B \in SA_{nm}$.*

Demostración. Sean A un conjunto semialgebraico de $\mathbb{R}^n$ y B un conjunto semialgebraico de $\mathbb{R}^m$. Por la proposición **1.5**, A se puede escribir como unión finita de conjuntos de la forma:

$$\{x \in \mathbb{R}^n \,|\, p^A(x) = 0,\, q_1^A(x) > 0,\, \dots,\, q_{r(A)}^A(x) > 0\},$$

de igual modo B:

$$\{x \in \mathbb{R}^m \,|\, p^B(x) = 0,\, q_1^B(x) > 0,\, \dots,\, q_{r(B)}^B(x) > 0\}.$$

El producto cartesiano distribuye sobre una unión finita, por lo tanto basta ver el caso

$$A = \{x \in \mathbb{R}^n \,|\, p^A(x) = 0,\, q_1^A(x) > 0,\, \dots,\, q_{r(A)}^A(x) > 0\},$$

$$B = \{x \in \mathbb{R}^m \,|\, p^B(x) = 0,\, q_1^B(x) > 0,\, \dots,\, q_{r(B)}^B(x) > 0\}.$$

Escribiendo la definición el producto

$$\begin{aligned} A \times B = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \,|\, p^A(x) = 0,\, p^B(y) = 0,\, q_1^A(x) > 0,\, \dots,\, q_{r(A)}(x) > 0, \\ q_1^B(x) > 0,\, \dots,\, q_{r(B)}^B(x) > 0\} \end{aligned}$$

se ve claramente que este es un conjunto semialgebraico □

La propiedad anterior fue un ejercicio simple. De hecho la propiedad (iv) es la única no trivial para afirmar que los conjuntos semialgebraicos forman una estructura o-minimal. Eso es el objeto del llamado teorema de Tarski-Seidenberg, que se presentará a continuación en distintas versiones.

Teorema 1.7 (Tarski-Seidenberg - primera forma. Ver [4]) *Existe un algoritmo finito para el cual, dado un sistema de ecuaciones e inecuaciones polinomiales con variables $a = (a_1, \dots, a_k)$ y x a coeficientes reales, de la forma:*

$$S(a,x) = \begin{cases} S_1(a,x) & \triangleright_1 & 0 \\ S_2(a,x) & \triangleright_2 & 0 \\ & \dots & \\ S_r(a,x) & \triangleright_r & 0 \end{cases}$$

en donde $\triangleright_i$ puede ser $<, \leq, >, \geq$ ó $=$, este produce una colección finita $R_1(a), \dots, R_l(a)$ de sistemas de ecuaciones o inecuaciones polinomiales en a. Tales que para cada $b \in \mathbb{R}^k$ el sistema $S(b,x)$ tiene solución x real sí y solo sí alguno de los $R_j(b)$ se satisface.

Aunque el teorema anterior tendrá solo una aplicación inmediata en el teorema siguiente, este último será bastante fuerte. La idea en términos más sencillos es que la fórmula $\exists\, x\, S(a,x)$ resulta ser equivalente a la disyunción $\bigvee_{i=1}^k C_i(x)$, esto es llamado *eliminación de cuantificadores*. Por ejemplo $\exists\, x$ tal que $ax^2 + bx + c = 0$ es equivalente a:

$$(a \neq 0 \,\wedge\, b^2 - 4ac \geq 0) \;\vee\; (a = 0 \,\wedge\, b \neq 0) \;\vee\; (a = b = c = 0).$$

Teorema 1.8 (Tarski-Seidenberg - segunda forma. Ver [4]) *Sea A un conjunto semialgebraico de $\mathbb{R}^n$ y $\pi : \mathbb{R}^n \to \mathbb{R}^{n-1}$ la proyección canónica sobre las primeras $n-1$ coordenadas. Entonces $\pi(A)$ es un conjunto semialgebraico de $\mathbb{R}^{n-1}$.*

Demostración. Tal conjunto A es unión finita de conjuntos de la forma $\{x \in \mathbb{R}^n \,|\, p(x) = 0,\, q_1(x) > 0,\, \ldots,\, q_r(x) > 0\}$, lo que se reduce al problema de mostrar que la proyección de un conjunto de esa forma, es de nuevo un conjunto semialgebraico. Observando que

$$\pi(A) = \{(x_1, ..., x_{n-1} \in \mathbb{R}^{n1} \,:\, \exists\, x_n \in \mathbb{R}\ (x_1, ..., x_n) \in A\},$$

se nota que está escrito de la forma adecuada para aplicar la primera versión de Tarski-Seidenberg, y así concluir que existe un sistema de ecuaciones o inecuaciones polinomiales cuyo conjunto de soluciones es el mismo que $\pi(A)$, de lo que se deduce que este es semialgebraico. $\square$

Observación La demostración anterior muestra lo siguiente: dado un conjunto semialgebraico $A \subset \mathbb{R}^n$, sin importar cuales sean las coordenadas que uno elija, proyectar el conjunto A sobre cualquier subconjunto de las n coordenadas disponibles resulta en un conjunto semialgebraico si es que A lo es. En efecto se puede elegir otra variable distinta sobre la cual se aplica la primera forma de Tarski-Seidenberg, ya que no importa sobre cuál variable eliminar para mantener la propiedad de ser semialgebraico y luego iterar.

Definición 1.9 (Fórmula de primer orden, Ver [4]) *Se llama fórmula de primer orden a cualquier fórmula obtenida bajo las siguientes reglas:*

(i) *Si $p = p(x_1, ..., x_n)$ es un polinomio a n variables entonces $p = 0$ y $p > 0$ son fórmulas de primer orden.*
(ii) *Conjunción, disyunción y negación de fórmulas de primer orden son también fórmulas de primer orden, i.e. si ψ y ϕ son fórmulas de primer orden entonces también lo son $\psi \vee \phi$, $\psi \wedge \phi$ y ϕ.*
(iii) *Si ψ es una fórmula de primer orden y x es una variable real entonces las fórmulas $(\exists\, x)\psi$ y $(\forall\, x)\psi$ son también fórmulas de primer orden.*

Observación Respecto de la definición anterior se puede mencionar lo siguiente.

(i) Las fórmulas de (i) y (ii) son de orden cero, pues no contienen cuantificadores.
(ii) Fórmulas que involucren ecuaciones e inecuaciones con funciones semialgebraicas son de primer orden. Esto es consecuencia de la proposición **1.5**.

Ahora se puede enunciar la tercera forma del teorema de Tarski-Seidenberg.

Teorema 1.10 (Tarski-Seidenberg - tercera forma, Ver [4]) *Sea $\psi(x_1, ..., x_n)$ una fórmula de primer orden. Entonces las tuplas $(x_1, ..., x_n) \in \mathbb{R}^n$ tales que se satisface $\psi(x_1, ..., x_n)$ definen un conjunto semialgebraico.*

Observación Con la segunda forma de Tarski-Seidenberg y los otros resultados anteriores se concluye que la estructura semialgebraica es una estructura o-minimal, pero más aún, es la más pequeña posible. Esto último es consecuencia inmediata de su definición.

Siendo la clase de conjuntos semialgebraicos, la estructura o-minimal más pequeña posible, amerita que se muestren algunas familias de funciones semialgebraicas.

Proposición 1.11 *Sea $A \subset \mathbb{R}^n$. Las siguientes funciones $f : A \to \mathbb{R}^m$ son semialgebraicas:*

(i) *La función $f = (p_1, ..., p_m)$ donde p_{i} es un polinomio a n variables para $\mathrm{i} = 1, ..., m$.*

(ii) *La función $f = (q_1, ..., q_m)$ donde q_{i} es una función racional a n variables para $\mathrm{i} = 1, ..., m$.*

(iii) *Sea $C \subset \mathbb{R}^n$ un conjunto no vacío y semialgebraico. Entonces la función*

$$\mathrm{dist}(x, C) := \inf_{y\in C} \sqrt{\sum_{\mathrm{i}=1}^{n} (x_{\mathrm{i}} - y_{\mathrm{i}})^2}$$

es semialgebraica.

(iv) *Si f es semialgebraica y positiva en $A \subset \mathbb{R}^n$ entonces también lo es $\sqrt{f}$.*

(v) *La función $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ es semialgebraica.*

Demostración. En esta demostración se verá lo útil de la tercera forma de Tarski-Seidenberg.

(i) Basta escribir

$$\mathrm{gph}(f) = \{(x, y) \in A\times B \, : \, y = f(x)\} = \{(x, y) \in A\times B \, : \, 0 = y_{\mathrm{i}} - p_{\mathrm{i}}(x)\, , \; \mathrm{i} = 1, ..., m\}.$$

Como $p_{\mathrm{i}} - y_{\mathrm{i}}$ es un polinomio, conjunto anterior es semialgebraico.

(ii) En este caso el conjunto

$$\mathrm{gph}(f) = \left\{(x, y) \in A \times B \, : \, y = \frac{p(x)}{q(x)}\right\},$$

donde p y q son polinomios, al ser reescrito de la forma

$$\mathrm{gph}(f) = \{(x, y) \in A \times B \, : \, q(x)y = p(x)\},$$

se ve nuevamente que aparece una ecuación polinomial por lo que el conjunto es semialgebraico.

(iii) El conjunto

$$\mathrm{gph}(\mathrm{dist}(x, C)) = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} \, : \, r = \mathrm{dist}(x, C)\}$$

se puede escribir como sigue:

$$\left\{(x, r) \in \mathbb{R}^n \times \mathbb{R} \, : \, \left(\forall\, y \in C, \quad r \le \sqrt{\sum_{\mathrm{i}=1}^{n} (x_{\mathrm{i}} - y_{\mathrm{i}})^2}\right) \wedge \right.$$
$$\left. \left(\forall\, y \in C, \, s \in \mathbb{R} \quad s \le \sqrt{\sum_{\mathrm{i}=1}^{n} (x_{\mathrm{i}} - y_{\mathrm{i}})^2} \implies s \le r\right)\right\}$$

lo que está adentro es una fórmula de primera orden, por lo tanto el conjunto es semialgebraico gracias a la tercera forma de Tarski-Seidenberg.

(iv) Se observa que $\operatorname{gph}(\sqrt{f}) = \{(x,y) \in A \times \mathbb{R}^n \,:\, y^2 = f(x)\}$, está definido por una fórmula de primer orden con objetos semialgebraicos.

(v) Es consecuencia de (iv) y que $\|x\| = \operatorname{dist}(x, \{0\})$.

$\square$

A continuación se mencionan las siguientes propiedades:

Proposición 1.12 *Sea $A \subset \mathbb{R}^n$.*

(i) *Composición de funciones semialgebraicas es da lugar a una función semialgebraica.*

(ii) *Si $f = (f_1, ..., f_k) : A \subset \mathbb{R}^n \to \mathbb{R}^k$, entonces f es semialgebraica sí y solo sí las funciones coordenadas lo son.*

(iii) *Si $f : A \to \mathbb{R}^k$ es invertible y semialgebraica entonces su inversa es semialgebraica.*

(iv) *Más aún, si $f : U \subset \mathbb{R}^n \to f(A) \subset \mathbb{R}^k$, es diferenciable y semialgebraica, entonces sus derivadas parciales y por ende gradiente son funciones semialgebraicas.*

Demostración. La demostración de estas propiedades no es más que definir apropiadamente fórmulas de primer orden, involucrando exclusivamente objetos semialgebraicos.

(i) Sean $g : A \subset \mathbb{R}^k \to \mathbb{R}^n$ y $f : B \subset g(A) \to \mathbb{R}^m$ ambas funciones semialgebraicas. El conjunto

$$\Big(\operatorname{gph}(g) \times \mathbb{R}^m\Big) \cap \Big(A \times \operatorname{gph}(f)\Big) = \{(x,y,z) \,:\, x \in A,\, y = g(x),\, z = f(y)\}$$

es semialgebraico y resulta que $\operatorname{gph}(f \circ g)$ es la eliminación de la variable y por proyección del conjunto anterior, lo que muestra que $f \circ g$ es semialgebraica.

(ii) Es directo de escribir la fórmula:

$$\operatorname{gph}(f) = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \,:\, f_1(x) = y_1,\, ...,\, f_k(x) = y_k\}.$$

(iii) Es evidente que $\operatorname{gph}(f^{-1}) = \{(x,y)\mathbb{R}^m \times \mathbb{R}^n : f(x) = y\}$ es semialgebraico, la fórmula que lo define es de primer orden con f semialgebraica.

(iv) La clave es presentar el ínfimo como fórmula de primer orden:

$$\begin{aligned}
\operatorname{gph}\Big(\frac{\partial f}{\partial x_{\mathrm{i}}}\Big) &= \Big\{(x,v) \,:\, v = \frac{\partial f}{\partial x_{\mathrm{i}}}\Big\} \\
&= \Big\{(x,v) \,:\, v = \lim_{t \to 0} \frac{f(x + t\mathrm{e}_{\mathrm{i}}) - f(x)}{t}\Big\} \\
&= \Big\{(x,v) \,:\, (\forall\, \varepsilon > 0)(\exists\, T\,)(\forall\, t \geq T\,) \Big|\frac{f(x + t\mathrm{e}_{\mathrm{i}}) - f(x)}{t} - v\Big| \leq \varepsilon\Big\}.
\end{aligned}$$

Esto último es una fórmula de primera orden que involucra objetos semialgebraicos.

$\square$

A continuación se muestran dos resultados: el primero dice que las funciones semialgebraicas eventualmente no crecen más rápido que un polinomio. El segundo, que es consecuencia del primero, afirma una desigualdad que será una herramienta importante para estudiar las órbitas de sistemas dinámicos gradiente y sus generalizaciones.

Proposición 1.13 (Ver [4]) *Sea $f : (a, \infty) \to \mathbb{R}$ una función semialgebraica. Existen $b \geq a$ y un entero $k \in \mathbb{N}$ tal que:*

$$|f(x)| \leq |x|^k, \quad \forall\, x \in (b, \infty).$$

Demostración. Como f es semialgebraica, su grafo lo es. Entonces por la proposición **1.5**, $\mathrm{gph} = \bigcup_{\mathrm{i}=1}^{m} A_{\mathrm{i}}$, donde $A_{\mathrm{i}} = \{x \in \mathbb{R} \,:\, p_{\mathrm{i}}(x) = 0,\, q_1^{(\mathrm{i})}(x) > 0,\, \ldots,\, q_{r(\mathrm{i})}^{(\mathrm{i})}(x) > 0\}$.
Por otro lado $\mathrm{gph}(f) \subset \mathbb{R}^2$, es decir, los polinomios de la definición tienen dos variables. Más aún todos los polinomios p_{i} tienen necesariamente grado estrictamente positivo en y. Sino fuera así, por la forma de las otras restricciones el grafo tendría un intervalo vertical $\{t\} \times (a, b)$ lo cual es absurdo para el grafo de una función.
Sea entonces $p(x) = \prod_{\mathrm{i}=1}^{m} p_{\mathrm{i}}(x) = a_0(x) + a_2(x)y + \ldots + a_r(x)y^r$. Necesariamente $a_r(x)$ es un polinomio no nulo y por lo tanto para b suficientemente grande, $a_r(x) \neq 0$ en (b, ∞). Un resultado clásico dice que si c es un cero de un polinomio como el de arriba entonces:

$$|c| \leq \max_{\mathrm{i}=1,..,r} \left(r \left| \frac{a_{\mathrm{i}}(x)}{a_r(x)} \right| \right)^{\frac{1}{\mathrm{i}}}.$$

Recordando que se está trabajando sobre el grafo de f, los puntos $y = f(x)$ son necesariamente ceros de alguno de los polinomios $p_{\mathrm{i}}(x)$. Por lo tanto son ceros de p y se deduce:

$$|f(x)| \leq \max_{\mathrm{i}=1,..,r} (r \frac{|a_{\mathrm{i}}(x)|}{|a_r(x)|})^{1/\mathrm{i}}, \quad \forall\, x \in (b', \infty)$$

con b' suficientemente grande. Como el lado derecho de la desigualdad se puede acotar por un polinomio para x también grande, se concluye que existe $b > b'$ tal que para cada $x \in (b, \infty)$ se cumple:

$$|f(x)| \leq |x|^N.$$

□

Teorema 1.14 (Desigualdad de Łojasiewicz, Ver [4]) *Sean $K \subset \mathbb{R}^n$ un compacto semialgebraico y f, $g : K \to \mathbb{R}$ funciones semialgebraicas y continuas. Si son tales que se satisface lo siguiente:*

$$\forall\, x \in K, \qquad f(x) = 0 \implies g(x) = 0$$

esto es que $f^{-1}(0) \subset g^{-1}(0)$, entonces existen un entero $k \in \mathbb{N}$ y una constante $C \geq 0$ tales que:

$$\forall\, x \in K, \qquad |g(x)|^k \leq C|f(x)|.$$

Demostración. El truco, tal como en varias demostraciones anteriores, reside en definirse el conjunto semialgebraico adecuado. En este caso dicho conjunto es

$$F_t = \{x \in K \,:\, |g(x)| = \frac{1}{t},\, t > 0\},$$

que resulta ser es semialgebraico pues $|g|$ es semialgebraica. Por otro lado, si F_t es no vacío como la función g no se anula en ese conjunto entonces tampoco lo hace f.
La función $\dfrac{1}{|f|}$ es continua en F_t y semialgebraica, de hecho alcanza un máximo en F_t, denotado $m(t)$. Para los casos en que F_t resulte ser vacío se fija $m(t) = 0$, con esto queda definida la función $m(t) : (0, \infty) \to \mathbb{R}$. Dicha función es semialgebraica. Además cabe notar que si todos los conjuntos F_t fueran vacíos, entonces la función g es 0 por lo que bastaría seguir la línea previa para concluir el resultado.
Lo que sigue es analizar si la función m es semialgebraica:

$$\operatorname{gph}(m) = \Big\{(t, y) \,:\, (\forall\, x \text{ tal que } t|g(x)| = 1,\, y \geq \frac{1}{|f(x)|}) \vee (y = 0 \;\text{ si } \forall\, x\;\; t|g(x)| \neq 1)\Big\}.$$

Se verifica que la fórmula que describe al conjunto anterior es de primer orden, y por lo tanto este último es semialgebraico.
Aplicando la proposición anterior se obtienen una constante b y un entero $n \in \mathbb{N}$ tales que

$$m(t) \leq t^n \;\; \forall\, t \in (b, \infty).$$

Reinterpretando los roles de $m(t)$ como el máximo de $1/|f(x)|$ en los puntos donde $t = 1/|g(x)|$ esto último es equivalente a:

$$\forall\, x \in K \;\, 0 < |g(x)| < \frac{1}{b} \implies \frac{1}{|f(x)} \leq \frac{1}{|g(x)|^n}.$$

Por lo que en este caso se obtiene la desigualdad buscada. Ahora el conjunto faltante para concluir la desigualdad es $\Big\{x \in K \,:\, \dfrac{1}{b} \leq |g(x)|\Big\}$.
Sea B el máximo de la función $\dfrac{|g(x)|^n}{f(x)}$ sobre este conjunto (se alcanza pues este es compacto y la función es continua). Se obtiene que $\dfrac{|g(x)|^n}{f(x)} \leq B$, de modo que para concluir la desigualdad basta tomar como constante $C = \max\{1, B\}$. $\square$

Ahora se volverá al caso más general. Todos resultados previos sobre funciones y otras propiedades se mantienen para estructuras o-minimales arbitrarias, con la salvedad de las dos desigualdes importantes, cuyas versiones para estructuras o-minimales genéricas se estudiarán en los próximos capítulos.

El detalle relevante que justifica que lo anterior es válido en estructuras o-minimales más generales, es que la propiedad de la proyección (que se demuestra para conjuntos semialgebraicos) será ahora un axioma (véase definición **1.1**). Dicho axioma es lo que reemplaza la necesidad de tener el teorema de Tarski-Seidenberg. Pronto se verá de todas formas el teorema potente que hace esto posible, se definirán las fórmulas de primer orden para estas estructuras y se tendrá el análogo a la tercera versión de Tarski-Seidenberg.

Por otro lado, lo hecho hasta ahora muestra una amplia cantidad de funciones o-minimales lo que sustenta su valor y hace que ser definible no sea una hipótesis exagerada al momento de plantear una proposición.

Se presentan primero algunos resultados cortos, que también serán útiles en el presente trabajo.

Proposición 1.15 *Si $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ es una función definible, y $B \subset \mathbb{R}^m$ es un conjunto definible restringido en gph(f) entonces $f^{-1}(B)$ también es un conjunto definible.*

Demostración. Basta notar que $B = \pi(\text{gph}(f)) \cap (A \times B)$, donde $\pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ proyecta sobre las primeras m coordenadas. □

Proposición 1.16 *La adherencia y el interior de un conjunto definible en $\mathbb{R}^n$ son definibles.*

Demostración. Si $A \subset \mathbb{R}^n$ es definible, entonces $\text{d}(x, A)$ es una función definible, como $\overline{A} = \text{d}(\cdot, A)^{-1}(0)$ por lo que se concluye que la adherencia de definibles es definible. Usando complemento se deduce que el interior de un conjunto definible también lo es. □

En lo que sigue se muestra la versión general de este importante resultado, que es análogo al caso semialgebraico.

Definición 1.17 (Fórmula de primer orden, versión definible. Ver [3]) *Una fórmula de primer orden se construye de acuerdo a las siguientes reglas:*

(i) *Si $p \in \mathbb{R}[x_1, ..., x_n]$ (polinomio a n variables reales), entonces $p(x_1, ..., x_n) = 0$ y $p(x_1, ..., x_n) > 0$ son fórmulas definibles de primer orden.*
(ii) *Si A es un subconjunto definible de $\mathbb{R}^n$ entonces $x \in A$ es una fórmula definible de primer orden.*
(iii) *Si $\Phi(x_1, ..., x_n)$ y $\Psi(x_1, ..., x_n)$ son fórmulas definibles de primer orden entonces su conjunción $\Phi \wedge \Psi$, disyunción $\Phi \vee \Psi$, negación $\neg\Psi$ e implicancia $\Phi \implies \Psi$ son fórmulas definibles de primer orden.*
(iv) *Si $\Phi(x, y)$ con $x = (x_1, ..., x_n)$ e $y = (y_1, ..., y_k)$ es una fórmula definible de primer orden y $A \in \mathcal{S}_k$ entonces las fórmulas $(\exists y \in A)\Phi(x, y)$ y $(\forall y \in A)\Phi(x, y)$ son también definibles de primer orden.*

Teorema 1.18 (Ver [3]) *Si $\Psi(x_1, ..., x_n)$ es una fórmula definible de primer orden, entonces*

$$\{(x_1, ..., x_n) \,:\, \Psi(x_1, ..., x_n)\}$$

es un conjunto definible.

Cabe reiterar que lo destacable de este teorema es su utilidad en cuanto a identificar rápidamente conjuntos definibles, lo que facilita y acorta considerablemente las demostraciones que se hagan.

Observación Si $f : A \subset \mathbb{R} \to \mathbb{R}$ es o-minimal, y si F es una primitiva de f, lamentablemente no necesariamente se cumple que esta última sea definible en la misma estructura o-minimal. Por ejemplo, sea

$$\begin{aligned} f : (0, \infty) &\to \mathbb{R} \\ x &\mapsto f(x) = \frac{1}{x} \end{aligned}$$

Claramente f es semialgebraica, pero su primitiva $F(x) = \ln x$ falla en serlo. Si $\ln(x)$ fuera semialgebraica, entonces su inversa e^x también lo sería, lo cual contradice la proposición **1.13** que la fuerza a tener crecimiento polinomial a partir de un instante. Sin embargo, ambas funciones $\ln x$ y e^x pertenecen a una estructura más grande, llamada $\mathbb{R}_{(an,exp)}$.

Teorema 1.19 (Teorema de monotonía. Ver [3]) *Sea $f : (a, b) \to \mathbb{R}$ o-minimal. Entonces existen $a = a_0 < a_1 < ... < a_k = b$ tal que en cada intervalo de la forma $(a_\mathrm{i}, a_{\mathrm{i}+1})$, f es constante o continua y estrictamente monótona.*

Ahora se echará un vistazo a una construcción que es de gran utilidad.

Definición 1.20 (Descomposición cilíndrica definible en celdas de $\mathbb{R}^n$, abreviadamente **dcdc**. Ver [3]) *Se define una **dcdc** de $\mathbb{R}^n$ como una partición de este en conjuntos definibles $(C_\lambda)_\lambda$ que se llamarán las celdas de la **dcdc** y deben satisfacer lo siguiente:*

- *Si la dimensión es $n = 1$ entonces una **dcdc** de $\mathbb{R}$ está dado por una colección finita de singletons $\{a_\mathrm{i}\}_{\mathrm{i}=1}^r$ con $a_1 < ... < a_r$, los intervalos comprendidos entre estos de la forma $(a_\mathrm{i}, a_{\mathrm{i}+1})$ y los intervalos $(-\infty, a_1)$ y (a_r, ∞).*
- *Para $n > 1$ una **dcdc** de $\mathbb{R}^n$ está construido a partir de una **dcdc** de $\mathbb{R}^{n-1}$, en donde para cada celda C de $\mathbb{R}^{n-1}$ existen funciones continuas y definibles $f_1^C < ... < f_r^C(C)$, definidas sobre C y a valores reales, tal que las celdas de $\mathbb{R}^n$ son los grafos de estas funciones: $\{(x, f_j^C(x)) : x \in C\}$ y las bandas $(f_j^C, f_{j+1}^C) := \{(x, y) : x \in C \text{ y } f_j^C < y < f_{j+1}^C)\}$, todo lo anterior para $0 \leq j \leq r(C)$ en donde $f_0^C = -\infty$ y $f_{r(C)+1}^C = \infty$.*

Definición 1.21 (Ver [3]) *Se define inductivamente la dimensión de una celda de la siguiente manera:*

- *La dimensión de un singleton es 0 y de un intervalo es 1.*
- *Si C es una celda en una **dcdc** de $\mathbb{R}^n$ entonces su dimensión es $\dim(\pi(C))$ si C es un grafo o $\dim(\pi(C)) + 1$ si C es una banda, donde $\pi : \mathbb{R}^n \to \mathbb{R}^{n-1}$ es la proyección canónica.*

Proposición 1.22 (Ver [3]) *Para cada celda C de una **dcdc** de $\mathbb{R}^n$ existe un homeomorfismo definible $\theta_C : C \to \mathbb{R}^{\dim(C)}$.*

En lo que sigue se verán tres teoremas bastantes importantes para la teoría general. Aunque no se les dará un gran uso explícito en los capítulos posteriores, de todos modos es adecuado revisarlos.

Teorema 1.23 (Uniformidad finita. Ver [3]) *Sea $A \subset \mathcal{S}_n$ tal que para cada $x \in \mathbb{R}^{n-1}$ el conjunto $A_x = \{y \in \mathbb{R} : (x, y) \in A\}$ es finito. Entonces existe un $k \in \mathbb{N}$ tal que $|A_x| \leq k$ para todo $x \in \mathbb{R}^{n-1}$.*

Teorema 1.24 (Continuidad por trozos. Ver [3]) *Sean $A \subset \mathbb{R}^n$ un conjunto definible y $f : A \to \mathbb{R}$ una función o-minimal. Existe una **dcdc** de $\mathbb{R}^n$ tal que A es unión de celdas de esta **dcdc**, y tal que para cada $C \subset A$ se tiene que $f\big|_C$ es continua.*

Teorema 1.25 (Descomposición de celdas. Ver [3]) *Sean $A_1, ..., A_k \subset \mathcal{S}_n$, existe una **dcdc** de $\mathbb{R}^n$ tal que cada A_j se escribe como unión de celdas. Si una **dcdc** de $\mathbb{R}^n$ satisface esta propiedad entonces se dirá que es una **dcdc** adaptada a $A_1, ..., A_k$.*

Demostración. La prueba se hará haciendo una inducción sobre la dimensión, el caso $n = 1$ es inmediato de que los únicos conjuntos definibles sobre $\mathbb{R}$ son uniones finitas de puntos e intervalos. Suponiendo el resultado para los conjuntos de $\mathbb{R}^{n-1}$, se prueba el caso n-dimensional como sigue:
El primer paso es considerar el siguiente conjunto:

$$A = \{(x, y) \in \mathbb{R}^{n-1} \times \mathbb{R} \, : \, y \in \partial A_{\mathrm{i},x} \, \text{para algún} \, \mathrm{i} = 1, ..., k\}$$

donde ∂X denota la frontera topológica del conjunto X.
Los conjuntos $A_{\mathrm{i},x}$ son definibles y por lo tanto sus interiores y adherencias también lo son. Entonces sus fronteras son conjuntos definibles, de lo que se deduce que el conjunto A es definible por estar definido a partir de una fórmula definible de primer orden.
Por otro lado el conjunto $A_x = \cup_{\mathrm{i}=1}^{k} \partial A_{\mathrm{i},x}$ es finito. Esto último es consecuencia de que cada conjunto $A_{\mathrm{i},x}$ está en $\mathcal{S}_1$ y por lo tanto es unión finita de puntos e intervalos, de lo que se deduce que $\partial A_{\mathrm{i},x}$ no tiene más opción que ser unión finita de puntos.
Gracias a lo anterior se puede aplicar el teorema de uniformidad finita al conjunto A para concluir que $|A_x| \leq l$ para cada $x \in \mathbb{R}^{n-1}$.
Para $x \in \mathbb{R}^{n-1}$ e $\mathrm{i} = 1, ..., l$ se define $f_\mathrm{i} : x \mapsto a_\mathrm{i}$ en donde a_i es el i-ésimo elemento de A_x (de acuerdo al orden usual de $\mathbb{R}$) si es que existe (por lo que las f_i no estarán definidas en todo $\mathbb{R}^n$ sino en subconjuntos definibles), esto está bien definido pues el conjunto es finito con a lo más l puntos. Además estas funciones son o-minimales, pues nuevamente se puede escribir su grafo como el conjunto de puntos tales que satisfacen una fórmula de primer orden. Sumado a esto, por la definición de las funciones f_i se obtiene que A es la unión de sus grafos.

Con todo lo anterior, se pueden clasificar los puntos $x \in \mathbb{R}^{n-1}$ de acuerdo a la siguiente información:

(i) El primer dato es la cardinalidad $|A_x|$.
(ii) Luego están los índices $j \in \{1, ..., k\}$ tales que $f_{\text{i}}(x) \in A_{j,x}$ para $\text{i} = 1, ..., |A_x|$.
(iii) Finalmente los índices $j \in \{1, ..., k\}$ tales que $(f_{\text{i}}(x), f_{\text{i}+1}(x)) \subset A_{j,x}$ para $\text{i} = 0, ..., |A_x|$, donde se definen $f_0(x) = -\infty$ y $f_{|A_x|+1}(x) = \infty$.

Todas las posibilidades anteriores son finitas y cada una constituye un conjunto definible en donde un $x \in \mathbb{R}^{n-1}$ puede estar, pues estas opciones se describen por fórmulas de primer orden. Aplicando la hipótesis inductiva se concluye la existencia de una **dcdc** de $\mathbb{R}^{n-1}$ adaptada a estos conjuntos, por lo que si dos puntos comparten celda entonces también tienen la misma información. Además usando el teorema de continuidad por trozos se puede asumir que: para cada celda C de la **dcdc** y para cada $\text{i} = 1, ..., l$ f_{i} no está definida en C, o sí lo está, siendo necesariamente continua. Con todo esto se construye una **dcdc** de $\mathbb{R}^n$ usando de la **dcdc** de $\mathbb{R}^{n-1}$ y las funciones f_{i} lo que concluye la demostración del teorema. □

Teorema 1.26 (Teorema de elección definible. Ver [3]) *Sean $A \subset \mathbb{R}^m \times \mathbb{R}^n$ un conjunto o-minimal,*
$\pi_{m+n,m} : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ la proyección canónica sobre las primeras m coordenadas. Entonces existe una función $f : \pi_{m+n,m}(A) \to \mathbb{R}^n$ tal que para cada $x \in \pi_{m+n,m}(A)$, el par $(x, f(x))$ pertenece a A.

Demostración. El problema se puede reducir al caso unidimensional, pues en el caso general se puede descomponer en las proyecciones sucesivas de $\mathbb{R}^{m+n}$ hasta llegar a $\mathbb{R}^n$.
Sea entonces una **dcdc** de $\mathbb{R}^{m+1}$ adaptada a A. La proyección $\pi(A)$ es la unión de las proyecciones de las celdas contenidas en A, por lo que se puede suponer que A es una celda de $\mathbb{R}^{m+1}$ y por lo tanto $\pi(A)$ es una celda de $\mathbb{R}^m$. Suponer que A es una celda es suponer dos casos:

- $A = \text{gph}(\zeta)$ donde $\zeta : \pi(A) \to \mathbb{R}$, esto por la definición de una **dcdc**. En tal caso se concluye tomando $f = \zeta$, o bien
- A es una banda $(\zeta_{\text{i}}, \zeta_{\text{i}+1})$, de donde surgen más opciones: A ambas funciones son finitas en A y en ese caso $f = \frac{1}{2}(\zeta_{\text{i}} + \zeta_{\text{i}+1})$ sirve, o alguna de ellas sea $-\infty$ o $+\infty$ y en tal caso se toma $f = \zeta_{\text{i}+1} - 1$ o $f = \zeta_{\text{i}} + 1$ respectivamente.

□

El teorema se ve más práctico y toma una forma elegante si está escrito como sigue:

Teorema 1.27 *Dados, A, B, C conjuntos definibles, si una fórmula del estilo:*

$$(\forall\, x \in A)(\exists\, y \in B)((x, y) \in C)$$

es cierta, entonces $y = y(x)$ define una función o-minimal de en A.

El próximo teorema proporcionará una herramienta potente para lo que sigue, y se le dará uso reiteradas veces.

Lema 1.28 (Selección de curvas. Ver [3]) *Sea A un definible de $\mathbb{R}^n$ y $a \in \overline{A}$. Entonces existe una curva continua, o-minimal $\gamma : [0,1) \to \mathbb{R}^n$ tal que $\gamma(0) = a$ y $\gamma((0,1)) \subset A$.*

Demostración. Sea $X = \{(t,x) \in \mathbb{R}\times\mathbb{R}^n \,:\, x \in A \text{ y } \|x-a\| < t\}$. Como $a \in \overline{A}$ se tiene que $\pi_{n+1,1}(X) = \{t \in \mathbb{R} \,:\, t > 0\}$, pues los puntos en $\overline{A}$ son exactamente los que están a distancia 0 de A. Aplicando el teorema de elección definible al conjunto X se obtiene una función $\delta : (0,\infty) = \pi_{n+1,1}(X) \to A$ tal que $\|\delta(t)-b\| < t$, por teorema de monotonía se puede tomar $\delta : (0,\varepsilon) \to A$ continua aún satisfaciendo que $\|\delta(t) - b\| < t$ y que por continuidad se puede extender a una función $\delta : [0,\varepsilon) \to A$ tal que además de lo anterior se tiene que $\delta(0) = a$. □

Se concluirá la sección con una definición y una proposición, también potentes, pero para el trabajo presente será usando tanto como los anteriores.

Definición 1.29 (Estratificación de Whitney, véase [2]) *Sea $Q \subset \mathbb{R}^n$ un conjunto. Una estratificación de Whitney C^p regular del conjunto Q es una partición de dicho conjunto en finitas variedades C^p regulares, que serán llamados estratos, satisfaciendo dos condiciones de compatibilidad:*

(i) *La condición de frontera: Para cualquier par de estratos L y M se cumple:*

$$L \cap \overline{M} \neq \emptyset \implies L \subset \overline{M} \setminus M,$$

es decir, si dos de estas variedades se intersectan entonces una es frontera de la otra.

(ii) *Para cualquier secuencia de puntos $\{x_n\}_n$ en un estrato M, tal que $x_n \to \bar{x} \in L$, si los vectores normales correspondientes a dichos puntos, $\nu_n \in N_M(x_i)$ convergen a un vector ν, entonces $\nu \in N_L(\bar{x})$. Esta condición puede pensarse como una regularidad en la forma en que las variedades de la estratificación están pegadas.*

Proposición 1.30 (Ver [2]) *Sea $A \subset \mathbb{R}^n$ un conjunto definible en una estructura o-minimal. Entonces A admite una Estratificación de Whitney C^p regular.*

Para cerrar este capítulo es relevante mencionar que la teoría de geometría o-minimal es más extensa que lo que fue recién expuesto. Se ha centrado la presentación en las herramientas que serán utilizadas en el presente trabajo.

Capítulo 2

Elementos de análisis variacional

En esta sección se presentan conceptos de análisis variacional que se utilizarán más adelante. Si bien la mayoría de los conceptos que siguen se pueden desarrollar sobre espacios normados o incluso sobre espacios vectoriales topológicos localmente convexos, esto no aporta en nada en la presente memoria. Por lo tanto el marco de presentación se limitará a $\mathbb{R}^n$.

A partir de todo lo que sigue $\overline{\mathbb{R}}$ denotará $\mathbb{R} \cup \{\infty\}$.

Definición 2.1 (Función semicontinua inferior, abreviadamente s.c.i.. Ver [9]) *Una función $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ se llama semicontinua inferior si su epígrafo es un conjunto cerrado.*

Definición 2.2 (Cono. Ver [9]) *Un subconjunto no vacío C de $\mathbb{R}^n$ se llama cono si para cada $x \in C$ y $\lambda \geq 0$ se tiene que $\lambda x \in C$.*

Definición 2.3 (Multiaplicación. Ver [9]) *Sean X, Y conjuntos. Una multiaplicación de X a Y, denotada $T : X \rightrightarrows Y$ es una función $T : X \to 2^Y$.*

Para el contexto de este trabajo solo se prestará anteción a las multiaplicaciones de la forma $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$.

Definición 2.4 (Ver [9]) *Sea $T : X \rightrightarrows Y$ una multiaplicación.*

(i) *El dominio de T es el conjunto siguiente:*

$$\mathrm{dom} T := \{x \in \mathbb{R}^n \, : \, T(x) \neq \emptyset\}.$$

(ii) *Se define el grafo de T como sigue:*

$$gph\, T := \{(x, y) \, : \, y \in T(x)\}.$$

Definición 2.5 (Ver [9]) *Una multiaplicación T se dice compacta (respectivamente convexa) si sus imágenes $T(x) \subset \mathbb{R}^m$ son conjuntos compactos (respectivamente convexos). Similarmente se dice cerrada (respectivamente definible) si su grafo es cerrado (respectivamente definible).*

Observación Si $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ es una multiaplicación, entonces su inversa dada por: $T^{-1}(y) = \{x : y \in F(x)\}$ es también una multiaplicación.

Definición 2.6 (Ver [9]) *Una multiaplicación $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ se dice positivamente homogénea si su grafo es un cono, esto es: $0 \in T(0)$ y $T(\lambda x) = \lambda T(x)$ para todo $\lambda > 0$ y $x \in \mathbb{R}^n$.*

Observación Si $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ es una multiaplicación positivamente homogénea, entonces su inversa $T^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ también es positivamente homogénea.

Definición 2.7 (Norma exterior. Ver [9]) *Para una multiaplicación $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ positivamente homogénea se define su norma exterior como:*

$$|T|^+ := \sup_{x \in B(0,1)} \sup_{y \in T(x)} \|y\|.$$

Observación Si T es una multiaplicación positivamente homogénea se puede mostrar que su norma exterior admite la siguiente definición equivalente:

$$|T|^+ = \inf\{\lambda > 0 \,:\, \|y\| \leq \lambda \|x\|\,,\; x \in T(y)\}.$$

Asimismo se tiene la siguiente fórmula para la norma exterior de su inversa:

$$|T^{-1}|^+ = \frac{1}{\inf_{\|x\|} \operatorname{dist}(0, T(x))}.$$

A continuación se revisarán algunos elementos clásicos de análisis no diferenciable:

Definición 2.8 (Vector tangente. Ver [9]) *Un vector $v \in \mathbb{R}^n$ se dice tangente a un conjunto $C \subset \mathbb{R}^n$ en un punto $\bar{x} \in C$ si para secuencias $\{x_n\}_n \subset C$, $x_n \to \bar{x}$ y $t_n \to 0$ se tiene que:*

$$\frac{x_n - \bar{x}}{t_n} \to v$$

Definición 2.9 (Cono tangente. Ver [9]) *Dado un conjunto $C \subset \mathbb{R}^n$ y un punto $\bar{x} \in C$, se define el cono tangente de C en el punto $\bar{x}$ como $T_C(\bar{x}) = \{v \in \mathbb{R}^n : v$ es vector tangente a $C\}$.*

Proposición 2.10 *Sea $\bar{x} \in C \subset \mathbb{R}^n$. Se tiene la siguiente caracterización de $T_C(\bar{x})$:*

$$T_C(\bar{x}) = \limsup_{t \searrow 0} \frac{C - \bar{x}}{t}.$$

Definición 2.11 (Cono normal. Ver [9]) *Sean $C \subset \mathbb{R}^n$ y $\bar{x} \in C$.*

(i) *Un vector $\nu \in \mathbb{R}^n$ se dice normal al conjunto C en el punto $\bar{x}$ en el sentido regular (normal regular) si para cada $x \in C$:*

$$\langle \nu, x - \bar{x} \rangle \leq o(|x - \bar{x}|).$$

Al conjunto de vectores normales en el sentido regular se le llama cono normal de Fréchet y se denota por $\hat{N}_C(\bar{x})$.

(ii) *Un vector* $\nu \in \mathbb{R}^n$ *se dice normal a* C *en* $\bar{x}$ *en el sentido general si existen secuencias* $\{x_n\}_n \subset C$ *con* $x_n \to \bar{x}$ *y* $\nu_n \subset N_C(\bar{x}_n)$ *con* $\nu_n \to v$. *Al conjunto de estos últimos se le llama cono normal límite.*

(iii) *Al conjunto* $N_C^c(\bar{x}) := \overline{conv(N_C(\bar{x}))}$ *se le llama cono normal de Clarke a* C *en el punto* $\bar{x}$.

Definición 2.12 (Derivada direccional. Ver [9]) *Sean* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *función y* $\bar{x}$ *un punto donde* $f(\bar{x}) < \infty$. *Se define la derivada direccional en el punto* $\bar{x}$ *como:*

$$\mathrm{d}f(\bar{x})(u) = \lim_{\substack{t\to 0\\ v\to u}} \frac{f(\bar{x}+tv) - f(\bar{x})}{t}.$$

El resultado que sigue muestra que las derivadas direccionales expresan geométricamente el cono tangente al epígrafo de la función. Esto no se utilizará directamente en el presente trabajo, pero es relevante revisar esta noción como una relación básica entre la derivada de una función y la geometría subyacente a esta.

Teorema 2.13 (Ver [9]) *Para* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *y todo punto* $\bar{x}$ *donde* $f(\bar{x}) < \infty$ *se cumple que:*

$$\mathrm{epi}(\mathrm{d}f)(\bar{x}) = T_{\mathrm{epi}(f)}(\bar{x}, f(\bar{x})).$$

Definición 2.14 (Subgradientes. Ver [9]) *Sea* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *y un punto* $\bar{x}$ *en donde* f *es finita.*

(i) *Un vector* $v \in \mathbb{R}^n$ *se dice subgradiente regular de* f *en* $\bar{x}$ *si satisface que para cada* $x \in \mathbb{R}^n$:

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x}\rangle + o(|x - \bar{x}|).$$

El conjunto de los subgradientes regulares en $\bar{x}$ *se denota por* $\hat{\partial} f(\bar{x})$, *a este conjunto se le llama subdiferencial de Fréchet.*

(ii) *Un vector* $v \in \mathbb{R}^n$ *se dice subgradiente generalizado de* f *en* $\bar{x}$ *si existen sucesiones* $\{x_n\}_n$ *y* $\{v_n\}_n$ *tales que* $f(x_n) \to f(\bar{x})$, $v_n \in \hat{\partial} f(x_n))$ *y* $v_n \to v$. *El conjunto de los subgradientes generalizados en* $\bar{x}$ *se denota por* $\partial f(\bar{x})$, *y se llama subdiferencial límite.*

Proposición 2.15 (Subdiferencial de Fréchet. Ver [9]) *Si* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *es semicontinua inferior y* $\bar{x}$ *es un punto en donde* f *es finita entonces se tiene la siguiente fórmula para el subdiferencial de Fréchet:*

$$\hat{\partial} f(x) = \Big\{ v \in \mathbb{R}^n \, : \, \liminf_{\substack{y\to x\\ y\neq x}} \frac{f(y) - f(x) - \langle v, y - x\rangle}{\|y - x\|} \Big\}.$$

Definición 2.16 (Derivada direccional de Clarke-Rockafellar. Ver [9]) *Dada una función* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *se define su derivada direccional de Clarke-Rockafellar en el punto* $\bar{x}$ *donde* f *es finita y con respecto a la dirección* v *como:*

$$\mathrm{d}^{\uparrow} f(\bar{x}, v) := \sup_{\varepsilon>0} \limsup_{\substack{y\to x\\ t\searrow 0^+}} \inf_{u\in B(v,\varepsilon)} \frac{f(y+tv) - f(y)}{t}.$$

El subdiferencial de Clarke-Rockafellar en un punto $\bar{x}$ *donde* f *es finita se define como sigue:*

$$\partial^{\uparrow}(\bar{x}) = \{v \in \mathbb{R}^n \, : \, \mathrm{d}^{\uparrow} f(\bar{x}, v) \geq \langle v, u\rangle, \ \forall\, u \in \mathbb{R}^n\}.$$

Proposición 2.17 (Ver [9]) *Sean $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ y $\bar{x}$ tal que $f(x) < \infty$. Entonces:*

$$\hat{\partial} f(\bar{x}) = \{v \in \mathbb{R}^n \mid \langle v, w\rangle \leq \mathrm{d}f(\bar{x})(w),\ \forall\ w \in \mathbb{R}^n\}.$$

El siguiente resultado proporciona representaciones de los subdiferenciales presentados, mediante los conos normales.

Teorema 2.18 (Ver [9]) *Sean $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ y $\bar{x}$ un punto donde f es finita, se tiene lo siguiente:*

$$\begin{aligned}
\hat{\partial} f(\bar{x}) &= \{v \in \mathbb{R}^n \mid (v, -1) \in \hat{N}_{epi\ f}(\bar{x}, f(\bar{x}))\}.\\
\partial f(\bar{x}) &= \{v \in \mathbb{R}^n \mid (v, -1) \in N_{epi\ f}(\bar{x}, f(\bar{x}))\}.\\
\partial^c f(\bar{x}) &= \{v \in \mathbb{R}^n \mid (v, -1) \in N^c_{epi\ f}(\bar{x}, f(\bar{x}))\}.
\end{aligned}$$

Proposición 2.19 (Ver [9]) *Sean $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ y $\bar{x}$ un punto donde f es finita. Si f es semicontinua inferior entonces:*

$$\partial^{\uparrow} f(x) = \{v \in \mathbb{R}^n\ :\ f^{\uparrow}(x, u) \geq \langle v, u\rangle\ \forall\, u \in \mathbb{R}^n\}.$$

Más aún $\partial^{\uparrow} f(x)$ es no vacío en un conjunto denso.

Proposición 2.20 (Ver [9]) *Sean $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ y $\bar{x}$ un punto donde f es finita. Si la función f es Lipschitz entonces se tiene la siguiente caracterización simplicada para la derivada direccional de Clarke-Rockafellar:*

$$f^{\circ}(x, v) = \limsup_{\substack{y \to x\\ t \searrow 0^+}} \frac{f(y + tv) - f(y)}{t}.$$

En este caso la función $f^{\circ}(x, \cdot)$ es sublineal y continua.

Teorema 2.21 (Ver [9]) *Si $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ es convexa y propia y todos los puntos están en su dominio entonces:*

$$\partial f(\bar{x}) = \{v \in \mathbb{R}^n \mid f(x) \geq f(\bar{x}) + \langle v, x - \bar{x}\rangle\ \forall\ x\} = \hat{\partial} f(\bar{x}).$$

Más elegantemente:

$$\partial f(\bar{x}) = \hat{\partial} f(\bar{x}) = \{v \in \mathbb{R}^n \mid f(\cdot) - \langle v, \cdot\rangle \text{ alcanza un mínimo global en } x\}.$$

Observación Para $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ semicontinua inferior, el grafo de su subdiferencial de límite, visto como una multiaplicación: $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ es la cerradura del grafo del subdiferencial de Fréchet.

Lo que sigue es una definición que extiende la noción de derivadas y subdiferenciales a las multiaplicaciones, si bien se verán algunos resultados que muestran su valor por sí mismas, su utilidad para el trabajo presente se verá explícitamente en lo que se llama *sweeping process.*

Definición 2.22 (Ver [9]) *Sea* $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *y un punto* $\bar{x} \in domS$. *Se define la coderivada regular de* S *en* $\bar{x}$ *sobre* $\bar{y} \in S(\bar{x})$ *como la multiaplicación* $\hat{D}^*S(\bar{x}|\bar{y}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *definida como:*

$$v \in \hat{D}^*S(\bar{x}|\bar{y})(u) \iff (v, -u) \in \hat{N}_{gph(S)}(\bar{x}, \bar{y}).$$

Similarmente se define la coderivada de Clarke como la multiaplicación $D_c^*S(\bar{x}|\bar{y}) : \mathbb{R}^n \to \mathbb{R}^m$ *definida como:*

$$v \in D_c^*S(\bar{x}|\bar{y})(u) \iff (v, -u) \in N_{gph(S)}^c(\bar{x}, \bar{y}).$$

Cabe destacar que esta generalización no se hace a mediante una fórmula explícita como en otros casos, sino que a través de una relación geométrica. Con este capítulo se cierra la revisión de contenidos básicos sobre los que se cimentan el trabajo que sigue.

Capítulo 3

Desigualdad de Łojasiewicz en el caso suave

En ese capítulo se presentarán aplicaciones en la teoría de sistemas dinámicos de tipo (sub)gradiente obtenidos de la combinación de la geometría o-minimal con en análisis variacional, dicha teoría se suele llamar *análisis variacional moderado.*
Krzysztof Kurdyka prueba resultados sobre el comportamiento asintótico de la dinámica del gradiente de una función definible, que se presenta a continuación
Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función definible de clase C^1. Se considera el sistema dinámico definido por su gradiente, es decir:

$$\dot{x}(t) = -\nabla f(x(t)).$$

El resultado que sigue es una generalización de la desigualdad de Lojasiewicz (teorema **1.14**), este último permitió concluir que las órbitas del sistema gradiente tienen longitud finita en los compactos.

Para empezar se enuncia el siguiente resultado que generaliza el teorema **1.14**.

Teorema 3.1 (Desigualdad de Kurdyka-Łojasiewicz - versión 1. Ver [6]) *Sean $K \subset \mathbb{R}^n$ un conjunto compacto y $f, g : K \to \mathbb{R}$ dos funciones definibles. Si los ceros de f son ceros de g, i.e. $f^{-1}(0) \subset g^{-1}(0)$ entonces existe una función definible estrictamente creciente y positiva $\sigma : \mathbb{R}_+ \to \mathbb{R}$ de clase C^1 tal que para cada $x \in K$ se tiene la siguiente desigualdad:*

$$|f(x)| \geq \sigma(g(x)).$$

Demostración. Sea $h : K \to \mathbb{R}$ la función dada por: $h(u) = (g(u), |f(u)|)$. Evidentemente h es definible y por lo tanto $h(K)$ es un subconjunto de $\mathbb{R}$ definible y además es un compacto pues h es continua. Por otro lado la propiedad sobre los ceros de las funciones $|f|$ y g indica que

$$h(K) \cap \{(x, y) \,:\, y = 0\} = (0, 0).$$

Aplicando el lema de selección de curvas (teorema **1.28**) junto con teorema de monotonía (ver teorema **1.19**) se obtiene una función definible estrictamente creciente $\sigma : \mathbb{R}_+ \to \mathbb{R}$ que es además de clase C^1 tal que $h(K) \subset \{y \geq \sigma(x),\, x \geq 0\}$. Esto último se traduce en que para cada $u \in K$ $g(u)$ y $f(u)$ son tales que $|f(x)| \geq \sigma(g(u))$. □

Cabe observar que el rol de σ en el caso semialgebraico lo hace la función x^k (ver teorema **1.14**).

Ahora corresponde darle un sentido a cómo se aplicará la desigualdad anterior, que será llamada desigualdad de Kurdyka-Łojasiewicz para sistemas dinámicos de tipo gradiente, $\dot{x}(t) = -\nabla f(x(t))$, en donde f es una función definible en una estructura o-minimal y por el momento C^1.

Una pregunta que surge al estudiar tales sistemas es si sus órbitas acotadas tienen longitud finita. La respuesta es positiva y la idea de la demostración es en esencia simple. Si se compone tal función f con una función estrictamente creciente y C^1, la nueva función $\sigma \circ f$ inducirá el mismo retrato de fases. Por lo tanto el comportamiento de las trayectorias del sistema gradiente original se puede estudiar a partir de este nuevo, es decir, las deducciones que se puedan hacer sobre el nuevo sistema son directamente válidas para el sistema original.

Observación Algunas observaciones básicas son las siguientes:

(i) Si $x(t)$ es una órbita del sistema gradiente, entonces es suave y su tangente es normal a los conjuntos de nivel de la función $[f = r] = \{x \in \mathbb{R}^n \,:\, f(x) = r\}$. Esto es consecuencia de que el gradiente de una función es ortogonal a sus conjuntos de nivel.

(ii) La función f evaluada a lo largo de sus trayectorias es decreciente. Esto es consecuencia de que $\frac{\mathrm{d}}{\mathrm{d}t}(f(x(t))) = \dot{x}(t) \cdot \nabla f(x(t)) = -\|\nabla f(x(t))\|^2 \leq 0$. Suponiendo que la función es propia, se deduce que $f(x(t))$ converge a un valor asintótico cuando $t \to \infty$.

(iii) El punto anterior abre la pregunta de si $f(x(t))$ converge a un valor crítico de la función, pero este tema será tratado pronto.

Esto se profundizará más en los capítulos posteriores.

El siguiente teorema es válido para estructuras o-minimales polinomialmente acotadas, es decir, en donde el crecimiento al *infinito* de cualquier función definible a valores reales se puede acotar de manera polinomial (por ejemplo la estructura de los conjuntos semialgebraicos). Si bien no se le dará un uso, cabe mencionarlo como un modelo de desigualdad para el gradiente, la llamada desigualdad de Łojasiewicz.

Teorema 3.2 (Desigualdad de Łojasiewicz. Ver [6]) *Sean $\Omega \subset \mathbb{R}^n$ abierto y acotado y $f : \Omega \to \mathbb{R}$ una función o-minimal. Suponiendo f diferenciable en $\Omega \setminus f^{-1}(0)$. Entonces existen constantes $C > 0$, $\rho > 0$ y $\theta \in [0, 1)$ tales que:*

$$\|\nabla f(x)\| \geq C|f(x)|^{\theta}$$

para cada $x \in \Omega$ tal que $|f(x)| \in (0, \rho)$.

El lema siguiente será bastante útil para la demostración del teorema próximo.

Lema 3.3 (Ver [6]) *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}$ una función definible tal que $f(x) \geq 0$ para cada $x \in A$ y $G : A \to \mathbb{R}^m$ otra función definible. Entonces la siguiente función es definible:*

$$\varphi : G(A) \to \mathbb{R}.$$
$$\varphi(y) = \inf_{x \in G^{-1}(y)} f(x).$$

Demostración. Basta representar el grafo mediante la siguiente fórmula:

$$\operatorname{gph}(\varphi) = \{(y, t) \, : \, y \in G(A), \, t = \inf_{x \in G^{-1}(y)} f(x)\},$$

y notar que la fórmula del ínfimo se reescribe como:

$$t = \inf_{x \in G^{-1}(y)} f(x) \iff \Big(t \leq f(x) \; \forall \, x \in G^{-1}(y)\Big) \wedge \Big(t' \leq f(x) \; \forall \, x \in G^{-1}(y) \implies t' \leq t\Big).$$

Lo anterior es una fórmula de primer orden con objetos definibles, por lo tanto se concluye. □

El resultado que sigue es un ejemplo de lo que es llamado desingularización de una función. Cabe destacar que no se trata de *remover* las singularidades de una función (los ceros de su gradiente), sino de cambiar las pendientes en una vecindad de los valores críticos. Esto se hace usando una función auxiliar cuyo propósito es *subir* los niveles de la función original, como fue mencionado más arriba.

Teorema 3.4 (Desigualdad de Kurdyka-Łojasiewicz versión 2. Ver [6]) *Sea $f : U \to \mathbb{R}$ una función definible y diferenciable, con U un abierto acotado de $\mathbb{R}^n$. Si $f > 0$ en todo U entonces existen constantes $c > 0$ y $\rho > 0$ y una función definible estrictamente creciente, positiva y C^1: $\Psi : \mathbb{R}_+ \to \mathbb{R}$ tal que:*

$$\|\nabla(\Psi \circ f)(x)\| \geq c$$

para cada $x \in U$ tal que $f(x) \in (0, \rho)$.

Demostración. Como f es definible también lo es ∇f y consecuentemente lo es $\|\nabla f\|$. Ahora a descartar un caso molesto, si $f^{-1}(r) = \emptyset$ para valores que acumulan en torno al 0, entonces el ínfimo de f sobre U es necesariamente mayor a una constante positiva y por ende, no habría nada que probar, pues se pierde el intervalo de interés. Por lo tanto el caso interesante es asumir que a partir de t suficientemente pequeño, $f^{-1}(r') \neq \emptyset$ en el intervalo $(0, r)$. Por lo anterior la siguiente función está bien definida para algún intervalo pequeño $(0, \varepsilon)$ y además es o-minimal:

$$\varphi(r) = \inf_{x \in f^{-1}(r)} \|\nabla f(x)\|.$$

El paso que sigue es mostrar que $\varphi(t) > 0 \; \forall \, t \in (0, \varepsilon)$ si ε es lo suficientemente pequeño. El razonamiento será por contradicción, suponiendo que no es así. Se define entonces el siguiente conjunto:

$$\Sigma = \{x \in U \, : \, \|\nabla f(x)\| < f(x)^2\}.$$

El conjunto anterior está definido por una fórmula de primer orden. Donde tanto f, como $\|\nabla f\|$ y f^2 son funciones definibles, por lo tanto el conjunto Σ es definible.
Entonces se tiene que en todos los intervalos $(0, \eta)$ con η suficientemente pequeño hay un valor $r \in (0, \eta)$ tal que $\varphi(r) = 0$. Se puede entonces definir una secuencia de valores $r_n \searrow 0$ tal que $\varphi(r_n) = 0 \quad \forall \, n \in \mathbb{N}$.
Como esto se está haciendo en puntos donde f admite preimagen se puede también elegir una secuencia $(x_n)_n \subset \Sigma$ tal que $f(x_n) = r_n$, es decir, $(x_n, r_n) \in \operatorname{gph}(f\big|_{\Sigma})$. Si a es un punto de acumulación de $(x_n)_n$, como $r_n \to 0$ se tiene que $(a, 0) \in \overline{\operatorname{gph}(f\big|_{\Sigma}) \setminus \{(a, 0)\}}$. A esto se puede aplicar el lema de selección de curvas para deducir la existencia de una curva definible:

$$\gamma : (-\delta, \delta) \to \mathbb{R}^n \times \mathbb{R},$$

que es además de clase C^1 y tal que $\gamma(0) = (a, 0)$ y $\gamma((0, \delta)) \subset \operatorname{gph}(f\big|_{\Sigma})$. Con una leve modificación se puede tomar tal intervalo simétrico, lo cual es útil para posteriormente acotar en norma, considerando el punto 0 está en el interior del intervalo. Lo mismo para la derivada de dicha función.
Como γ llega al grafo de f restringido a Σ, se puede escribir $\gamma(s) = (\alpha(s), (f \circ \alpha)(s))$. Esto permite definir $g(s) = (f \circ \alpha)(s)$ definida sobre el intervalo $(0, \delta)$, como $\alpha(s) \in \Sigma$ se tiene que $g(s) \to 0$ cuando $s \to 0$ y lo mismo para $g'(s)$. Por otro lado si δ es lo suficientemente pequeño, gracias al teorema de monotonía se puede asumir que g y g' son monótonas, más aún se comprueba que son crecientes. Nuevamente como $\alpha \in \Sigma$ se tiene que:

$$0 < g'(s) \leq C g(s)^2, \quad \forall \, s \in (0, \delta),$$

donde $C = \|\alpha'\|_\infty$ sobre $[0, \delta]$. Tomando δ levemente más pequeño de ser necesario, usando que α es C^1 y que en 0 también alcanza un valor. Además, aplicando el teorema de valor medio sobre g en $(0, s)$ se obtiene:

$$g(s) = s g'(s') \leq s g'(s).$$

Lo último gracias a que g es creciente. Combinando ambas desigualdades se obtiene finalmente:

$$0 < g'(s) \leq A s^2 (g'(s))^2, \; \forall \, s \in (0, \delta).$$

Esto último es una contradicción pues $\lim_{s\to 0} h'(s) \to 0$ y la desigualdad se puede reescribir como $0 < \frac{1}{C} \leq s^2 g'(s)$. Ahora entonces es un hecho que $\varphi(t) > 0$ en $(0, \varepsilon)$ si ε es lo suficientemente pequeño. Sea ahora:

$$\Gamma = \{x \in U \setminus f^{-1}(0) \,:\, f(x) < \varepsilon,\, \|\nabla f(x)\| \leq 2\varphi(f(x))\}.$$

El conjunto Γ es definible, pues su definición involucra una fórmula de primer orden con funciones definibles. Bajo el mismo razonamiento previo se deduce la existencia de $b \in \overline{U}$ tal que $(b, 0) \in \overline{\mathrm{gph}(f\big|_{\Gamma}) \setminus \{(b, 0)\}}$.
Como consecuencia del lema de selección de curvas en $\mathrm{gph}(f\big|_{\Gamma})$ en el punto $(b, 0)$ se obtiene una función definible $\zeta : (-\delta, \delta) \to \mathbb{R}^n$ de clase C^1 tal que $\zeta(0) = (b, 0)$ y $\zeta((0, \delta)) \subset \mathrm{gph}(f\big|_{\Gamma})$. Tal como antes se escribe $\zeta(s) = (\beta(s), (f \circ \beta)(s))$. Análogamente a lo anterior se define $h(s) = (f \circ \beta)(s)$. Por la definición de β se tiene que $\lim_{s\to 0} h(s) = 0$ y más aún $h(s) > 0$ en todo $(0, \delta)$.
Por una aplicación del teorema de monotonía, junto con el teorema de la función inversa, se tiene que si δ' es lo suficientemente pequeño la función $h : (0, \delta') \to \mathbb{R}$ es un difeomorfismo sobre su imagen, algún intervalo de la forma $(0, \rho)$. La función candidata para el resultado final será $\psi = h^{-1}$. A continuación se ve que es la adecuada:
Tomando δ' pequeño se puede acotar $\|\zeta'(s)\|$ uniformemente en $(0, \delta')$, puesto que ya es C^1 en un intervalo más grande $(-\delta, \delta)$. Dado $x \in U$ tal que $f(x) \in (0, \rho)$, si $s = \psi(f(x))$ entonces:

$$\|\nabla(\psi \circ f)(x)\| = \psi'(f(x))\|\nabla f(x)\|.$$

Como consecuencia de la definición del conjunto Γ se obtiene:

$$\nabla(\psi \circ f)(x) \geq \psi'(f(x))\frac{1}{2}\|\nabla f(\beta(s))\|$$

Como $\|\nabla f(\beta(s))\|\, \|\beta'(s)\| \geq \langle \nabla f(\beta(s)), \beta'(s)\rangle$ se tiene finalmente que:

$$\nabla(\psi \circ f)(x) \geq \frac{\psi'(f(x))}{2\tilde{C}}(f \circ \beta)'(s) = \frac{1}{2\tilde{C}},$$

para todo $x \in [0 < f < \delta']$. Lo que concluye el resultado. □

Definición 3.5 (Ver [6]) *Sean $U \subset \mathbb{R}^n$ abierto y $f : U \to \mathbb{R}$ diferenciable. Un valor $s \in \mathbb{R}$ se llama valor crítico de f si existe $x \in [f(x) = s]$ tal que $\nabla f(x) = 0$*

La siguiente definición extiende la noción de valores críticos de una función lo que será utilizado para el próximo teorema importante de la sección.

Definición 3.6 (Ver [6]) *Sean $U \subset \mathbb{R}^n$ abierto y $f : U \to \mathbb{R}$ diferenciable. Un valor $s \in \mathbb{R}$ se llama valor crítico asintótico de f si existe una sucesión $\{x_n\}_n \subset U$ tal que $f(x_n) \to s$ y $\nabla f(x_n) \to 0$.*

El siguiente resultado es una versión fuerte del teorema de Sard en el contexto de la geometría o-minimal. Dicho resultado permite apreciar el sencillo y poco patológico comportamiento de las funciones y en las estructuras o-minimales.

Teorema 3.7 (Teorema de Sard definible. Ver [6]) *Sean U es un abierto acotado de $\mathbb{R}^n$ y f una función definible. Entonces el conjunto de sus valores críticos, tanto usuales como asintóticos, es finito.*

Demostración. Si los valores críticos (asintóticos) de f son un conjunto discreto, entonces deben ser finitos, como consecuencia de la o-minimalidad de f, ya que el conjunto de puntos críticos asintóticos (escribiéndolos a través de una fórmula de primer orden y una función definible) es definible.
Entonces ahora basta probar que los valores críticos son aislados. Sea r un valor crítico de f. Como U es acotado, se puede aplicar el teorema anterior localmente en torno al valor crítico. El teorema arroja una desigualdad del estilo:

$$\|\nabla(\psi \circ f)(x)\| = \psi'(f(x))\|\nabla f(x)\| \geq c.$$

Acotando inferiormente el gradiente de f por una constante positiva y el inverso de la derivada de ψ, por ende f no tiene puntos críticos asintóticos en un conjunto de la forma $(r-\rho, r) \cup (r, r+\rho)$. Esto muestra que tales puntos son discretos y por lo tanto finitos. $\square$

El siguiente teorema es el primer resultado, sobre el comportamiento asintótico de las trayectorias acotadas que del sistema gradiente.
Pero antes hay que detallar conceptos importantes.

Observación Suponiendo que la función f es propia, es decir preimagen de un conjunto compacto bajo esta es un compacto, se tiene que todas las órbitas del sistema gradiente son curvas acotadas definidas sobre $[0, \infty)$.

Observación (Parametrizaciones alternativas) Uno podría reparametrizar las curvas, por ejemplo, en cuanto a longitud de arco:

$$s(t) = \int_a^t \|\dot{x}(t')\| \mathrm{d}t'.$$

De modo que el sistema gradiente reparametrizado por su longitud toma la forma:

$$\dot{y}(s) = -\frac{\nabla f(y(s))}{\|\nabla f(y(s))\|}.$$

Pero la reparametrización más adecuada para lo que sigue es de acuerdo a los niveles de la función, al igual que en el caso de arriba, se debe tener el reparo de que esto debe ser fuera de los puntos críticos.
Sea $[a, b)$ un intervalo de tiempo en el cual solo $x(b)$ es un punto crítico asociado a un valor crítico (r_∞, la función $\rho(t) = f(x(t))$ es un difeomorfismo entre (a, b) y $(r_\infty, r_0]$ con r_0 el valor de f en $x(a)$, por lo que reparametrizando como $x(r) = x(\rho^{-1}(r))$, se cumple la siguiente ecuación:

$$\dot{x}(r) = \frac{\nabla f(x(r))}{\|\nabla f(x(r))\|^2}.$$

Bajo esta parametrización de las órbitas, la longitud de tales se calcula entre dos niveles r_1 y r_2 de la función bajo la fórmula:

$$|x(a, b)| = \int_{r_1}^{r_2} \frac{1}{\|\nabla f(x(r))\|} \mathrm{d}r.$$

De lo anterior se ve explícitamente la relevancia de acotar inferiormente las normas de los gradientes. Esto será utilizado en el teorema que viene más adelante.

La siguiente definición será para precisar lo que será una órbita parametrizada por los valores de la función f en este contexto.

Definición 3.8 (Ver [6]) *Sea $f : U \subset \mathbb{R}^n$ una función C^1. Dado el sistema gradiente $\dot{x}(t) = -\nabla f(x(t))$, una trayectoria u órbita entre los valores $a, b \in \mathbb{R}$ es una función $\gamma : (a, b) \to U$ maximalmente diferenciable que satisface la ecuación diferencial:*

$$\dot{\gamma}(r) = \frac{\nabla f(\gamma(r))}{\|\nabla f(\gamma(r))\|^2}.$$

Para $(a', b') \subset (a, b)$, $|\gamma(a', b')|$ denota la longitud de la curva entre los instantes a' y b', es decir:

$$|\gamma(a', b')| = \int_{a'}^{b'} \|\gamma'(r)\| \mathrm{d}r.$$

Si una órbita está definida sobre un intervalo específico conocido, su longitud total de denotará simplemente por $|\gamma|$.

Teorema 3.9 (Ver [6]) *Sean U un conjunto abierto acotado de $\mathbb{R}^n$ y $f : U \to \mathbb{R}$ una función de clase C^1 y definible. Entonces:*

(i) *existe una constante $C > 0$ tal que el largo de todas las trayectorias del gradiente está acotado por C, es decir $\displaystyle\int_0^\infty \|\dot{x}(t)\| \mathrm{d}t \leq C$.*

(ii) *existe una función $\sigma : \mathbb{R}_+ \to \mathbb{R}_+$ continua, definible y estrictamente creciente tal que $\lim_{t\to 0} \sigma(t) = 0$, de modo que si γ es una trayectoria del gradiente y $a, < b$ son instantes de valores de f con $a < b$, entonces, utilizando la parametrización que se tiene por niveles se tiene la siguiente cota:*

$$|\gamma(a, b)| \leq \sigma(|f(\gamma(b)) - f(\gamma(a))|).$$

Demostración. Para lo que sigue se supondrá que f sea acotada. Puesto que si no lo fuera, se podría cambiar por $\phi \circ f$ donde $\phi(t) = \frac{t}{1+t^2}$, así la imagen estará contenida en $(-1, 1)$, y mejor aún, componer con tal ϕ no altera las trayectorias del sistema gradiente, y $\phi \circ f$ es o-minimal pues ϕ es semialgebraica y f también.
Tal como en una demostración anterior se considera la función $\varphi : (-1, 1) \to \mathbb{R}$ dada por:

$$\varphi(t) = \begin{cases} \inf\{\|\nabla f(x)\| \,:\, x \in f^{-1}(t)\}, & \text{si } f^{-1}(r) \neq \emptyset \\ 1, & \text{si } f^{-1}(r) = \emptyset \end{cases}$$

Esta función es o-minimal. Más aún, se puede interpretar como una aplicación que, a un nivel r, le asigna el peor caso posible de la norma del gradiente de f, y ese peor caso es precisamente el que se quiere controlar, esto porque:

$$|\gamma(a, b)| = \int_a^b \|\gamma'(r)\| \mathrm{d}r = \int_a^b \frac{1}{\|\nabla f(\gamma(r))\|} \mathrm{d}r \leq \int_a^b \frac{1}{\varphi(r)} \mathrm{d}r.$$

Lo que sigue será descomponer $(-1,1)$ en finitos sub-intervalos, en los cuales la función f se comporte adecuadamente por ser o-minimal, luego usar reiteradamente el teorema de valor medio para acotar la longitud de la curva.
Entonces aplicando el teorema de monotonía sobre la función φ se encuentran $r_0 = -1 < r_1 < ... < r_n = 1$ tales que f es C^1 en cada sub-intervalo $(r_{\mathrm{i}}, r_{\mathrm{i}+1})$, en donde φ adicionalmente va a ser constante o estrictamente monótona. Aquí la observación útil es que si φ es constante en $(r_{\mathrm{i}}, r_{\mathrm{i}+1})$ entonces no puede ser igual a 0, esto gracias a que los valores críticos asintóticos son un conjunto discreto (teorema **3.7**).
Se tienen las siguientes posibilidades para cada intervalo $(r_{\mathrm{i}}, r_{\mathrm{i}+1})$:

(i) Existe una cota inferior $C_{\mathrm{i}} > 0$ de φ, i.e. $\forall\, t \in (r_{\mathrm{i}}, r_{\mathrm{i}+1})\ \varphi(r) \geq C_{\mathrm{i}}$,

(ii) o bien r_{i} o $r_{\mathrm{i}+1}$ es un valor crítico asintótico de f, entonces aplicando el teorema anterior en el intervalo $(t_{\mathrm{i}}, t_{\mathrm{i}+1})$ se obtiene la existencia de una constante C_{i} y una función $\psi_{\mathrm{i}} : (r_{\mathrm{i}}, r_{\mathrm{i}+1}) : \mathbb{R}$ estrictamente creciente, acotada y C^1 tal que:

$$\|\nabla(\psi_{\mathrm{i}} \circ f)(x)\| \geq C_{\mathrm{i}}$$

para cada $x \in f^{-1}((r_{\mathrm{i}}, r_{\mathrm{i}+1}))$.

Entonces el razonamiento que sigue será tratar cada posibilidad por separado para concluir la desigualdad.

Sean γ una trayectoria de $-\nabla f$ y γ_{i} la trayectoria limitada al intervalo $(r_{\mathrm{i}}, r_{\mathrm{i}+1})$. Para el primer caso se tiene lo siguiente:

$$|\gamma_{\mathrm{i}}| \leq \frac{1}{C_{\mathrm{i}}} |\psi_{\mathrm{i}}(r_{\mathrm{i}}) - \psi_{\mathrm{i}}(r_{\mathrm{i}+1})|.$$

Esto último sale inmediatamente de que en $|\gamma(r_{\mathrm{i}}, r_{\mathrm{i}+1})| = \displaystyle\int_{r_{\mathrm{i}}}^{r_{\mathrm{i}+1}} \frac{1}{\|\nabla f(x(r))\|} \mathrm{d}r$.
Mientras que para el segundo caso la desigualdad de arregla como sigue:

$$\|\nabla(\psi_{\mathrm{i}} \circ f)(x)\| = \psi_{\mathrm{i}}'(f(x)) \|\nabla f(x)\| \geq C_{\mathrm{i}}\,,\ x \in U, f(x) \in (r_{\mathrm{i}}, r_{\mathrm{i}+1}).$$

Sigue que:

$$\frac{1}{\|\nabla f(x)\|} \leq \frac{1}{C_{\mathrm{i}}} |\psi_{\mathrm{i}}'(f(x))| \leq \frac{1}{C_{\mathrm{i}}} |\psi_{\mathrm{i}}(r_{\mathrm{i}}) - \psi_{\mathrm{i}}(r_{\mathrm{i}+1})|,$$

donde se usó la desigualdad de valor medio, junto con el hecho de que ψ es creciente.
De este modo:

$$|\gamma_{\mathrm{i}}| \leq \frac{1}{C_{\mathrm{i}}} |\psi_{\mathrm{i}}(r_{\mathrm{i}}) - \psi_{\mathrm{i}}(r_{\mathrm{i}+1})|.$$

Todo esto gracias a que la composición de f con las funciones ψ_{i} no alteran las trayectorias del sistema, finalmente al sumar sobre los intervalos $(r_{\mathrm{i}}, r_{\mathrm{i}+1})$ se llega a la cota para la longitud de γ:

$$|\gamma| = \sum_{\mathrm{i}=0}^{k-1} |\gamma_{\mathrm{i}}| \leq \frac{1}{C_{\mathrm{i}}} \sum_{\mathrm{i} \in I_1} |r_{\mathrm{i}} - r_{\mathrm{i}+1}| + \frac{1}{C_{\mathrm{i}}} \sum_{\mathrm{i} \in I_2} |\psi_{\mathrm{i}}(r_{\mathrm{i}}) - \psi_{\mathrm{i}}(r_{\mathrm{i}+1})| = C,$$

donde I_1 denota a los índices del primer caso e I_2 a los del segundo. Eso demuestra la primera parte del teorema.

La segunda parte es más breve:
Una función σ adecuada se construye a partir de lo que sigue:

$$\sigma_{\mathrm{i}}(r) = \begin{cases} (1/C_{\mathrm{i}}) \sup\{|\psi_{\mathrm{i}}(p) - \psi_{\mathrm{i}}(q)| \,:\, p, q \in (r_{\mathrm{i}}, r_{\mathrm{i}+1}),\, r = p - q\}, & \text{si i} \in I_1 \\ r/C_{\mathrm{i}}, & \text{si i} \in I_2 \end{cases}$$

Cada σ_{i} es continua, o-minimal y estrictamente creciente, y se puede extender con las mismas propiedades a todo $\mathbb{R}$, por lo que definiendo $\sigma = \sup_{\mathrm{i}} \sigma_{\mathrm{i}}$, mayora la desigualdad que cada una de las σ_{i} alcanza sobre su intervalo $(r_{\mathrm{i}}, r_{\mathrm{i}+1})$, por lo tanto, globalmente:

$$|\gamma(a,b)| \leq \sigma(|f(\gamma(b)) - f(\gamma(a))|).$$

Se verifica fácilmente que la función σ es definible, continua y estrictamente creciente.

$\square$

Capítulo 4

Desigualdad de Łojasiewicz en el caso no suave

Una generalización natural del sistema gradiente $\dot{x}(t) = -\nabla f(x(t))$, consiste en considerar una función no necesariamente diferenciable, conjuntamente con un subdiferencial con propiedades trabajables, por lo que un planteamiento razonable del sistema es a través de la inclusión diferencial:

$$\dot{x}(t) \in -\partial f(x(t)), \text{ en casi cada } t,$$

en donde la solución que se busca es una curva absolutamente continua $x(t)$.
Al igual que antes se buscará el cambio de coordenadas de acuerdo a los niveles de la función, para explicitar la relación entre los subgradientes y la longitud de las órbitas. Se verá que entonces que un sistema más general adopta la forma de un *sweeping process*:

$$\dot{x}(r) \in -N_{[f \leq r]}(x(r)), \text{ en casi cada } r.$$

Se buscará estimar la longitud de las órbitas, a través de una generalización no suave de la desigualdad de tipo Łojasiewicz. Se recuerda que la desigualdad de Łojasiewicz se puede escribir de la siguiente forma (de hecho incluye el caso analítico).

Proposición 4.1 (Ver [1]) *Si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es analítica (respectivamente semialgebraica y de clase C^1) y $\bar{x} \in U$ es un punto crítico de f entonces existe $\theta \in [0, 1)$ tal que la función*

$$\frac{|f - f(\bar{x})|^\theta}{\|\nabla f\|}$$

es acotada en torno al punto $\bar{x}$.

Observación Lo anterior es consecuencia del teorema **1.14**, a través del teorema **3.7** que afirma que los valores críticos son aislados.

Evidentemente, teniendo tal resultado para los puntos críticos, se cumple que las curvas tienen longitud finita, pues se puede aplicar el teorema **3.9** con la función

$$\Psi(r) = \frac{(r - f(\bar{x}))^{1-\theta}}{1-\theta},$$

que es una función estrictamente creciente y C^1, para acotar la longitud de la curva entre niveles r_1 y r_2 de la función:

$$|x| = \int_{r_1}^{r_2} \frac{1}{\|\nabla f(x(r))\|} \mathrm{d}r.$$

Tal como se hizo antes, se obtiene:

$$\|\nabla(\Psi \circ f)(x)\| \geq C,$$

lo que es equivalente a:

$$\Psi'(f(x))\|\nabla f(x)\| \geq C.$$

Reemplazando derivada de la función Ψ se obtiene que:

$$|f(x) - f(\bar{x})|^{-\theta}\|\nabla f(x)\| \geq C,$$

para finalmente recordenar la desigualdad como:

$$\frac{1}{C} \geq \frac{|f - f(\bar{x})|^{\theta}}{\|\nabla f(x)\|}.$$

El paso siguiente será tener la desigualdad de Łojasiewicz para f no suave, e hipótesis menos fuertes como exigir que sea semicontinua inferior, suponiendo que la función sea o-minimal. Para esto será necesario redefinir lo que se entiende por punto crítico asintótico, y extender los resultados adaptando la forma de utilizar las herramientas de geometría o-minimal.

Observación (Ver [8]) Dejando de lado por un momento la o-minimalidad, el hecho que la proposición **4.1** sea válida para funciones analíticas podría inducir a pensar que también es cierto para funciones de clase C^∞. Este no es el caso, el contraejemplo de hecho está basado en la clásica función C^∞ pero no analítica, que a x le asigna $\mathrm{e}^{\frac{1}{x(1-x)}}$ sobre $(0,1)$ y 0 fuera de dicho intervalo. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ definida en coordenadas polares como sigue:

$$f(r\cos(\theta), r\sin(\theta)) = \begin{cases} \mathrm{e}^{\frac{1}{r^2-1}}, & \text{si } r < 1 \\ 0, & \text{si } r = 1 \\ \mathrm{e}^{\frac{-1}{r^2-1}} \sin(\frac{1}{r-1} - \theta), & \text{si } r > 1 \end{cases}$$

Se puede demostrar que existen trayectorias del sistema gradiente inducido por f, de modo que su ω-límite es el círculo centrado en 0 y de radio 1.
Cabe recordar que el ω-límite se puede definir como el conjunto de puntos de acumulación de la trayectoria cuando $t \to \infty$, y de hecho consiste exclusivamente en los puntos críticos de f. Por lo tanto, dicha trayectoria tiene longitud infinita acercándose a un conjunto de cardinalidad del continuo de puntos críticos (pero aún de medida 0).

Volviendo a la generalización del sistema gradiente y sus resultados, a continuación se exponen las primeras versiones más generales de lo visto en la sección previa:

Definición 4.2 (Pendiente no suave. Ver [1]) *Sea $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ una función s.c.i., se define su pendiente no suave como sigue:*

$$m_f(x) = \inf\{\|v\| \,:\, v \in \partial f(x)\},$$

con la convención usual que $\inf\limits_{\emptyset} = \infty$.

La pendiente no suave mide el peor caso para las normas de los subgradientes, por lo tanto lo razonable será buscar acotarla inferiormente tal como se hizo para la norma de los gradientes en la sección previa.

Definición 4.3 (Puntos críticos generalizados. Ver [1]) *Dada una función semicontinua inferior $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, el conjunto de sus puntos críticos generalizados se define como:*

$$\mathrm{crit}(f) := \{x \in \mathbb{R}^n \,:\, 0 \in \partial f(x)\}.$$

Observación Si $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ es semicontinua inferior y convexa, como el grafo del subdiferencial ∂f es cerrado, $\mathrm{crit}(f)$ es también un conjunto cerrado.
Más aún m_f es una función semicontinua inferior, puesto que siendo cerrado el grafo del subdiferencial, el epígrafo de m_f es cerrado, y en este caso se tiene que $\mathrm{crit}(f) = m_f^{-1}(0)$

Proposición 4.4 (Ver [1]) *Sea $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ una función. Si f es definible, entonces las multiaplicaciones $\hat{\partial} f$, ∂f y $\partial_c f$ son definibles (respectivamente subdiferenciales de Fréchet, límite y de Clarke). Como consecuencia de lo anterior, m_f es una función definible y por lo tanto* $\mathrm{crit}(f)$ *es un conjunto definible.*

Demostración. En esta demostración se verán de paso varios resultados que son útiles de recordar para lo que sigue más adelante.
Primero, si $Q \subset \mathbb{R}^n$ es un conjunto o-minimal entonces el cono normal sobre un punto $\bar{x} \in Q$ también lo es, transformando la fórmula en una de primer orden:

$$\begin{aligned} N_Q(\bar{x}) &= \Big\{ v \in \mathbb{R}^n \,|\, \langle \nu, x - \bar{x} \rangle \le o(|x - \bar{x}|)\,,\ \forall\, x \in Q \Big\} \\ &= \Big\{ v \in \mathbb{R}^n \,|\, \limsup_{\substack{x \to \bar{x} \text{ en } Q \\ x \ne \bar{x}}} \frac{\langle v, x - \bar{x} \rangle}{|x - \bar{x}|} \le 0\,,\ \forall\, x \in Q \Big\}. \end{aligned}$$

Como el límite superior se puede descomponer en un ínfimo y un supremo sobre la intersección de Q con una bola sin un punto, siendo todo esto o-minimal, la fórmula que define al cono normal sobre un conjunto definible, es de primer orden y por lo tanto tal cono es o-minimal.
Recordando entonces que:

$$\hat{\partial} f(x) = \{v \in \mathbb{R}^n \,|\, (v, -1) \in \hat{N}_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x}))\},$$

es inmediato ver que el subdiferencial en un punto es un conjunto o-minimal.
Por otro lado viendo al subdiferencial como una multiaplicación:

$$\mathrm{gph}(\hat{\partial} f) = \Big\{(x, v) \in \mathbb{R}^n \times \mathbb{R}^n \, : \, v \in \hat{\partial} f(x)\Big\}.$$

Se ve inmediatamente que el conjunto anterior es o-minimal, deglosando la pertenencia del vector $v \in \mathbb{R}^n$ al subdiferencial en una fórmula de primer orden.
Similarmente se hace para el cono y subdiferencial límite, pero la demostración es bastante más larga y tediosa. Como el cono normal de Clarke se obtiene tomando adherencia de la envoltura convexa al cono límite, será o-minimal cuando este último lo sea. En efecto, como ya se vio, adherencia de un conjunto definible da lugar a un conjunto definible. Se verifica que sucede lo mismo para envolturas convexas, esto mediante el teorema de Caratheodory, este establece que si un punto x está en la envoltura convexa de un conjunto $X \subset \mathbb{R}^n$, entonces existe un subconjunto de X, X' que consiste en a lo más $n+1$ puntos, tal que x está en la envoltura convexa de X'.
Teniendo entonces que ∂f es o-minimal, el siguiente conjunto:

$$\mathrm{gph}(m_f) = \Big\{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \inf\{\alpha = \|v\| \, : \, v \in \partial f(x)\Big\}.$$

También es o-minimal, pues ya está establecido que descomponer el ínfimo en su definición arroja una fórmula de primer orden.
De igual modo

$$\mathrm{crit}(f) := \{x \in \mathbb{R}^n \, : \, 0 \in \partial f(x)\},$$

es o-minimal, lo que concluye el resultado. □

Lo que sigue será generalizar los teoremas de la capítulo anterior, en el marco del análisis no diferenciable.

Proposición 4.5 (Ver [1]) *Sea $f : \mathbb{R}^n \to \mathbb{R}$ función semialgebraica y continua. Sea $a \in \mathbb{R}^n$ un punto crítico generalizado de f. Entonces existe $\theta \in [0, 1)$ tal que la función:*

$$\frac{|f - f(a)|^\theta}{m_f}$$

está acotada en torno al punto a.

Demostración. Basta recurrir nuevamente a la desigualdad de Łojasiewicz para el caso semialgebraico. La función m_f es una función semialgebraica cuando f lo es. Tal como en el caso anterior, en un entorno suficientemente pequeño en torno al punto a, si $f(x) - f(a)$ se anula, m_f se anula. □

A continuación se presenta otra versión del teorema de Sard, el cual es importante si se quiere probar una desigualdad de Łojasiewicz, puesto que debe ser válida en cada intervalo de valores regulares con extremos posiblemente valores críticos.

Teorema 4.6 (Teorema de Sard no suave - caso definible. Ver [1]) *El conjunto de los valores críticos generalizados de una función $f : U \subset \mathbb{R}^n \to \mathbb{R}$ definible y semicontinua inferior, es finito.*

El teorema que sigue entonces es el caso no suave de la desigualdad de Łojasiewicz. Una vez demostrado, se pueden sacar las conclusiones deseables sobre las órbitas del sistema subgradiente.

Teorema 4.7 (Ver [1]) *Sean $U \subset \mathbb{R}^n$ un conjunto abierto, acotado y no vacío y $f : U \subset \mathbb{R}^n \to \mathbb{R}$ una función definible y semicontinua inferior. Si $f > 0$ en todo U entonces existen constantes $C > 0$ y $\rho > 0$ y una función $\Psi : \mathbb{R}_+ \to \mathbb{R}$ positiva, estrictamente creciente y de clase C^1 tal que:*

$$\Psi'(f(x))m_f(x) \geq C$$

para cada $x \in U$, $f(x) \in (0, \rho)$. Al igual que antes el teorema se adapta para usarse localmente en torno a valores críticos generalizados de la función f.

Demostración. La prueba es bastante similar a la versión C^1, por lo que se indicarán las modificaciones pertinentes:
Primero se hace la observación de que existe $\varepsilon > 0$ tal que la siguiente función en el intervalo $(0, \varepsilon)$ solo consite en valores regulares, que rescata el peor caso:

$$\varphi(r) = \text{ínf}\{m_f(x) \,:\, x \in f^{-1}(r)\}.$$

Claramente tal función es o-minimal.
Lo que sigue es verificar que φ es positiva en un intervalo $(0, \varepsilon')$, con $\varepsilon' < \varepsilon$. Esto se hace considerando el siguiente conjunto:

$$\{x \in U \,:\, m_f(x) < (f(x))^2\},$$

el que es evidentemente definible, y construido de manera que, al aplicarle el lema de selección de curvas al grafo de la función f restringida a este, se construye una nueva función de la que deriva una contradicción.
Cabe notar que una posible complicación saldría de la no diferenciabilidad de f, pero al tener la curva $\gamma_1(s) = (\gamma(s), (f \circ \gamma)(s))$ que da el lema, la contradicción aparece al construir desigualdades sobre la función $(f \circ \gamma)$, que a su vez aparecen por el teorema de valor medio. Nada de eso es problema, ya que, gracias al teorema de monotonía, como $(f \circ \gamma)$ es una función real de una variable y definible, entonces sobre un intervalo suficientemente pequeño es de clase C^1. Esto último es consecuencia directa de elegir el primer intervalo de la división que entrega el teorema de monotonía (teorema **1.19**).

Teniendo probado lo anterior el paso siguiente era trabajar con el conveniente conjunto:

$$\{x \in U \setminus f^{-1}(0) \,:\, f(x) < \varepsilon\,,\; m_f(x) \leq 2\varphi(f(x))\}.$$

Dicho conjunto también es o-minimal: nuevamente se aplica el lema de selección de curvas al grafo de f restringida en el conjunto, y con pasos similares a los del párrafo anterior, se obtiene una curva $\delta_1(s) = (\delta(s), (f \circ \delta)(s))$. Al igual que antes, la no diferenciabilidad no es un problema gracias al teorema de monotonía. Tal como en la demostración del teorema **3.4** se construye un difeomorfismo $\Psi(r)$ para r en un intervalo de la forma $(0, \rho)$. Igual que antes $g(s) = (f \circ \delta)(s)$, y $\Psi(r) = g^{-1}(r)$.

El paso final es comprobar que Ψ es la función buscada. Para esto la clave es acotar $\|\delta\|$ en el intervalo donde está definida, de modo que para cada $x \in U$ con $r = f(x) \in (0, \rho)$ y $s = \Psi(r) = g^{-1}(r)$ y utilizando una cota superior C de $\|\delta'(s)\|$:

$$\Psi'(f(x))m_f(x) \geq \Psi'(t)\frac{1}{2}m_{f\circ\delta}(s) \geq \frac{\Psi'(t)}{2C}(f\circ\delta)'(s) = \frac{1}{2C}.$$

Se usa simplemente que la curva se construyó por el lema de selección de curvas sobre el conjunto elegido más arriba junto con que $f \circ \delta$ es C^1 en un intervalo suficientemente pequeño. □

Lo que sigue es aplicar el teorema **4.7** al sistema dinámico subgradiente, para obtener un resultado sobre sus trayectorias. Aunque primero hay que aclarar cuáles son las trayectorias (órbitas) de interés en este caso.

Definición 4.8 (Ver [1]) *Una trayectoria (órbita) del sistema subgradiente, es una curva absolutamente continua $x : [0,T) \to \mathbb{R}^n$ que satisface las siguientes condiciones:*

$$\begin{cases} 0 \in \dot{x}(t) + \partial f(x(t)) & \textit{c.t.p. en } (0,T) \\ \partial f(x(t)) \neq \emptyset & \forall\, t \in [0,T) \end{cases}$$

Una trayectoria x se llamará maximal si no tiene una extensión sobre su dominio tal que se sigan satisfaciendo las condiciones del sistema.

Observación Un par de ideas a tener en mente son las siguientes:

(i) Como las trayectorias estudiadas son curvas absolutamente continuas, su longitud se puede definir tal cual como en la sección anterior:

$$|x(t_0,t)| = \int_{t_0}^{t} \|\dot{x}(s)\| \mathrm{d}s.$$

(ii) Cabe recordar que las funciones absolutamente continuas son diferenciables casi en todas partes y que son recuperables de la integración de su derivada salvo una constante.

El siguiente teorema garantiza la existencia de trayectorias bajo condiciones razonables. Cabe mencionar que se hace necesario por estar lidiando con inclusiones diferenciales, a diferencia del caso anterior en el sistema dinámico era una EDO común.

Teorema 4.9 (Existencia y unicidad de órbitas. Ver [1]) *Sea $f : \mathbb{R}^n \to \mathbb{R}$ satisfaciendo lo siguiente:*

(i) *f es semicontinua inferior y convexa.*

(ii) *f está acotada inferiormente.*

Entonces para cada $x_0 \in \mathbb{R}^n$ tal que $\partial f(x_0) \neq \emptyset$ existe una única curva absolutamente continua $x : [0,T) \to \mathbb{R}^n$, $T \in (0,\infty)$, que es solución maximal del sistema:

$$\begin{cases} 0 \in \dot{x}(t) + \partial f(x(t)) & \textit{c.t.p. en } (0,T) \\ \partial f(x(t)) \neq \emptyset & \forall\, t \in [0,T) \\ x(0) = x_0 & \end{cases}$$

Más aún la función $f \circ x$ es absolutamente continua.

A continuación se presentan corolarios de gran utilidad para varios desarrollos próximos.

Corolario 4.10 (Ver [1]) *Sea $x : [0,T) \to \mathbb{R}^n$ una trayectoria del sistema satisfaciendo las dos condiciones del teorema anterior. Se cumple lo siguiente:*

(i) *Para casi todo $t \in (0,T)$:*

$$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ x)(t) = \langle \dot{x}(t), v \rangle, \;\; \forall\, v \in \partial f(x(t)).$$

(ii) *Para casi todo $t \in (0,T)$ la aplicación $v \mapsto \langle \dot{x}(t), v \rangle$ es constante en $\partial f(x(t))$.*

(iii) *La trayectoria x puede extenderse a una trayectoria maximal $\bar{x} \in W^{1,2}(\mathbb{R}_+, \mathbb{R}^n)$.*

Demostración. Las funciones $t \mapsto x(t)$ y $t \mapsto f(x(t))$ son absolutamente continuas. Consecuentemente se puede tomar un conjunto común donde ambas son diferenciables, de la forma $(0,T) \setminus \mathcal{N}$, donde $\mathcal{N}$ tiene medida nula. Entonces para cada $t \in (0,T) \setminus \mathcal{N}$, se tiene $x(t) \in \mathrm{d}om\partial f$ y $\hat{\partial} f(x(t)) = \partial f(x(t))$ gracias a la convexidad de f. Entonces sigue que:

$$\partial(f \circ x)(t) = \{(f \circ x)'(t)\} = \left\{\frac{\mathrm{d}}{\mathrm{d}t}(f \circ x)(t)\right\} = \{\langle \dot{x}(t), v \rangle\,,\; v \in \partial f(x(t))\}.$$

De aquí se obtienen las dos primeras conclusiones. Para la tercera, basta notar que:

$$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ x)(t) = -\|\dot{x}(t)\|^2\,,\; \forall t \in (0,T) \setminus \mathcal{N}.$$

De modo que f es Lyapunov para el sistema dinámico y:

$$\int_0^T \|\dot{x}(t)\|^2 \mathrm{d}t = f(x_0) - f(x(T)) < \infty.$$

Así se ve que $\dot{x} \in L^2((0,T); \mathbb{R}^n)$, por lo tanto $x \in W^{1,2}((0,T); \mathbb{R}^n)$. Como el grafo de ∂f es cerrado, entonces $x(T) \in \mathrm{d}om(\partial f)$, por lo tanto gracias al resultado de existencia y unicidad anterior, se puede extender la trayectoria anterior a un intervalo $[0, T+\delta)$. Esto demuestrea que la extensión maximal de x está definida en $(0, \infty)$. □

Corolario 4.11 (Ver [1]) *Sea x una trayectoria maximal del sistema anterior con condición inicial $x(0) = x_0 \in \mathrm{dom}(\partial f)$. Entonces para casi todo $t \in \mathbb{R}_+$ se cumple lo siguiente:*

(i) $\|\dot{x}(t)\| = m_f(x(t))$.

(ii) $\frac{\mathrm{d}}{\mathrm{d}t}(f \circ x)(t) = -(m_f(x(t)))^2$. *Es decir, f decrece a lo largo de sus trayectorias.*

Se enuncia por fin el teorema conclusivo de la sección, que extiende en gran medida el resultado análogo suave citado anteriormente.
De hecho la demostración será algo distinta, pudiendo hacerse como en la sección anterior.

Teorema 4.12 *Sea f una función convexa, semicontinua inferior, acotada inferiormente y además o-minimal. Entonces cualquier trayectoria acotada tiene longitud finita y por lo tanto converge a un punto crítico de f.*

Demostración. Sea x una trayectoria maximal acotada. Gracias al corolario **4.10** la trayectoria queda definida en todo $\mathbb{R}_+$. Además por el corolario **4.11** y el hecho de que f se supone acotada inferiormente se tiene que:

$$\lim_{t\to\infty} f(x(t)) = \beta.$$

Por otro lado, reemplazando f por $f-\beta$, como no se altera el subdiferencial, se puede asumirse que $\beta = 0$.
Continuando con las simplificaciones se puede asumir adicionalmente que $f(x(t)) \neq 0$ para todo $t > 0$, esto es producto de que si una trayectoria se encuentra con un punto crítico generalizado, *i.e.* $m_f(x(t_0)) = 0$, el corolario **4.11** garantiza que la trayectoria se detiene, esto es que $x(t) = x(t_0)$ para todo $t \geq t_0$. Más aún, gracias a que la trayectoria está acotada, su ω-límite es no vacío, es decir existe una secuencia de tiempos $(t_n)_n$ con $t_n \to \infty$ tal que:

$$\lim_{t_n\to\infty} x(t_n) = a.$$

Notar que $f(a) = 0$ por el hecho de que $(f \circ x)$ es continua. Como ∂f tiene grafo cerrado se deduce que $a \in \mathrm{d}om(\partial f)$.
Hasta ahora no hay evidencia de que a sea o no un punto crítico de f. De cualquier modo se tiene la desigualdad de Łojasiewicz. Sigue que existen constantes $\varepsilon > 0$, $C > 0$, $\rho > 0$ y una función $\Psi : \mathbb{R}_+ \to \mathbb{R}$ positiva, estrictamente creciente y C^1 tal que:

$$m_f(x) \geq C\Psi'(f(x)),\ \ \forall\, x \in B(a,\varepsilon) \text{ tal que } f(x) \in (0,\rho).$$

Ahora bien, la función $\tilde{h} = \Psi(f \circ x)$ es absolutamente continua y como Ψ es estrictamente creciente, también lo es su inversa, esto junto con que f decrece a lo largo de las trayectorias resulta en que $x(t) \to a$ y $\tilde{h}$ es estrictamente decreciente y converge a 0 cuando $t \to \infty$. Sigue que existe un instante t_0 a partir del cual:

$$\frac{\tilde{h}(t) - \tilde{h}(t_0)}{C} \leq \frac{\varepsilon}{3}$$

y también:

$$\|x(t_0) - a\| \leq \frac{\varepsilon}{3}.$$

Se define el siguiente instante, al estilo de un tiempo de parada:

$$T_{out} = \inf\{t \geq t_0 \,:\, x(t) \notin B(a,\varepsilon)\}.$$

Algo simple de ver es que por continuidad de la trayectoria $t_0 < T_{out} \leq \infty$, algo no tan simple es que $T_{out} = \infty$, esto después de hacer unos cálculos previos:
Es fácil ver de todo el siguiente cálculo da una cota útil:

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{h}(t) = (\Psi^{-1})'(f\circ x)(t)\frac{\mathrm{d}}{\mathrm{d}t}(f\circ x)(t) \leq -\Psi'(f\circ x)(t)(m_f(x(t)))^2 \leq -Cm_f(x(t)),$$

en donde se usó que $(f \circ x)(t) = (m_f(x(t)))^2$ y que $-\Psi'(f(x))m_f(x) \leq -C$. Con esto se obtiene la siguiente cota:

$$\int_{t_0}^{t} m_f(x(s))\mathrm{d}s \leq \frac{1}{C}\int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}t}\tilde{h}(s)\mathrm{d}s = \frac{\tilde{h}(t)-\tilde{h}(t_0)}{C} \leq \frac{\varepsilon}{3}$$

Pero nuevamente, por el corolario **4.11**, en particular el hecho que $\|\dot{x}(t)\| = m_f(x(t))$ c.t.p. se cumple lo siguiente:

$$\int_{t_0}^{t} \|\dot{x}(s)\|\mathrm{d}s \leq \frac{\varepsilon}{3} \ \forall t \in [t_0, T_{out}].$$

Finalmente si T_{out} fuese finito entonces:

$$\|x(T_{out}) - a\| \leq \left|\left(x(t_0) + \int_{t_0}^{T_{out}} \|\dot{x}(s)\|\mathrm{d}s\right) - a\right| \leq \frac{2\varepsilon}{3}.$$

Lo anterior contradice el hecho de que en el tiempo T_{out} la trayectoria sale de la bola, por lo que finalmente se concluye que:

$$\int_{t_0}^{\infty} \|\dot{x}(s)\|\mathrm{d}s \leq \frac{\varepsilon}{3}.$$

Esto es que la trayectoria tiene longitud finita y por lo tanto converge. Así $\lim_{t\to\infty} x(t) = a$ y $m_f(x(t))$ tiene a 0 como punto límite. Nuevamente por el hecho de que el grafo de ∂f es cerrado se concluye de que a es un punto crítico de f. $\square$

Capítulo 5

Sweeping process y desingularización

Haciendo un breve recuento de lo realizado hasta ahora, el primer sistema dinámico de interés fue $\dot{x} = -\nabla f(x)$. Ya se vieron distintos casos en que sus órbitas tienen longitud finita y por lo tanto convergen a puntos críticos de f. Esto se pudo concluir gracias a la desigualdad de Łojasiewicz o a su generalización por Kurdyka en el caso o-minimal general.
El paso siguiente, que se vio en el capítulo anterior, fue el de generalizar el sistema al caso no diferenciable. Eso resultó en estudiar la inclusión diferencial $\dot{x} \in -\partial f(x)$ c.t.p., nuevamente probando una versión de la desigualdad de Łojasiewicz no suave, para luego usarla en concluir que la longitud de las trayectorias acotadas es finita y las órbitas convergen a un punto crítico.

En este capítulo se utiliza un enfoque aún más amplio de lo ya hecho. Primero que todo cabe observar que bajo una reparametrización adecuada, a través de los niveles de la función, el sistema subgradiente se puede transformar en otra inclusión diferencial:

$$\dot{x}(r) \in -N_{[f \leq r]}(x(r))\,,\ r \in [0, \eta) \text{ c.t.p..}$$

Esto motiva estudiar sistemas más generales del estilo:

$$\dot{x}(r) \in -N_{S(r)}(x(r))\,,\ r \in [a, b) \text{ c.t.p.,}$$

en donde S es una multiaplicación. La inclusión diferencial anterior se llama *sweeping process* y fue introducido por Moreau.

En lo que sigue del capítulo se hará, a grandes rasgos, el proceso ya explicado anteriormente, encontrar una desigualdad para algún tipo de derivada de multiaplicaciones o-minimales. En este caso será la coderivada o más bien su norma exterior. Luego esto se relacionará con las trayectorias del *sweeping process* para poder concluir, bajo las hipótesis adecuadas, que tienen longitud finita.

Entonces como ya se mencionó el primer objeto de interés será la coderivada de Clarke $D_c^*S(\bar{x}|\bar{y})$, definida en el final del primer capítulo. La primera definición será la de valor crítico asintótico basada en el concepto anterior.

Definición 5.1 (Ver [5]) *Sea $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ una multiaplicación. Un valor $\bar{z} \in S(\mathbb{R}^n)$ se llama valor crítico asintótico de Clarke de la multiaplicación S, sobre un conjunto $\mathcal{U} \subset \mathbb{R}^n$, si existe una secuencia $((x_n, z_n)) \subset gph(S)$ en donde $(z_n)_n \subset \mathcal{U}$ tal que $y_n \to \bar{z}$ y:*

$$|D_c^* S^{-1}(y_n|x_n)|^+ \to \infty.$$

A continuación se presenta nuevamente un teorema de tipo Sard para dichos valores críticos.

Teorema 5.2 (Teorema de Sard para multiaplicaciones. Ver [5]) *Sea $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ una multiaplicación definible de grafo cerrado. El conjunto de valores críticos asintóticos de Clarke asociado es un conjunto definible de dimensión a lo más $m-1$.*

Teorema 5.3 (Ver [5]) *Sean $S : \mathbb{R} \rightrightarrows \mathbb{R}^n$ una multiaplicación definible con grafo cerrado y $x : (a, b) \to \mathbb{R}$ una curva absolutamente continua que satisface la inclusión de sweeping. Entonces se satisface la siguiente estimación:*

$$\|\dot{\gamma}(r)\| \le |D_c^* S(r|\gamma(r))|^+, \quad c.t.p.\ para\ r \in (a, b).$$

Este teorema puede resultar bastante útil si se busca acotar la longitud de una trayetoria, lo que será el caso si la norma exterior de la coderivada resulta ser integrable.

Definición 5.4 (Talweg. Ver [5]) *Sean $T : (a, b) \rightrightarrows \mathbb{R}^n$ y $\mathcal{U} \subset \mathbb{R}^n$. Se define el talweg de la multiaplicación T en $\mathcal{U}$ como la función $\varphi : (a, b) \to \overline{\mathbb{R}}$ dada por:*

$$\varphi(r) := \sup_{x \in S(r) \cap \mathcal{U}} |D_c^* T(r|x)|^+.$$

Esta definición debiese recordar a la función φ definida en capítulos anteriores, véase la demostración del teorema **4.7**. Recordando lo hecho previamente, dicha función φ era la que en cada nivel t de la función f, entregaba el peor caso de la norma de la derivada. Aquí no es muy diferente, se mostrará más adelante que el inverso de la coderivada está relacionado con el subdiferencial.

Lema 5.5 (Ver [5]) *Sea $T : (a, b) \rightrightarrows \mathbb{R}^n$ una multiaplicación definible a valores acotados, y $\mathcal{U} \subset \mathbb{R}^n$ un conjunto acotado. Entonces existe $\varepsilon > 0$ tal que la función talweg de la multiaplicación T en $\mathcal{U}$ es finita sobre $(a, a + \varepsilon)$.*

Teorema 5.6 (Ver [5]) *Sea $T : (a, b) \rightrightarrows \mathbb{R}^n$ una multiaplicación definible, a conjuntos no vacíos y cerrados. Entonces su función talweg es integrable (definida sobre un conjunto no vacío, acotado y definible $\mathcal{U} \subset \mathbb{R}^n$), esto es:*

$$\int_a^b \varphi(r) \mathrm{d}r < \infty.$$

Demostración. La demostración sigue el espíritu de lo hecho previamente, solo se mostrarán los pasos clave:
Primero definirse el conjunto adecuado:

$$\mathcal{V}(r) = \left\{ x \in S(r) \cap U \ : \ |D_c^* S(r|x)^+ \geq \frac{1}{2}\varphi(r) \right\}, \ r \in (a,b).$$

Notar que todo es definible, y que esos conjuntos son no vacíos salvo para finitos valores r. Mediante el lema de selección de curvas se obtiene una curva $r \mapsto \delta(r)$ tal que $\theta(r) \subset \mathcal{V}(r)$ para cada $r \in (a,b)$. Ocupando una estratificación de Whitney del grafo de S, se puede mostrar que:

$$t = \langle \dot{\theta}(r), D_c^* S^{-1}(\theta(r)|r)(t) \rangle$$

para r c.t.p. en (a,b).
Por otro lado, ocupando la fórmula para la norma exterior de la inversa de una multiaplicación. Se tiene que:

$$\inf_{|t|=1} \mathrm{d}ist(0, D_c^* S^{-1}(\theta(r)|r)(t)) = \frac{1}{|D_c^* S(r|\theta(r))|^+}.$$

Si ni 1 ni -1 están en el dominio de $S^{-1}(\theta(r)|r)$, entonces, recordando que θ era la curva obtenida mediante el lema de selección de curvas se tiene que:

$$\frac{1}{2}\varphi(r) \leq |D_c^* S(r|\theta(r))|^+ = 0.$$

Mientras que en el buen caso, para los r tal que $|D_c^* S(r|\theta(r))|^+ \neq 0$, se obtiene de las ecuaciones anteriores que:

$$1 \leq \|\dot{\theta}(r)\| \inf_{|t|=1} \mathrm{d}ist(0, D_c^* S^{-1}(\theta(r)|r)(t)) = \frac{\|\dot{\theta}(r)\|}{|D_c^* S(r|\theta(r))|^+} \leq \frac{2}{\varphi(r)}\|\dot{\theta}(r)\|.$$

□

Como es inferible, del resultado anterior se obtiene la conclusión esperable.

Corolario 5.7 (Ver [5]) *Si $T : \mathbb{R} \rightrightarrows \mathbb{R}^n$ es una multiaplicación a conjuntos definibles con grafo cerrado y (a,b) es un intervalo no necesariamente acotado de $\mathbb{R}$ entonces toda curva absolutamente continua $\gamma : (a,b) \to \mathbb{R}^n$ que satisface la inclusión de sweeping:*

$$\dot{\gamma}(r) \in -N^c_{S(r)}(\gamma(r)),$$

c.t.p. en (a,b), tiene longitud finita.

Todavía falta concluir el teorema principal, siguiendo la misma línea que en los capítulos anteriores.

Teorema 5.8 (Ver [5]) *Sean $S : \mathbb{R} \rightrightarrows \mathbb{R}^n$ una multiaplicación a definible y de grafo cerrado, $U \subset \mathbb{R}^n$ un conjunto acotado. Se tiene que para cada $a \in \mathbb{R}$ existe $b > a$ y una función estrictamente creciente y continua $\sigma : [a,b) \to \mathbb{R}$ que es además C^1 en (a,b), $\sigma(a) = 0$ y tal que se satisface lo siguiente:*

$$|D_c^*(S \circ \sigma)(r|x)|^+ \leq 1$$

para cada $r \in (a,b)$ y cada $x \in S(\sigma(r)) \cap U$.

Demostración. Puesto que la demostración es bastante similar a lo hecho antes, por lo que se mencionaran sólo los pasos clave:
Primero notar que si no existe $b > a$ tal que $(a, b) \cap \mathrm{dom} S \neq \emptyset$ entonces no hay nada que probar. Por lo tanto se asume la existencia de un b cumpliendo esa condición.
Definir el talweg φ sobre el conjunto $S(U)$ restringido al intervalo de interés (a, b). En este caso el talweg se puede suponer continuo en (a, b) pueso que es definible. Por otro lado si φ es 0 cerca de a el teorema también es trivial, por lo que se supone distinto de 0 en todo (a, b). Así se puede definir:

$$\Phi(r) = \int_a^r \varphi(t)\mathrm{d}t, \ r \in [a, b).$$

Se ve que tal función está bien definida, es estrictamente creciente y C^1 en (a, b) y se muestra que:

$$|D_c^*(S \circ \Psi(t|x))|^+ = \frac{|D_c^* S(\Psi(t)|x)|^+}{\varphi(\Psi(t))} \leq 1,$$

con lo cual se concluye. $\square$

Las aplicaciones de este resultado en el caso de la multiaplicación $S(r) = [f \leq r]$ (conjuntos de subnivel de f) se verán en el siguiente capítulo. Es importante que por el momento solo se estudia el comportamiento asintótico de las órbitas, y aún no se ha dado un teorema que garantice su existencia. Tales resultados se presentarán a continuación sin demostración.

Definición 5.9 (Ver [5]) *Una multiaplicación $S : [0, \eta) \rightrightarrows \mathbb{R}^n$ se dirá L-Lipschitz si:*

$$S(r) \subset S(t) + L|t - r|\mathcal{B},$$

en donde $\mathcal{B}$ es la bola unitaria de $\mathbb{R}^n$.

Teorema 5.10 (Ver [5]) *Sean $S : [0, \eta) \rightrightarrows \mathbb{R}^n$ una multiaplicación y un punto $x_0 \in S(0)$. Se considera el siguiente sistema:*

$$\begin{cases} -\dot{x}(r) \in N^c_{S(r)}(x(r)) & c.t.p. \ en \ [0, \eta) \\ x(r) \in S(r) & \forall\, t \in [0, \eta) \\ x(0) = x_0 & \end{cases}$$

Si S es L-Lipschitz, a valores cerrados y no vacíos, entonces para cada $x_0 \in S(0)$ existe una curva L-Lipschitz $x : [0, \eta) \to \mathbb{R}^n$ satisfaciendo el sistema.

Capítulo 6

Sweeping process v/s sistema gradiente

En el presente capítulo se pretende exponer resultados obtenidos durante el trabajo de tesis bajo el contexto de los capítulos anteriores.

En lo que sigue se mostrará cómo del estudio asintótico de las órbitas del *sweeping process* se recuperan los resultados más generales del capítulo 2. Esto es, una desigualdad tipo Łojasiewicz y que las trayectorias del sistema subgradiente tengan longitud finita.

Pero primero hay que precisar el concepto de desingularización:

Definición 6.1 *Sean $f : \mathbb{R}^n \to \mathbb{R}$ una función, $\bar{x} \in \mathbb{R}^n$ tal que $f(\bar{x}) = 0$ y $m_f(\bar{x}) = 0$. Se dice que f es desingularizable en torno al valor crítico 0 si existen constantes $C > 0$, $\varepsilon > 0$ y $\rho > 0$ y una función $\Psi : \mathbb{R}_+ \to \mathbb{R}$ positiva, estrictamente creciente y de clase C^1 tal que:*

$$\Psi'(f(x))m_f^{\uparrow}(x) \geq C,$$

para cada $x \in B(\bar{x}, \varepsilon)$, $f(x) \in (0, \rho)$.
Aquí $m_f^{\uparrow}$ mantiene se define similarmente como en la definición 4.2, pero ahora para el subdiferencial de Clarke-Rockafellar, y es simple de verificar que sigue siendo una función o-minimal.

Observación Si una función es desingularizable en torno a sus valores críticos, entonces las trayectorias absolutamente continuas del sistema dinámico subgradiente:

$$\begin{cases} 0 \in \dot{x}(t) + \partial^{\uparrow} f(x(t)), & c.t.p. \text{ en } (0,T) \\ \partial^{\uparrow} f(x(t)) \neq \emptyset, & \forall\, t \in [0,T) \end{cases}$$

tienen longitud finita y convergen a puntos críticos. Este resultado se deduce fácilmente de lo hecho previamente. Si bien tal demostración se pudo haber hecho tal como en el capítulo 2, definiendo conjuntos adecuados y usando el teorema de selección de curvas, la demostración del capítulo 3 hace uso directo de una desigualdad tipo Łojasiewicz. En este caso se asume al decir que f es desingularizable, por lo tanto se puede quitar la hipótesis de o-minimalidad sobre la función f y reemplazar por la hipótesis de tener valores críticos aisladaos desingularizables.

De igual modo se debe definir lo mismo para multiaplicaciones:

Definición 6.2 *Sean $S : \mathbb{R} \rightrightarrows \mathbb{R}^n$ una multiaplicación a conjuntos definibles y de grafo cerrado y $U \subset \mathbb{R}^n$ un conjunto acotado y no vacío. Se dice que la multiaplicación S es desingularizable en U si se tiene que para cada $a \in \mathbb{R}$ existe $b > a$ y una función estrictamente creciente y continua $\sigma : [a, b) \to \mathbb{R}$ que es además C^1 en (a, b) y $\sigma(a) = a$ tal que se satisface lo siguiente:*

$$|D_c^*(S \circ \sigma)(r|x)|^+ \leq 1$$

para cada $r \in (a, b)$ y cada $x \in S(\sigma(r)) \cap U$.

Proposición 6.3 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ una función convexa, continua y definible en alguna estructura o-minimal. Se tiene entonces que f admite una desingularización, es decir: para cada $a \in \mathbb{R}$ existe $b > a$ y una función estrictamente creciente y continua $\Psi : [a, b) \to \mathbb{R}$ que es además C^1 en (a, b) y satisface $\Psi(a) = a$ tal que:*

$$\operatorname{dist}(0, \partial^\uparrow(\Psi \circ f)(x)) = \Psi'(f(x)) m_f^\uparrow(x) \geq 1.$$

Demostración. Es una conclusión directa de aplicar el teorema de desingularización del capítulo anterior (ver teorema **5.8**). Una función f satisfaciendo tales hipótesis tiene epígrafo cerrado y definible. Como $\operatorname{epi}(f) = \operatorname{gph}(S)$, se deduce que el *sweeping* de f satisface las hipótesis del teorema **5.8**. El otro detalle radica en la observación de que el subdiferencial de f en $x$coincide con la preimagen de $\{-1\}$ por la coderivada de S en $(x, f(x))$:

$$\begin{aligned}
D_c^* S(x|f(x))^{-1}(-1) &= \{y \,:\, (-1, -y) \in N^c_{gph(S)}(x, f(x))\} \\
&= \{y \,:\, (-1, -y) \in \widetilde{N^c_{\operatorname{epi}(f)}}(x, f(x))\} \\
&= \{y \,:\, (-y, -1) \in N^c_{\operatorname{epi}(f)}(x, f(x))\} \\
&= \{y \,:\, -y \in \partial^\uparrow f(x)\} \\
&= -\partial^\uparrow f(x),
\end{aligned}$$

en donde $\widetilde{N_{\operatorname{epi}(f)}}(x, f(x))$ es el conjunto que se obtiene cambiando las coordenadas del cono de Clarke $N^c_{\operatorname{epi}(f)}(x, f(x))$ en el epígrafo de f. Lo anterior entonces es simplemente consecuencia de escribir la definición del cono normal de Clarke.
Un desarrollo casi idéntico muestra que:

$$D_c^*(S \circ \sigma)(x|(\sigma^{-1} \circ f)(x))^{-1}(-1) = -\partial^\uparrow(\Psi \circ f)(x),$$

donde Ψ es la inversa de la función σ y por lo tanto también es C^1, estrictamente creciente. El otro paso relevante es utilizar la siguiente propiedad que fue mostrada en el capítulo previo:

$$|S \circ \sigma(r|x)|^+ = \frac{1}{\inf_{|t|=1} dist(0, (S \circ \sigma)^{-1}(x|r)(t))}.$$

De modo que juntando todo lo anterior y reemplazando con $r = f(x)$ se concluye la siguiente desigualdad:

$$\operatorname{dist}(0, \partial^\uparrow(\Psi \circ f)(x)) \geq 1.$$

Esto es lo mismo que:

$$m_{\Psi \circ f}(x) \geq 1.$$

Lo anterior se puede ver como una desingularización, salvo que hace falta utilizar una regla de cálculo subdiferencial para que lo sea, y más aún, para que se pueda concluir que las trayectorias acotadas del sistema subgradiente tienen longitud finita.
Primero hay que notar que si se tiene una regla de la cadena para la derivada direccionale de Clarke-Rockafellar entonces sigue que hay una regla de la cadena para el subdiferencial de Clarke-Rockafellar. Esto es consecuencia de que si g es una función s.c.i. entonces:

$$\partial^{\uparrow} g(x) = \{v \in \mathbb{R}^n \,:\, \mathrm{d}^{\uparrow} g(x;u) \geq \langle v, u\rangle,\ \forall\, u \in \mathbb{R}^n\}.$$

Por lo tanto para $g = \Psi \circ f$, si efectivamente $\mathrm{d}^{\uparrow}(\Psi \circ f)(x;u) = \Psi'(f(x)) f^{\uparrow}(x;u)$ entonces reemplazando en la fórmula del subdiferencial se deduce:

$$\begin{aligned}
\partial^{\uparrow}(\Psi \circ f)(x) &= \{v \in \mathbb{R}^n \,:\, \Psi'(f(x))\mathrm{d}^{\uparrow} f(x;u) \geq \langle v, u\rangle,\ \forall\, u \in \mathbb{R}^n\} \\
&= \{\Psi'(f(x))w \in \mathbb{R}^n \,:\, \Psi'(f(x))\mathrm{d}^{\uparrow} f(x;u) \geq \langle \Psi'(f(x))w, u\rangle,\ \forall\, u \in \mathbb{R}^n\} \\
&= \{\Psi'(f(x))w \in \mathbb{R}^n \,:\, \mathrm{d}^{\uparrow} f(x;u) \geq \langle w, u\rangle,\ \forall\, u \in \mathbb{R}^n\} \\
&= \Psi'(f(x))\partial^{\uparrow}(f)(x).
\end{aligned}$$

Por ende hay que corroborar que en este caso se tiene regla de la cadena para la derivada direccional de Clarke, siempre que Ψ sea un difeomorfismo creciente:

$$\begin{aligned}
\mathrm{d}^{\uparrow}(\Psi \circ f)(x;u) &= \sup_{\varepsilon>0} \limsup_{\substack{x'\to x \\ t\searrow 0}} \inf_{\nu\in B(u,\varepsilon)} \frac{\Psi(f(x'+t\nu)) - \Psi(f(x'))}{t} \\
&= \sup_{\varepsilon>0} \limsup_{\substack{x'\to x \\ t\searrow 0}} \inf_{\nu\in B(u,\varepsilon)} \Big(\frac{\Psi(f(x'+t\nu)) - \Psi(f(x'))}{f(x'+t\nu) - f(x')}\Big)\Big(\frac{f(x'+t\nu) - f(x')}{t}\Big)
\end{aligned}$$

Observar aquí que la expresión:

$$\Big(\frac{\Psi(f(x'+t\nu)) - \Psi(f(x'))}{f(x'+t\nu) - f(x')}\Big)\Big(\frac{f(x'+t\nu) - f(x')}{t}\Big),$$

con $t > 0$ y $\varepsilon > 0$ fijos es continua en función de ν, entonces el ínfimo se alcanza en la bola, en un vector $\nu(\varepsilon)$, el cual es en particular ínfimo de

$$\frac{f(x'+t\nu) - f(x')}{t},$$

puesto que la función Ψ es estrictamente creciente, por lo tanto:

$$\begin{aligned}
&\inf_{\nu\in B(u,\varepsilon)} \Big(\frac{\Psi(f(x'+t\nu)) - \Psi(f(x'))}{f(x'+t\nu) - f(x')}\Big)\Big(\frac{f(x'+t\nu) - f(x')}{t}\Big) \\
&= \inf_{\nu\in B(u,\varepsilon)} \frac{\Psi(f(x'+t\nu)) - \Psi(f(x'))}{f(x'+t\nu) - f(x')} \inf_{\nu\in B(u,\varepsilon)} \frac{f(x'+t\nu) - f(x')}{t}
\end{aligned}$$

Por otro lado el límite superior de:

$$\Big(\frac{\Psi(f(x'+t\nu))-\Psi(f(x'))}{f(x'+t\nu)-f(x')}\Big)\Big(\frac{f(x'+t\nu)-f(x')}{t}\Big)$$

se puede aproximar por una sucesión de pares (x'_n, t_n). Además como el límite de la primera fracción se alcanza, se tiene que el límite superior del producto acota superiormente al producto de los límites superiores. De este modo:

$$\begin{aligned}&\limsup_{\substack{x'\to x\\ t\searrow 0}} \inf_{\nu\in B(u,\varepsilon)} \Big(\frac{\Psi(f(x'+t\nu))-\Psi(f(x'))}{f(x'+t\nu)-f(x')}\Big)\Big(\frac{f(x'+t\nu)-f(x')}{t}\Big)\\ &=\limsup_{\substack{x'\to x\\ t\searrow 0}} \inf_{\nu\in B(u,\varepsilon)} \frac{\Psi(f(x'+t\nu))-\Psi(f(x'))}{f(x'+t\nu)-f(x')} \limsup_{\substack{x'\to x\\ t\searrow 0}} \inf_{\nu\in B(u,\varepsilon)} \frac{f(x'+t\nu)-f(x')}{t}.\end{aligned}$$

Más aún, por el teorema del valor medio: $\Psi(f(x'+t\nu(\varepsilon)))-\Psi(f(x'))=\Psi'(r)$, con $r\in(f(x'), f(x'+t\nu(\varepsilon)))$ (o al revés dependiendo de cual sea el mayor), por continuidad de f sobre el segmento, se tiene que existe $s\in(0,t)$ tal que $r=f(x'+s\nu(\varepsilon))$.
Por esto último, y gracias a la continuidad, se puede tomar el límite superior en la expresión:

$$\frac{\Psi(f(x'+t\nu(\varepsilon)))-\Psi(f(x'))}{f(x'+t\nu(\varepsilon))-f(x')}$$

este ya no depende de $\varepsilon>0$, puesto que $t\to 0$. Esto, junto con el hecho de que el $\nu(\varepsilon)$ es el mismo que minimiza a cada uno por separado resulta en lo siguiente:

$$\begin{aligned}&\limsup_{\substack{x'\to x\\ t\searrow 0}} \inf_{\nu\in B(u,\varepsilon)} \Big(\frac{\Psi(f(x'+t\nu))-\Psi(f(x'))}{f(x'+t\nu)-f(x')}\Big)\Big(\frac{f(x'+t\nu)-f(x')}{t}\Big)\\ &=\limsup_{\substack{x'\to x\\ t\searrow 0}} \Psi'(f(x'+s\nu(\varepsilon)))\limsup_{\substack{x'\to x\\ t\searrow 0}} \inf_{\nu\in B(u,\varepsilon)} \frac{f(x'+t\nu)-f(x')}{t}\\ &=\Psi'(f(x))limsup_{\substack{x'\to x\\ t\searrow 0}} \inf_{\nu\in B(u,\varepsilon)} \frac{f(x'+t\nu)-f(x')}{t}.\end{aligned}$$

Finalmente, como consecuencia de tomar supremo sobre $\varepsilon>0$,

$$\mathrm{d}^{\uparrow}(\Psi\circ f)(x;u)=\Psi'(f(x))\mathrm{d}^{\uparrow}f(x;u)$$

Por otro lado

$$\begin{aligned}m_{\Psi\circ f}(x)&=\{\|v\|\,:\,v\in\partial^{\uparrow}(\Psi\circ f)(x)\}=\{\|v\|\,:\,v\in\Psi'(f(x))\partial^{\uparrow}f(x)\}\\ &=\Psi'(f(x))\{\|v\|\,:\,v\in\partial^{\uparrow}f(x)\}=\Psi'(f(x))m_f^{\uparrow}(x)\end{aligned}$$

De lo que sigue la conclusión final:

$$\Psi'(f(x))m_f^{\uparrow}(x) \geq 1$$

$\square$

El sencillo resultado anterior motiva preguntas naturales que vinculan la desingularización de una función con su proceso de *sweeping* y más allá lo que es posible afirmar del acotamiento de la longitud de las curvas del sistema dinámico definido por la inclusión diferencial del *sweeping*.
Entonce la pregunta interesante, que deja de lado las herramientas de geometría o-minimal es la siguiente:

f es desingularizable $\iff$ el *sweeping process* correspondiente a f es desingularizable.

Proposición 6.4 *El enunciado anterior tiene respuesta positiva, en el caso en que f es continua.*

La demostración se obtiene por inspección del teorema anterior, al mostrar como pasar de coderivada a subdiferencial y al revés. Por el teorema de la función inversa, σ y Ψ siempre serán C^1 y además estrictamente crecientes. Más aún, tales funciones existen si se supone la existencia de una desingularización de f en un punto crítico o una desingularización de su *sweeping process*. Por lo tanto los pasos realizados durante la demostración son reversibles y permiten concluir una desingularización a partir de la otra, pasando de la coderivada $D_c^*(S \circ \sigma)(x|f(x))$ en puntos de la forma $(x, f(x))$ al subdiferencial $-\partial^{\uparrow} f(x)$ en el punto x, cabe observar que tales puntos son los únicos relevantes, si se cambia $f(x)$ por $r \neq f(x)$ entonces el par $(x, f(x))$ pertenece al interior del epígrafo de f o simplemente no pertenece a este, por lo tanto el cono normal de Clarke $N^c_{\mathrm{epi}(f)}(x|f(x))$ es vacío.

Conclusión

En este trabajo se revisaron aspectos de la geometría o-minimal y del análisis variacional, que proporcionaron las herramientas y el marco teórico para obtener resultados sobre el comportamiento asintótico de sistemas de descenso (gradiente, subgradiente y *sweeping process*). Las demostraciones presentadas en esta memoria fueron desarrolladas en el marco más general posible, extendiendo ligeramente lo realizado en otras publicaciones, como por ejemplo [1] (marco subanalítico global) o [6] (caso diferenciable).

A lo largo de lo expuesto y los resultados de esta memoria se aprecia la potencia de la geometría o-minimal y su sinergia con el análisis variacional para el estudio del comportamiento asintótico de las órbitas de los sistemas vistos, en torno a los valores críticos. Una herramienta indispensable, y característica de la geometría o-minimal es el llamado lema de selección de curvas. Este se usa reiteradamente sobre conjuntos inteligentemente construidos para obtener desigualdades cruciales, tal como se pudo apreciar en demostraciones de los capítulos 3 y 4. En general el esquema para obtener los resultados fue a grandes rasgos el mismo: a partir de un resultado para los valores críticos tipo Sard, se establece una versión de la desigualdad de Kurdyka-Łojasiewicz para luego aplicar dicho resultado a las órbitas del sistema en estudio.

Finalmente se mostró a la desingularización como una propiedad esencial para acotar la longitud de las órbitas de los sistemas abordados en el trabajo. Más aún se concluyó que es equivalente tener desingularización para una función continua como tenerla para su *sweeping process* correspondiente.

Bibliografía

[1] Jerome Bolte, Aris Daniilidis, and Adrian Lewis. The łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.* **17** (2006), no. 4, 1205–1223.

[2] Jerome Bolte, Aris Daniilidis, Adrian Lewis, and Masahiro Shiota. Clarke subgradients of stratifiable functions. *SIAM J. Optim.* **18** (2007), 556–572.

[3] Michel Coste. *An introduction to o-minimal geometry.* RAAG Network, 1999.

[4] Michel Coste. *An introduction to semialgebraic geometry.* RAAG Network, 2002.

[5] Aris Daniilidis and Dmitriy Drusvyatskiy. Sweeping by a tame process. **preprint** 20p, 2015.

[6] Krzysztof Kurdyka. On gradients of functions definible in o-minimal structures. *Ann. Inst. Fourier (Grenoble)* **48** (1998), no. 3, 769-783.

[7] Jean Jaques Moreau. Evolution problem associated with a moving convex set in a hilbert space. *J. Differential Equations* **26** (1977), no. 3, 347–374.

[8] Jacob Palis and Welington de Melo. *Geometric Theory of Dynamical Systems: An Introduction.* Springer, 1982.

[9] Tyrrell Rockafellar and Roger Wets. *Variational Analysis.* Springer, 2009.

[10] Lou van den Dries and Chris Miller. Geometric categories and o-minimal structures. *Duke Math. J.* **84** (1996), no. 2, 497–540.