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

SEUDO-MÉTRICAS INDUCIDAS POR FUNCIONES DE TIPO LEGENDRE Y MÉTODOS DINÁMICOS EN OPTIMIZACIÓN.

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

CRISTOPHER ADRIÁN HERMOSILLA JIMÉNEZ

PROFESOR GUÍA :
FELIPE ÁLVAREZ DAZIANO

MIEMBROS DE LA COMISIÓN:
ROBERTO COMINETTI COTTI-COMETTI
HÉCTOR RAMÍREZ CABRERA

SANTIAGO DE CHILE
AGOSTO 2011

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL MATEMÁTICO
POR : CRISTOPHER HERMOSILLA J.
FECHA : 1 DE JUNIO DE 2011
PROF. GUIA: Sr. FELIPE ÁLVAREZ D.

SEUDO-MÉTRICAS INDUCIDAS POR FUNCIONES DE TIPO LEGENDRE Y MÉTODOS DINÁMICOS EN OPTIMIZACIÓN.

El objetivo de la presente memoria es proponer un nuevo método para resolver una clase general de problemas de optimización, a saber, dado un conjunto convexo y abierto $C \subseteq \mathbb{R}^n$, una función diferenciable $f : \mathbb{R}^n \to \mathbb{R}$, una matriz de rango completo $A \in \mathcal{M}_{m\times n}$ (con $m \leq n$) y un vector $b \in \operatorname{Im} A$, buscamos resolver algorítmicamente el problema

$$(P_0) \qquad \min\{f(x) : \ x \in \operatorname{cl} C,\ Ax = b\}.$$

Para esto, tomamos herramientas de la Geometría Riemanniana, las mezclamos con el método de máximo descenso y nos preguntamos qué sucede si miramos este algoritmo bajo la lupa de otra métrica, una no necesariamente Euclideana. Si bien la idea de usar métricas variables para resolver este tipo de problemas no es nueva, nuestro trabajo sí lo es, pues nos interesamos en una en particular, una que es inducida por el cuadrado de la matriz Hessiana de una cierta función barrera cuyo dominio coincide con C. Esta métrica tiene la gran gracia de proveernos de una isometría, fácil de calcular, entre el conjunto C, visto como variedad, y un espacio Euclideano apropiado.

En el capítulo 1 de esta memoria damos una descripción introductoria de las herramientas de la Geometría Riemanniana que usamos para desarrollar nuestra teoría. En el capítulo 2 definimos formalmente la *métrica Hessiana cuadrada de Legendre* sobre un dominio convexo. Estudiamos también sus principales propiedades y consecuencias. En el capítulo 3 introducimos un nuevo método de optimización para resolver de forma algorítmica un problema más simple que (P_0), el de minimizar la función f sólo sobre la adherencia del conjunto C. También introducimos una nueva noción de dualidad y presentamos algunos teoremas de convergencia. En el capítulo 4 generalizamos este método, con el fin de resolver algorítmicamente el problema (P_0). Por otra parte, en el capítulo 5 abordamos la pregunta de en qué casos nuestra métrica coincide con la inducida por la Hessiana de otra función barrera. Primeramente, planteamos el problema para el caso separable, obteniendo condiciones necesarias y suficientes, para luego pasar a un caso más general, donde sólo obtuvimos una condición necesaria. Finalmente, usando este criterio mostramos que el problema es en realidad muy restrictivo respecto al conjunto C, lo cual nos hace conjeturar que esta pregunta no es fácil de responder y que la respuesta es en general negativa.

Cabe destacar que la noción de dualidad que aquí introducimos crea un lazo entre las propiedades de carácter Riemanniano y las de carácter Euclideano, en particular, permite transformar problemas no convexos en otros que sí lo son. Más aún, esta noción nos muestra que es posible resolver ciertos problemas de optimización con restricciones aplicando métodos de optimización irrestricta sobre un problema dual adecuado.

AGRADECIMIENTOS

Quisiera comenzar por agradecer a todos quienes he tenido el honor de tener como profesores y que siempre me motivaron a dar lo mejor de mí. Gracias por su invaluable orientación y por la formación que me permitieron obtener. En especial quisiera agradecer a mis profesores del Instituto Nacional, quienes me introdujeron en el maravilloso mundo del pensamiento matemático.

Mi más sincero agradecimiento a quien fuera mi profesor guía en esta memoria, Felipe Álvarez, por sus buenos consejos, por su dedicación y porque me permitió conocer y encantarme con el mundo de la investigación científica. Agradezco también a los profesores que tuvieron la gentileza de ser parte de la comisión que evaluará esta memoria: Roberto Cominetti y Héctor Ramírez.

A todos los miembros del Departamento de Ingeniería Matemática de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Quienes cumplen una silenciosa, pero fundamental labor dentro del Departamento.

Al Centro de Modelamiento Matemático, que no tan sólo financió esta memoria, si no que me permitió dar mis primeros pasos como matemático a través de los distintos trabajos de investigaciones en que pude participar.

A mis padres, porque ellos me dieron la motivación inicial para convertirme en el hombre que hoy soy. A Sandra y Marcela porque sin su paciencia y apoyo, tal vez no podría estar escribiendo estas palabras.

A mis amigos y a mi hermana, quienes siempre se han preocupado por mí.

Finalmente quisiera agradecer a Aris Daniilidis por su hospitalidad, por darse el tiempo de entender mi trabajo y por compartir conmigo sus impresiones acerca de esta memoria.

En memoria de mi abuelo.

Índice General

Introducción

Uno de los problemas más recurrentes en la ingeniería y economía moderna es el de encontrar el valor óptimo de una cierta función objetivo $f : \mathbb{R}^n \to \mathbb{R}$, sobre un conjunto convexo intersectado con un subespacio afín de $\mathbb{R}^n$, es decir, problemas que se pueden escribir de la forma:

$$(P_0) \qquad \min\{f(x) : \; x \in \operatorname{cl} C, \; Ax = b\}.$$

Existen diversos métodos que permiten resolver algorítmicamente esta clase de problemas, pero todos ellos tienen que lidiar con alguna dificultad, ya sea de factibilidad o porque en cada iteración tienen que resolver otro problema de optimización; ver por ejemplo los métodos que aparecen en [10] y [14]. Esto es inconveniente pues puede llegar a significar un gran costo computacional a la hora de calcular cada iteración. Cabe destacar que este problema no es parte de algoritmos clásicos de la optimización sin restricciones, por lo que sería ideal contar con un método que rescate esto.

Nuestro principal objetivo es proponer una nueva manera de resolver el problema (P_0), recuperando la simplicidad de los algoritmos de la optimización irrestricta. Para esto, exploramos métodos de métrica variable, con el fin de cambiar la geometría del espacio y ocupar esto a nuestro favor. Proponemos estudiar el método de máximo descenso bajo la luz de una métrica Riemanniana especial, una que es inducida por el cuadrado de la matriz Hessiana de una función barrera particular. La clase de funciones barrera que ocuparemos son las conocidas por el nombre de *funciones de tipo Legendre*. Estas funciones fueron introducidas por Rockafellar en [19, Sección 26] y han sido utilizadas por otros autores con fines similares. Tal es el caso de Álvarez, Bolte y Brahic que en [1] estudiaron los flujos de gradiente Riemanniano con respecto a la métrica inducida por la matriz Hessiana de una de estas funciones, una especie de versión continua del método de máximo descenso.

En el capítulo 1 describimos las herramientas de la Geometría Riemanniana que usamos a través de toda esta memoria. Este primer capítulo tiene como objetivo facilitar el entendimiento de los subsiguientes y a pesar de que esta teoría es bastante conocida a nivel mundial, este capítulo sirve para fijar ideas, pues se definen conceptos que no son tan conocidos, como por ejemplo el de convexidad sobre una variedad Riemanniana. También se dan las definiciones formales de los conceptos más utilizados en nuestro estudio, como el de variedad Riemanniana, geodésica o isometría.

El capítulo 2 comienza recordando nociones básicas del Análisis Convexo, para luego dar a conocer las principales suposiciones teóricas en que se basa este estudio. En particular, en este capítulo damos una descripción detallada de las funciones de tipo Legendre y explicamos cómo con éstas podemos asociar a un conjunto abierto y convexo la métrica Riemanniana especial de qué hemos hablado, la cual llamamos en adelante *métrica Hessiana cuadrada de Legendre*. Luego, explicamos las principales consecuencias que se obtienen al dotar a este conjunto, que en adelante notaremos por C, de esta métrica cuando la función de Legendre que la induce, que en adelante notaremos por g, tiene la propiedad que el interior de su dominio coincide precisamente con el conjunto C. El ejemplo más sencillo que podemos considerar es la función $g(x) = \frac{1}{2}|x|^2$ y $C = \mathbb{R}^n$. Un hecho notable es que en este caso, la métrica Riemanniana coincide con la Euclideana.

Cabe destacar que es variada la literatura que existe con respecto a las métricas Riemannianas asociadas a una función barrera como, por ejemplo, el caso de funciones auto-concordantes cuyo uso está asociado a algoritmos de punto interior; para más detalles ver por ejemplo [13] o [18]. Sin embargo, la ventaja que presenta nuestra métrica es que nos provee de una isometría entre el conjunto C y un espacio Euclideano, que en muchos casos puede ser considerado como todo el espacio $\mathbb{R}^n$, lo cual es muy provechoso a la hora de mirar bajo otra óptica un problema no lineal que sea de la forma:

$$(P) \qquad \min\{f(x) : \ x \in C\}.$$

Más aún, la isometría anteriormente mencionada es fácil de calcular y depende exclusivamente de la función g, pues corresponde al gradiente de esta función. Este hecho en sí, es el pilar fundamental de todo lo que se hace en los capítulos venideros, pues permite calcular en forma explícita las geodésicas de nuestra variedad Riemanniana. Esto último es una gran ventaja si queremos comparar nuestro análisis con lo hecho por otros autores que también estudian algoritmos de búsqueda lineal sobre geodésicas; ver por ejemplo lo expuesto en [17] o [20].

El capítulo 3 está dedicado a la elaboración del método que buscamos en un caso más simple, cuando el subespacio afín es todo $\mathbb{R}^n$, esto es, que sirva para resolver el problema (P). Para esto, primero notamos que el algoritmo de máximo descenso clásico es en sí un método que sigue geodésicas Riemannianas, en efecto, recordemos que este método consiste en, dado un punto de iteración x_k y una dirección de descenso d_k, buscar sobre el rayo $\{x_k + td_k, \ t \geq 0\}$ un punto que haga que la función descienda lo suficiente. Este rayo no es otra cosa que una geodésica en $\mathbb{R}^n$. Luego, basándonos en esta idea, y ya que podemos calcular las geodésicas de C visto como variedad Riemanniana, es que construimos nuestros algoritmo haciendo que éste busque, a lo largo de las geodésicas de nuestra variedad, un punto que haga que la función objetivo descienda lo suficiente. Cabe destacar que esto inmediatamente hace surgir dos cuestiones, cómo escoger la dirección de descenso, que en este caso corresponde a la velocidad de la geodésica en el punto de iteración, y cómo escoger el paso del algoritmo.

Una de las propiedades importantes de la métrica Hessiana cuadrada de Legendre que fue expuesta primeramente por Álvarez et al en [1], es que las geodésicas de C, dotado de ésta métrica, minimizan asintóticamente una función lineal sobre el conjunto C. Esto nos dice que la dirección de descenso debe rescatar al menos información de primer orden de la función objetivo. Es así como se puede pensar en direcciones de descenso tales como $d_k = -\nabla f(x_k)$ y $d_k = -\nabla_G f(x_k)$, donde esta última corresponde al gradiente Riemannaniano de f con respecto a la métrica Hessiana cuadrada de Legendre.

Por otra parte, con la isometría en mente proponemos un problema dual a (P) que consiste en minimizar otra función, que llamamos φ y que está definida por la ecuación $f = \varphi \circ \nabla g$, donde g es la función de Legendre que induce la métrica Hessiana cuandrada. El conjunto sobre el que se minimiza esta función es denotado por C^* y corresponde al interior del dominio de g^*, la conjugada de Fenchel de g. Esto implica que el problema dual se puede escribir como:

$$(D) \qquad \inf\{\varphi(y) : \ y \in C^*\}$$

Con esta definición de φ es que se obtiene una relación entre propiedades Riemannianas y Euclideanas, por ejemplo, f es una función convexa en el sentido Riemanniano si y sólo si φ es convexa en el sentido clásico. Más aún, se tiene que el gradiente Riemanniano de f coincide con el gradiente Euclideano de φ. Es por esto último que en esta memoria se optó por desarrollar el algoritmo con el gradiente Riemanniano de la función objetivo como dirección de descenso, pues en este caso el método puede ser visto como un método de máximo descenso en el espacio dual C^*, lo cual facilita enormemente el análisis de convergencia, pues se rescata todo lo que se sabe del método del gradiente; ver por ejemplo los resultados obtenidos en [5] y [10]. En particular, se obtiene un teorema de convergencia en el caso que f es convexa en el sentido Riemanniano cuando el paso es acotado. Más aún esta noción de dualidad resuelve inmediatamente el problema de cómo escoger el paso del algoritmo, pues permite adaptar a nuestra situación todos los métodos de búsqueda lineal conocidos y estudiados para los métodos clásicos; ver por ejemplo los métodos definidos en [14, Capítulo 3].

Ahora bien, con la finalidad de abarcar problemas un poco más generales que (P), en el capítulo 4 generalizamos los resultados anteriormente obtenidos a una subvariedad de C, que corresponde a la restricción a un subespacio afín propio de $\mathbb{R}^n$, esto es para poder abarcar problemas del estilo de (P_0), donde A es una matriz de rango completo de dimensiones $m \times n$ (con $m \leq n$) y b un vector en el rango de A. En particular, con este análisis es posible resolver problemas sobre el simplex n-dimensional, pues basta notar que al tomar $C = \{x > 0\}$, $m = 1$, $A = [1 \ \dots 1]$ y $b = 1$ se obtiene el interior relativo de este conjunto, y esto, a su vez, permite abarcar distintos problemas de optimización que aparecen ingeniería moderna, como el de encontrar distribuciones óptimas de probabilidades .

En el capítulo 5 se cambia el enfoque del estudio, pues se plantea una pregunta que tiene un carácter más teórico que práctico, a saber, queremos responder la pregunta de en qué casos la métrica Hessiana cuadrada de Legendre coincide con la métrica Riemanniana utilizada por Álvarez et al en [1]. Esta métrica depende directamente de la matriz Hessiana de una función de Legendre. Por lo tanto, el problema se puede plantear como sigue: dada una función de tipo Legendre h, encontrar otra función de Legendre g, tal que $\nabla^2 h = (\nabla^2 g)^2$. En tal caso a la función g le llamamos *raíz cuadrada de Legendre*. Si bien, esta pregunta es interesante en sí, también tiene un motivación teórica, pues es natural preguntarse si nuestra teoría coincide con la expuesta en [1], es decir, es sólo otra expresión de lo ya conocido, o bien es un camino totalmente nuevo.

La estrategia para responder esta pregunta fue comenzar por estudiar el caso cuando la función g se puede escribir en forma separable, para luego ver que se puede decir en un caso más general. En el primer caso, encontramos una condición necesaria y suficiente sobre la función h que nos permitió responder satisfactoriamente esta pregunta. Cabe destacar que esto también nos sirvió para obtener una fórmula sencilla para la distancia geodésica. Además, dada la simplicidad del criterio que encontramos, pudimos estudiar en profundidad diversos ejemplos prácticos.

Por otra parte, en el caso general sólo pudimos obtener una condición necesaria sobre la función h. Sin embargo, esta condición nos permitió descartar casos de funciones que a priori uno pensaría que sí existe una tal función g. Es el caso cuando el conjunto C es un poliedro. En [1] se probó que, para un poliedro definido por una matriz de rango completo, ciertas funciones, que son separables con respecto a las desigualdades que definen al poliedro, son de Legendre. Nosotros en este capítulo mostramos que en el caso 2-dimensional, esta clase de funciones tiene una raíz de Legendre si y sólo si los lados del poliedro son paralelos o perpendiculares entre sí.

Finalmente, queremos destacar que algorítmicamente, en esta memoria, por simplicidad y elegancia, sólo se estudió el caso del método de máximo descenso, pero nada impide estudiar otros métodos de este estilo usando la misma metodología. Por otra parte, el nexo que crea la noción de dualidad entre las propiedades Riemannianas y las Euclideanas, puede ser explotada aún más. Por ejemplo, tratando problemas con estructuras más generales, como puede ser el caso cuando la función objetivo es quasi-convexa en el sentido Riemanniano; el capítulo 5 sirve para reforzar esta última idea. Estos argumentos nos hacen pensar que la teoría que proponemos es nueva y diferente de la expuesta por Álvarez et al, pues al parecer los resultados estas teorías sólo coinciden en casos muy particulares. Creemos que nuestra teoría aún tiene mucho potencial y que puede ser explorada en distintas direcciones, ya sea desde el punto de vista algorítmico o bien, desde la noción de dualidad, en las formas que ya mencionamos.

Capítulo 1

Basic notions of Riemannian Geometry

The Riemannian Geometry is a branch of the Differential Geometry that study manifolds with an inner product on the tangent space at each point which varies smoothly from point to point. In particular, this gives local notions of angle, length of curves, surface area, and volume. It is a very broad and abstract generalization of the Differential Geometry of surfaces in $\mathbb{R}^3$; see for example [6]. The reader who is familiar with the standard notions of this theory may skip this chapter because the purpose of this is to give a brief introduction of the basic concepts used in this thesis in order to make easier the reading. The following definitions were taken mainly from [7, 9, 20].

Let us start recalling with the most basic definition. A differentiable manifold of dimension n is a set M such that there exists a family of injective mappings $x_\alpha : U_\alpha \to M$ of open sets $U_\alpha \subseteq \mathbb{R}^n$ into M, that satisfies:

- $\bigcup_\alpha x_\alpha(U_\alpha) = M$.
- For any pair α, β with $x_\alpha(U_\alpha) \cap x_\beta(U_\beta) = W \neq \emptyset$, the sets $x_\alpha^{-1}(W)$ and $x_\beta^{-1}(W)$ are open sets in $\mathbb{R}^n$ and the mapping $x_\beta^{-1} \circ x_\alpha$ is differentiable.

Hereinafter, let M be an n-dimensional differentiable manifold. We say that a vector v is *tangent* to M at $x \in M$ if there exists a curve c drawn on M such that $c(0) = x$ and $c'(0) = v$. The tangent space to M at $x \in M$ is the set of all these vector and is denoted by T_xM. Let TM be the disjoint union of the tangent spaces to M at all points of M, this object is called the *tangent bundle* of M. Let us recall that a vector field on M is a smooth map X from M to TM such that, for any $x \in M$, $X(x) \in T_xM$. We denote by $\Xi(M)$ the vector space of vector fields on M.

On the other hand, as in the classical Differential Calculus, given a $\mathcal{C}^1$ function $f : M \mapsto \mathbb{R}$ we denote by d_xf its differential or tangent map $d_xf : T_xM \to \mathbb{R}$ at $x \in M$. We denote by $C^1(M)$ the set of all these function. Let us recall that, using the idea of parametrization, $X \in \Xi(M)$ can be also seen as a map from $C^1(M)$ into the set of functions on M.

A $\mathcal{C}^k$ metric on M, $k \geq 0$, is a family of scalar products $(\cdot,\cdot)_x$ on each T_xM, $x \in M$, such that $(\cdot,\cdot)_x$ depends in a $\mathcal{C}^k$ way on x. The couple $(M,(\cdot,\cdot))$ is called a $\mathcal{C}^k$ Riemannian manifold. This structure permits to identify T_xM with its dual space, that is, the cotangent space T_xM^*, and thus to define a notion of gradient vector. Indeed, given f in $\mathcal{C}^1(M)$, the gradient of f is denoted by $\nabla_{(\cdot,\cdot)} f$ and is uniquely determined by the following two conditions:

- **Tangency:** for all $x \in M$, $\nabla_{(\cdot,\cdot)} f(x) \in T_xM$,
- **Duality:** for all $x \in M$, $v \in T_xM$, $d_xf(v) = (\nabla_{(\cdot,\cdot)} f(x), v)_x$.

An *affine connection* on TM is a bilinear map $\nabla : \Xi(M) \times \Xi(M) \to \Xi(M)$ denoted by $(X,Y) \mapsto \nabla_X Y$, such that, for $X,Y \in \Xi(M)$ and $f \in \mathcal{C}^\infty(M)$, we have

$$\nabla_{fX} Y = f\nabla_X Y \quad \text{and} \quad \nabla_X(fY) = X(f)Y + f\nabla_X Y.$$

A vector field X along a curve $c : [a,b] \to M$ is a differentiable mapping that associates with every $t \in [a,b]$ a tangent vector $X(t) \in T_{c(t)}M$. In a differentiable manifold M with affine connection ∇, there exists a unique correspondence which associates to a vector field X along a differentiable curve $c : [a,b] \to M$ another vector field $\frac{DX}{dt}$ along c, called the covariant derivative of X along c such that:

$$\frac{D}{dt}(X+Y) = \frac{DX}{dt} + \frac{DY}{dt} \quad \text{and} \quad \frac{D}{dt}(fX) = \frac{df}{dt}X + f\frac{DX}{dt},$$

where Y is a vector field along c and f is a differentiable function on $[a,b]$. Moreover, if $Y \in \Xi(M)$ and $X(t) = Y(c(t))$ then $\frac{DX}{dt} = \nabla_{\frac{dc}{dt}} Y$.

In a differentiable manifold M with affine connection ∇, a parametrized curve $\gamma : [a,b] \to M$ is called *geodesic* if

$$\frac{D}{dt}\left(\frac{d\gamma}{dt}\right) = 0.$$

The manifold $(M,(\cdot,\cdot))$ is called *geodesically complete* if any geodesic $\gamma(t)$ starting from $x \in M$ is defined for all values of the parameter $t \in \mathbb{R}$.

Let $c : [0,1] \to M$ be a piecewise differentiable curve and let $0 = a_0 < a_1 < ... < a_n = 1$ be a partition such that $c|_{(a_i,a_{i+1})}$ is of class $\mathcal{C}^1$, then the length of c is given by

$$\ell(c) = \sum_{i=0}^{n-1} \int_{a_i}^{a_{i-1}} \sqrt{\left(\frac{dc}{dt}, \frac{dc}{dt}\right)} dt$$

We can associate to a Riemannian manifold $(M,(\cdot,\cdot))$ a distance d given by

$$dist(x,y) = \inf\{\ell(\gamma) \mid \gamma \text{ piecewise differentiable curve joining } x \text{ and } y\}.$$

Every geodesic in $(M,(\cdot,\cdot))$ has minimizing properties, that is, let $\gamma : [0,1] \to M$ be a geodesic, then for any piecewise differentiable curve $c : [0,1] \to M$ joining $\gamma(0)$ and $\gamma(1)$, we have $\ell(\gamma) \leq \ell(c)$ and if equality holds then $\gamma([0,1]) = c([0,1])$.

With the distance $dist$, we can endow M with the structure of a metric space, moreover, $(M,(\cdot,\cdot))$ is called complete if the associated metric space $(M, dist)$ is complete. The theorem of Hopf-Rinow - see [9, Corollary 2.105]- asserts that the following are equivalent:

1. $(M,(\cdot,\cdot))$ is a complete Riemannian manifold.
2. $(M,(\cdot,\cdot))$ is geodesically complete.

Let M and N be two Riemannian manifolds, a diffeomorphism $F : M \to N$ (that is, f is a differentiable bijection with a differentiable inverse) is called an *isometry* if

$$(u,v)_x = (d_xF(u), d_xF(v))_{f(x)}, \quad \forall x \in M,\ \forall u,v \in T_xM. \tag{1.1}$$

An isometry $F : M \to N$ is distance preserving, that is,

$$dist_M(x,y) = dist_N(F(x),F(y)), \quad \forall x,y \in M.$$

Finally, let $(M,(\cdot,\cdot))$ be a complete Riemannian manifold, a subset A of M is called *totally convex* if A contains every geodesic γ_{xy} of M whose endpoints x and y belong to A. A function $f : A \to \mathbb{R}$ is called convex on a totally convex set A if

$$f(\gamma_{xy}(t)) \leq (1-t)f(x) + tf(y), \quad \forall t \in [0,1],\ \forall x,y \in A,$$

for every geodesic $\gamma_{xy} : [0,1] \to A$ with $\gamma_{xy}(0) = x$ and $\gamma_{xy}(1) = y$.

Capítulo 2

Squared Hessian metrics induced by Legendre functions

2.1 Preliminaries

Throughout this paper we adopt the standard notation of Convex Analysis theory; see [19]. We say that an extended-real-valued function $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ belongs to the class $\Gamma_0(\mathbb{R}^n)$ when g is lower semicontinuous, proper (i.e. $g \not\equiv +\infty$) and convex. For a function $g \in \Gamma_0(\mathbb{R}^n)$, its *effective domain* is defined by $\operatorname{dom} g = \{x \in \mathbb{R}^n \mid g(x) < +\infty\}$. When $g \in \Gamma_0(\mathbb{R}^n)$, its *Fenchel-Legendre conjugate* is given by

$$g^*(y) = \sup\{\langle x, y\rangle - g(x) \mid x \in \operatorname{dom} g\},$$

and its *subdifferential* is the set-valued mapping $\partial g : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ given by

$$\partial g(x) = \{y \in \mathbb{R}^n \mid \forall z \in \mathbb{R}^n,\ g(x) + \langle y, z - x\rangle \leq g(z)\},$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product in $\mathbb{R}^n$. We set $\operatorname{dom} \partial g = \{x \in \mathbb{R}^n \mid \partial g(x) \neq \emptyset\}$.

We denote by $\operatorname{int} A$, $\operatorname{cl} A$ and ∂A the interior, the closure and the boundary of a set $A \subset \mathbb{R}^n$, respectively.

According to [19, Section 26], a function $g \in \Gamma_0(\mathbb{R}^n)$ is called *essentially smooth* if

- $\operatorname{int} \operatorname{dom} g \neq \emptyset$.

- g is differentiable on $\operatorname{int}\operatorname{dom} g$, that is, $\partial g(x) = \{\nabla g(x)\}$ for every $x \in \operatorname{int}\operatorname{dom} g$.
- $|\nabla g(x_k)| \to +\infty$ for every sequence $\{x_k\} \subseteq \operatorname{int}\operatorname{dom} g$ such that $x_k \to \bar{x}$ for some $\bar{x} \in \partial\operatorname{dom} g$.

Here, $|\cdot|$ stands for the Euclidean norm in $\mathbb{R}^n$ and $\nabla g(x)$ for the Euclidean gradient of g at x.

Notice that by [19, Theorem 26.1], $g \in \Gamma_0(\mathbb{R}^n)$ is essentially smooth if and only if $\partial g(x) = \{\nabla g(x)\}$ if $x \in \operatorname{int}\operatorname{dom} g$ and $\partial g(x) = \emptyset$ otherwise; in particular, $\operatorname{dom}\partial g = \operatorname{int}\operatorname{dom} g$. Besides, g is called *essentially strictly convex* if it is strictly convex on every convex subset of $\operatorname{dom}\partial g$. An important characterization is the following (see [19, Section 26]): g is essentially smooth if and only if its conjugate g^* is essentially strictly convex. If $g \in \Gamma_0(\mathbb{R}^n)$ is essentially smooth and essentially strictly convex, we say that g is of *Legendre type* or simply a *Legendre function*. Consequently, g is Legendre if and only if g^* is so, in which case

$$\nabla g : \operatorname{int}\operatorname{dom} g \to \operatorname{int}\operatorname{dom} g^*$$

is a bijection with

$$(\nabla g)^{-1} = \nabla g^*.$$

For a generalization of these concepts to abstract Banach spaces and for some other characterizations and properties, see [3].

From now on, let $C \subset \mathbb{R}^n$ be a given open and nonempty convex set, and we assume that $g \in \Gamma_0(\mathbb{R}^n)$ satisfies the following:

$$(H_0;C) \quad \begin{cases} 1. \quad g \text{ is a Legendre function.} \\ 2. \quad \operatorname{int}\operatorname{dom} g = C. \\ 3. \quad g|_C \in C^2 \text{ and } \forall x \in C,\ G(x) := \nabla^2 g(x) \in \mathbb{S}^n_{++}. \end{cases}$$

Here, $\mathbb{S}^n_{++}$ is the class of symmetric and positive-definite $n \times n$ real matrices and $G(x) = \nabla^2 g(x)$ stands for the Hessian of g at x so that

$$G(x)_{ij} = \frac{\partial^2 g}{\partial x_i \partial x_j}(x), \quad i,j = 1,\ldots,n.$$

For such a function $g \in \Gamma_0(\mathbb{R}^n)$, the *squared Hessian* mapping $C \ni x \mapsto G(x)^2$ endows C with the Riemannian metric given by

$$\forall x \in C,\ \forall u,v \in \mathbb{R}^n,\ (u,v)^{G^2}_x := \langle G(x)^2 u, v\rangle = \langle G(x)u, G(x)v\rangle.$$

and we say that $(\cdot,\cdot)^{G^2}_x$ is the *squared Hessian Riemannian metric* on C induced by the Legendre function g. Notice that since C is an open set, then the tangent space T_xC has been identified with $\mathbb{R}^n$ for all $x \in C$. We denote by I_n the identity matrix of rank n; consequently, the metric $(\cdot,\cdot)^{I_n}_x$ is exactly the Euclidean metric $\langle\cdot,\cdot\rangle$. For sake of simplicity, when a Riemannian metric $(\cdot,\cdot)_x$ is induced by a matrix A, such as in the above cases, we write $(M,(\cdot,\cdot)_x)$ as (M,A).

2.2 Basic properties

Now we show why is interesting to study the squared Hessian Riemannian metric and what are its more remarkable consequences. Despite its simplicity, the next lemma is the basis of our analysis.

Lemma 2.2.1 *Suppose $g \in \Gamma_0(\mathbb{R}^n)$ satisfies $(H_0;C)$ and set*

$$C^* := \operatorname{int}\operatorname{dom} g^*. \tag{2.1}$$

Then

$$\nabla g : (C, G^2) \longrightarrow (C^*, I_n)$$

is an isometry between C, endowed with the squared Hessian Riemannian metric induced by g, and C^* endowed with the Euclidean metric.

Proof : Since g is a Legendre function with a non singular Hessian, it follows from the inverse function theorem that ∇g is a diffeomorphism such that $(\nabla g)^{-1} = \nabla g^*$; see [19, Theorem 26.5]. A simple calculation show us the required result. Indeed, let $u, v \in \mathbb{R}^n$ and $x \in C$,

$$(u,v)_x^{G^2} = \langle (\nabla^2 g(x))^2 u, v\rangle = \langle \nabla^2 g(x) u, \nabla^2 g(x) v\rangle = \langle d_x(\nabla g)(u), d_x(\nabla g)(v)\rangle.$$

Recalling (1.1), we get that $\nabla g : (C, G^2) \longrightarrow (C^*, I_n)$ is an isometry and the proof is complete. □

The interest of the preceding result is not the proposition on its own, but its consequences.

Proposition 2.2.1 (Geodesics) *Suppose $g \in \Gamma_0(\mathbb{R}^n)$ satisfies $(H_0;C)$. The geodesic in the squared Hessian Riemannian manifold (C, G^2) starting from $x_0 \in C$ with initial velocity $v_0 = -\nabla^2 g(x_0)^{-1} d_0$ for any $d_0 \in \mathbb{R}^n$ is given by*

$$x(t) = \nabla g^*(\nabla g(x_0) - t d_0),\ t \in [0, T_{max}). \tag{2.2}$$

In particular, we have that

$$x(t) = \operatorname{arg\,min}\left\{\langle d_0, x\rangle + \frac{1}{t} D_g(x, x_0)\right\},\ t \in (0, T_{max}), \tag{2.3}$$

where D_g is the so called Bregman pseudo-distance *induced by g which is defined by*

$$D_g(x,y) = g(x) - g(y) - \langle \nabla g(y), x - y\rangle, \quad \forall x, y \in C.$$

Moreover, $\{x(t) \mid t \in [0, T_{max})\}$*, is the* steepest descent trajectory *for the linear function* $f(x) = \langle d_0, x\rangle$ *on* (C, G)*, that is,*

$$\begin{cases} \dot{x}(t) = -\nabla^2 g(x(t))^{-1} d_0, \ t \in [0, T_{max}), \\ x(0) = x_0. \end{cases} \tag{2.4}$$

Proof : By virtue of Lemma 2.2.1, we have that $y(t) = \nabla g(x(t))$ is a geodesic in the Euclidean space C^*, that is, $\exists y_0, c_0 \in \mathbb{R}^n$, such that

$$y(t) = y_0 - tc_0.$$

Therefore, $y_0 = \nabla g(x_0)$. Furthermore, as $\dot{y}(t) = \nabla^2 g(x(t))\dot{x}(t)$, we deduce that $c_0 = -\dot{y}(0) = -\nabla^2 g(x_0)v_0 = d_0$. As g is Legendre, we have that $(\nabla g)^{-1} = \nabla g^*$ on C, and since $x(t) = (\nabla g)^{-1}(y(t))$ we get the desired formula. Moreover, it follows from (2.2) that $x = x(t)$ satisfies

$$td_0 + \nabla g(x) - \nabla g(x_0) = 0. \tag{2.5}$$

This is the first-order stationary condition for the following minimization problem:

$$\min_{x \in C} t\langle d_0, x\rangle + g(x) - \langle \nabla g(x_0), x\rangle,$$

which is equivalent, in terms of optimal solutions, to (2.3) for $t > 0$. By convexity, for $x(t)$ solves (2.3), and by strict convexity of g on C, such a solution is unique, for $t \in (0, T_{max})$.

Finally, differentiating (2.5) with respect to t we get that $d_0 + \nabla^2 g(x(t))\dot{x}(t) = 0$. Hence $x(\cdot)$ solves (2.4), which is the steepest descent trajectory for $f(x) = \langle d_0, x\rangle$ on the set C endowed with the Hessian Riemannian metric given by G, as we claimed. □

Remark 2.2.1 *The fact that for a linear objective function the Hessian Riemannian steepest descent trajectory* (2.4) *can be viewed as a sort of* central path *(*cf. (2.3)*) has been already noticed in [4, 8, 12] for the log-barrier and in [11] for a fairly general function g. For a complete study of such trajectories including the case of a nonlinear objective function f, see [1]. In particular, see [1, Theorem 4.1] for sufficient conditions to ensure that* $T_{max} = +\infty$*. The goal in this paper is to exploit the connection with the geodesics of* (C, G^2)*, which is inspired by [1, Theorem 5.11].*

Since Riemannian isometries are distance preserving, and using (2.5), we get the following:

Corollary 2.2.1 (Geodesic distance) *The geodesic distance in* (C, G^2) *between two points* $x_0, x_1 \in C$ *is given by*

$$dist(x_0, x_1) = |\nabla g(x_0) - \nabla g(x_1)|. \tag{2.6}$$

In particular, if $x(t)$ is the geodesic in (C, G^2) starting from $x_0 \in C$ with initial velocity $v_0 = -\nabla^2 g(x_0)^{-1} d_0$ then

$$dist(x(t), x_0) = t|d_0|, \quad t \geq 0. \tag{2.7}$$

On the other hand, the completeness of a Riemannian manifold as a metric space is a very desirable property. There are some conditions that ensure this property in the so called *self-concordant case*; see, for instance [13, 15]. Here we present a sort of dual characterization for squared Hessian Riemannian metrics on convex sets.

Corollary 2.2.2 (Completeness) *The metric space associated with (C, G^2) is complete if and only if $C^* = \mathbb{R}^n$ for C^* given by* (2.1).

Proof : By virtue of the Hopf-Rinow theorem (see for instance [9]), it suffices to prove that (C, G^2) is geodesically complete iff $C^* = \mathbb{R}^n$. But we have the explicit formula (2.2) for the geodesic of (C, G^2) starting from $x_0 \in C$ with initial velocity $v_0 = -\nabla^2 g(x_0)^{-1} d_0$. So the geodesic is defined for every $t \in \mathbb{R}$ if and only if $\nabla g(x_0) - t d_0 \in C^*$ for every $t \in \mathbb{R}$. Since C is an open set, we have no restriction over the vector d_0, so C^* must be the entire space $\mathbb{R}^n$. $\square$

Capítulo 3

Gradient-like algorithms for solving optimization problems

3.1 Linear search along geodesics

The steepest descent algorithm is one of the simplest method for solving unconstrained optimization problems. Nevertheless, in the case of constrained optimization problems just few algorithms are so simple as this. The main goal of this section is to introduce a simple method based on the steepest descent algorithm in order to solve a constrained optimization problem of the form:

$$(P) \quad \min\{f(x) \mid x \in \operatorname{cl} C\}.$$

We assume f is smooth enough on C and $\inf_C f > -\infty$. We also denote by $S := \arg\min\{f(x) \mid x \in \operatorname{cl} C\}$, this set may be empty. Suppose $g \in \Gamma_0(\mathbb{R}^n)$ satisfies $(H_0;C)$. Assume that the conjugate function g^* is known and set $C^* = \operatorname{int}\operatorname{dom} g^*$. Now, for a given $x_0 \in C$, let us consider the linearized optimization problem

$$(P_L;x_0) \quad \min\{\langle d_0, x\rangle \mid x \in \operatorname{cl} C\}$$

where

$$d_0 = \nabla_G f(x_0) = \nabla^2 g(x_0)^{-1} \nabla f(x_0),$$

that is, d_0 is the gradient of f with respect to the metric induced by $G = \nabla^2 g$ at x_0. Since the geodesics of (C, G^2) are steepest descent trajectories in (C, G) for some appropriate linear function (cf. Proposition 2.2.1), if $T_{max} = +\infty$, it follows that $(P_L;x_0)$ can be solved asymptotically by the trajectory

$$\mathbb{R}_+ \ni t \mapsto x(t) = \nabla g^*(\nabla g(x_0) - t d_0), \tag{3.1}$$

as long as $t \to +\infty$; see [1, Theorem 4.7].

Exemple 3.1.1 *For $C = \mathbb{R}^n$, if $g(x) = \frac{1}{2}|x|^2$ then $g^*(y) = \frac{1}{2}|y|^2$, and so the trajectory (3.1) amounts to*

$$\mathbb{R}_+ \ni t \mapsto x(t) = x_0 - t\nabla f(x_0). \tag{3.2}$$

Let us recall here that the standard Euclidean steepest descent method consists in minimizing iteratively the function f along the path (3.2). Motivated by such a *linear search* technique, we propose an algorithm that follows the trajectory (3.1) in order to get a significant decrease on the objective function f as well as being strictly feasible by ensuring $x(t) \in C$. The general scheme for such an algorithm is the following:

1. Choose $x_0 \in C$.
2. **If** $d_0 = \nabla_G f(x_0) = 0$ then x_0 solves the problem $\to$ **stop**,
   else $\to$ take $x(t) = \nabla g^*(\nabla g(x_0) - t d_0)$.
3. Choose step $t_0 > 0$ such that $f(x_0) > f(x(t_0))$ and $x(t_0) \in C$.
4. Set $x_0 = x(t_0)$, return 2.

For simplicity, in the sequel we assume that f is a $\mathcal{C}^2$ function and that $T_{max} = +\infty$. Most of the times we will assume $f + \chi_{\mathrm{cl}\,C}$ is coercive, where χ_A is the characteristic function of the set A. Consider, for any $x_0 \in C$, the following sequence

$$x_{k+1} = \nabla g^*(\nabla g(x_k) - t_k d_k) \qquad \forall k \geq 0 \tag{3.3}$$

where $d_k = \nabla_G f(x_k)$ and $t_k > 0$ is such that $x(t_k) \in C$ and $f(x_{k+1}) < f(x_k)$ for all $k \geq 0$ when $d_k \neq 0$.

Proposition 3.1.1 *The sequence* (3.3) *is well-defined and $\{f(x_k)\}_{k \geq 0}$ is decreasing.*

Proof : Let us fix $k \geq 0$ such that $x_k \in C$; it is enough to prove the existence of a step-size parameter $t_k > 0$ small enough such that $f(x_k) > f(x(t_k))$ and $x(t_k) \in C$. Let us define $\phi(t) = f(x(t))$, with $x(t) = \nabla g^*(\nabla g(x_k) - t d_k)$, it follows that:

$$\begin{aligned}
\frac{d}{dt}\phi(t) &= \langle \nabla f(x(t)), \frac{d}{dt}x(t) \rangle \\
&= -\langle \nabla f(x(t)), \nabla^2 g^*(\nabla g(x_k) - t d_k) d_k \rangle \\
&= -\langle \nabla f(x_k) + o(|x(t) - x_k|), \nabla^2 g^*(\nabla g(x_k) - t d_k) d_k \rangle \\
&= -\langle \nabla f(x_k), \nabla^2 g^*(\nabla g(x_k) - t d_k) d_k \rangle + o(|x(t) - x_k|)
\end{aligned}$$

Since $\nabla^2 h(x) \in \mathbb{S}^n_{++}$ for all $x \in C$, we have $\nabla^2 g(x) \in \mathbb{S}^n_{++}$ for all $x \in C$ and therefore

$$\nabla^2 g^*(y) = [\nabla^2 g(\nabla g^*(y))]^{-1} \in \mathbb{S}^n_{++} \text{ for all } y \in C^*.$$

Consequently, by continuity of $x(\cdot)$ and bringing to mind that $x(0) = x_k \in C$, we have the existence of certain $\delta > 0$ such that $\frac{d}{dt}\phi(t) < 0$ and $x(t) \in C$ for all $t \in [0, \delta)$. Finally, by taking $t_k \in (0, \delta)$ we get that the general algorithm is well-defined. □

The reason why we choose the descent direction d_k as the Riemannian Gradient of the objective function is the dual scheme we propose in the next section, because, as we will prove, the Riemannian Gradient of a function, can be seen as a Euclidean Gradient under the Legendre change of coordinates and this fact is the base of all we do. Nevertheless, it is possible to explore the same kind of algorithm with another descent direction. Indeed, indirectly, Auslender and Teboulle in [2] did it with d_k as the Euclidean Gradient of the objective function. The algorithm they propose is

$$x_{k+1} \in \arg\min\{\lambda_k \langle g_k, x\rangle + d(x, x_k) : \ x \in \operatorname{cl} C\} \quad k = 0, 1, 2, \ldots$$

where $g_k \in \partial f(x_k)$ and d is a proximal distance; see [2, Definition 2.1]. Using d as the Bregman distance induced by a Legendre Function we get the same kind of iteration we propose since the geodesics of (C, G^2) are central path in the sense of Proposition 2.2.1. They also prove under suitable hypothesis that this algorithm is valued-convergent; for more detail see [2, Theorem 4.1].

3.2 Dual scheme

In order to study the convergence of the method, we start by noticing that the sequence (3.3) can be seen as an Euclidean descent method in a certain *dual set*. Define $\varphi : C^* \to \mathbb{R}$ by

$$\varphi(y) = f(\nabla g^*(y)),$$

then

$$\nabla\varphi(y) = \nabla^2 g^*(y) \nabla f(\nabla g^*(y)), \quad \forall y \in C^*.$$

Let

$$y_k = \nabla g(x_k)$$

and notice that the sequence (3.3) changes into

$$y_{k+1} = y_k - t_k d_k, \quad \forall k \geq 0. \tag{3.4}$$

Let us recall that $\nabla^2 g^*(y) = [\nabla^2 g(x)]^{-1}$ for any $x \in C$ such that $\nabla g(x) = y$, hence

$$\nabla\varphi(y) = \nabla_G f(x).$$

This means that the Euclidean gradient of φ is equal to the Riemannian gradient of f in (C,G). Therefore, we have that the algorithm can be seen as an Euclidean steepest descent method for the function φ, thas is,

$$y_{k+1} = y_k - t_k \nabla\varphi(y_k) \qquad \forall k \geq 0. \tag{3.5}$$

Let us recall the so-called Zoutendijk condition in this context: $\exists r > 0$, $\forall k \geq 0$ such that

$$(Z) \quad \varphi(x_{k+1}) \leq \varphi(x_k) - r|\nabla\varphi(x_k)|^2 \cos^2\theta_k, \quad \text{with } \cos\theta_k = -\frac{\langle \nabla\varphi(x_k), d_k \rangle}{|\nabla\varphi(x_k)||d_k|}.$$

This condition is widely used to ensure the convergence (in some sense) of methods of descent directions. In this case $\cos\theta_k = 1$, so if we choose a step-size t_k satisfying (Z), for example with the rule of Cauchy, Armijo or Wolfe, see [10, 14], if φ is bounded below we have in particular

$$\sum_{k\in\mathbb{N}} |\nabla\varphi(y_k)|^2 < +\infty,$$

and consequently $\nabla\varphi(y_k) \to 0$. On the other hand, if we set $y(t) = \nabla g(x(t))$, we get $f(x(t)) = \varphi(y(t))$ for all $t \geq 0$ where $x(\cdot)$ is well defined. So every way we can pick the step-size t_0 for the Euclidean gradient method gives us a way to pick the step-size in the Riemannian method.

These remarks motivate the introduction of a sort of *dual problem* associated with the initial problem (P), which is given by

$$(D) \quad \inf\{\varphi(y) \mid y \in C^*\}.$$

By virtue of the Legendre coordinate transformation $y = \nabla g(x)$, we have that

$$\forall x \in C,\ \exists y \in C^* \text{ such that } f(x) = \varphi(y) \tag{3.6}$$

$$\forall y \in C^*,\ \exists x \in C \text{ such that } f(x) = \varphi(y) \tag{3.7}$$

In fact, it is straightforward to prove the following proposition

Proposition 3.2.1 *We have that*

$$\inf\{f(x) \mid x \in C\} = \inf\{\varphi(y) \mid y \in C^*\}.$$

Moreover, $\{x_k\}$ is a minimizing sequence for (P) if and only if $\{y_k\}$ with $y_k = \nabla g(x_k)$ is a minimizing sequence for (D)

Remark 3.2.1 *The duality scheme proposed here is different from the classical notions of duality in Convex Analysis theory via perturbations or Lagrangians functions. The reason is very simple: in the classical cases, the convexity of the problem is preserved under duality. Instead, in our case, the dual (resp. primal) objective function is not necessarily convex even if the primal (resp. dual) objective function is convex.*

Since C^* is an open set, is natural to ask why the dual problem proposed do not look for a solution on the boundary of C^*, there are two answers, the first one is, if we assume the Hessian Riemannian manifold (C, G^2) is complete, then C^* is the whole space $\mathbb{R}^n$, and the second one, if $f + \chi_{\mathrm{cl}\,C}$ coercive then the function φ blows up on the boundary of C^* as we will see in Corollary 3.2.1.

Some other desirable properties are valid if we assume f convex and S is bounded. For instance, it might happen that the function φ had a local minimum that was not global, since ∇g^* does not preserve necessarily the convexity of f, but as $\nabla^2 g^*$ is invertible this can not happen.

Proposition 3.2.2 *We have $\bar{x} \in S \cap C$ if and only if $\bar{y} = \nabla g(\bar{x}) \in C^*$ is a solution of (D). Moreover, let $\bar{y}$ be a local minimum of φ then $\bar{y} \in C^*$ and it is global.*

Proof : First, let us note, by the one-to-one relationship between each $x \in C$ and each $y \in C^*$ described in (3.6) and (3.7), that if $\bar{x} \in S \cap C$ we have

$$\varphi(\bar{y}) = f(\bar{x}) \leq f(x) \qquad \forall x \in C,$$

it follows that $\varphi(\bar{y}) \leq \varphi(y)$ for all $y \in C^*$, so $\bar{y}$ solves (D). Similarly, we get that if $\bar{y}$ solves (D)

$$f(\bar{x}) \leq f(x) \qquad \forall x \in C,$$

and by continuity we extend the result to the entire clousure of C.

Let $\bar{y}$ be a local minimum of φ and suppose $\bar{y} \in C^*$, then $\nabla\varphi(\bar{y}) = 0$, but $\nabla\varphi(\bar{y}) = \nabla^2 g^*(\bar{y})\nabla f(\bar{x})$ where $\bar{x} = \nabla g^*(\bar{y})$. Since $\nabla^2 g^*$ is invertible on C^* we have $\nabla f(\bar{x}) = 0$ and as f is differentiable and convex we have $\bar{x} \in S \cap C$, so $\bar{x}$ is a global minimun of f and therefore $\bar{y}$ is a global minimun of φ.

Now, if $\bar{y} \in \partial C^*$ we can take a sequence $\{y_k\} \subseteq C^*$ such that $y_k \to \bar{y}$, then fix $x_k = \nabla g^*(y_k)$ then since g^* is a Legendre function $|x_k| \to \infty$, but since y_k is close enough to the local minimum $\bar{y}$ we can assume $\varphi(y_k)$ is uniformly bounded and so $f(x_k)$, therefore there exists $\nu \in \mathbb{R}$ such that $x_k \in [f \leq \nu]$ for every $k \geq 0$, and since f is convex on $\mathrm{cl}\,C$ and S is bounded we get a contradiction. $\square$

Let us stop for a moment in the problem (D). From the definition of this problem it is not clear if, in case the minimum is attained, it could be on the boundary of C^* or not. The answer is that under suitable hypothesis, for example if $f+\chi_{\text{cl}\,C}$ is coercive, the dual problem in case of have solution, it is in the interior of C^* because the dual function blows up on the boundary of C^*.

Corollary 3.2.1 *If $f+\chi_{cl\,C}$ is coercive then $\varphi|_{\partial C^*}=+\infty$.*

Proof : If there is a point $y^*\in\partial C^*$ such that $\varphi(y^*)<+\infty$ then we can take a sequence $\{y_k\}$ such that $y_k\to y^*$, and using the same argument from above we may find a sequence in C, namely $\{x_k\}$ such that $|x_k|\to+\infty$. Contradictorily, the sequence is inside of a level set, which is supposed to be bounded. So there is no such y^*. □

3.3 Convexity properties and value convergence

In this section we investigate the particular case where the function f is convex on (C,G^2) in the Riemannian sense. Now we recall a convergence result for the gradient method in the convex case; see, for instance [5] and the references therein for related topics. For the sake of completeness, we provide the proof.

Proposition 3.3.1 *Let $C^*\subseteq\mathbb{R}^n$ be nonempty, open and convex. Let $\varphi\in\Gamma_0(\mathbb{R}^n)$ be such that $\varphi|_{C^*}\in\mathcal{C}^1(C^*)$. Suppose that φ is bounded from below so that in particular $\inf_{C^*}\varphi>-\infty$, and assume that $\nabla\varphi$ is Lipschitz on the level sets of φ. Let $\{y_k\}\subset C^*$ be a sequence such that $\{\varphi(y_k)\}$ is nonincreasing and*

$$y_{k+1}=y_k-h_k\nabla\varphi(y_k),\quad k=0,1,2,...$$

starting from $y_0\in\mathbb{R}^n$, $h_k\in[0,h_{max}]$ with $0<h_{max}<\frac{2}{L}$, where L is the Lipschitz constant of $[\varphi\leq\varphi(y_0)]$ and $\sum h_k=\infty$. Then

1. *$\varphi(y_k)\to\inf_{C^*}\varphi$ as long as $k\to\infty$.*
2. *If $\arg\min\varphi_{C^*}\neq\emptyset$, $\exists y_\infty\in\arg\min\varphi_{C^*}$ such that $y_k\to y_\infty$.*

Proof : First of all, let us note that since φ is convex, $Conv\{y_k\}\subseteq[\varphi\leq\varphi(y_0)]$, then we can assume that $\nabla\varphi$ is L lipschitz in $Conv\{y_k\}$.

Let $v(t):=\varphi(y_k+t(y_{k+1}-y_k))$, $t\in[0,1]$, then

$$v(1)=v(0)+\int_0^1\frac{dv}{dt}(t)dt.$$

In other words we have:

$$\begin{aligned}
\varphi(y_{k+1}) &= \varphi(y_k) + \int_0^1 \langle \nabla\varphi(y_k + t(y_{k+1}-y_k)), y_{k+1}-y_k\rangle dt \\
&= \varphi(y_k) + \int_0^1 \langle \nabla\varphi(y_k + t(y_{k+1}-y_k)) - \nabla\varphi(y_k) + \nabla\varphi(y_k), y_{k+1}-y_k\rangle dt \\
&= \varphi(y_k) + \langle \nabla\varphi(y_k), y_{k+1}-y_k\rangle + \int_0^1 \langle \nabla\varphi(y_k + t(y_{k+1}-y_k)) - \nabla\varphi(y_k), y_{k+1}-y_k\rangle dt \\
&\leq \varphi(y_k) - h_k|\nabla\varphi(y_k)|^2 + \int_0^1 Lt|y_{k+1}-y_k|^2 dt \\
&= \varphi(y_k) - \left[1 - \frac{L}{2}h_k\right] h_k|\nabla\varphi(y_k)|^2 \\
&\leq \varphi(y_k) - \left[1 - \frac{L}{2}h_{max}\right] h_k|\nabla\varphi(y_k)|^2
\end{aligned}$$

As φ is bounded from below, $\{\varphi(y_k)\}$ is convergent to a real number. On the other hand, it follows that

$$\left[1 - \frac{L}{2}h_{max}\right] \sum_{k=0}^{N} |\nabla\varphi(y_k)|^2 h_k \leq \varphi(y_0) - \varphi(y_{N+1}) \leq \varphi(y_0) - \inf_{C^*}\varphi < \infty, \quad \forall N \in \mathbb{N}.$$

The latter implies that $\{|\nabla\varphi(y_k)|^2 h_k\}$ is summable. Next, let $z \in C^*$, then

$$\begin{aligned}
|y_{k+1}-z|^2 - |y_k - z|^2 &= \langle y_{k+1} + y_k - 2z, y_{k+1} - y_k\rangle \\
&= |y_{k+1}-y_k|^2 + 2\langle y_k - z, y_{k+1}-y_k\rangle \\
&\leq |y_{k+1}-y_k|^2 + 2h_k[\varphi(z) - \varphi(y_k)],
\end{aligned}$$

where we have used the convexity of φ. Let us set $\delta_k := |y_{k+1}-y_k|^2$, which is a summable sequence; indeed,

$$\frac{1}{h_{max}} \sum_{k=0}^{N} |y_{k+1}-y_k|^2 \leq \sum_{k=0}^{N} \frac{1}{h_k}|y_{k+1}-y_k|^2 = \sum_{k=0}^{N} |\nabla\varphi(y_k)|^2 h_k \leq \varphi(y_0) - \inf_{C^*}\varphi < \infty, \quad \forall N \in \mathbb{N}.$$

Also we know that

$$2\sum_{k=0}^{N} h_k[\varphi(y_k) - \varphi(z)] \leq |y_0 - z|^2 - |y_{N+1} - z|^2 + \sum_{k=0}^{N} \delta_k \leq |y_0 - z|^2 + \sum_{k=0}^{N} \delta_k.$$

This implies, for each $z \in C^*$, that we have

$$\sum_{k=0}^{\infty} h_k[\varphi(y_k) - \varphi(z)] < \infty,$$

and so $\liminf_{k\to\infty} \varphi(y_k) \leq \varphi(z)$. Therefore, $\liminf_{k\to\infty} \varphi(y_k) \leq \inf_{C^*}\varphi$ but $\{\varphi(y_k)\}$ is convergent and by hypothesis $\{y_k\} \subset C^*$, so it limit must be $\inf_{C^*}\varphi$ as claimed.

Let us assume $S^* = \arg\min_{C^*} \varphi \neq \emptyset$ and set $z \in S^*$. Then

$$|y_{k+1} - z|^2 - |y_k - z|^2 \leq \delta_k + 2h_k[\varphi(z) - \varphi(y_k)] \leq \delta_k.$$

Since δ_k is summable, the term $\theta_k = |y_k - z|^2 + \sum_{j=k}^{\infty} \delta_j$ is convergent for each $z \in S^*$, therefore the limit $\lim_{k\to\infty} |y_k - z|$ exists. On the other hand, we know that $\{y_k\}$ is bounded, so it must have just one accumulation point, indeed, let $\bar{y}$ and $\tilde{y}$ be two accumulation points, then $\lim_{k\to\infty} |y_k - \bar{y}| = \lim_{j\to\infty} |y_{k_j} - \bar{y}| = 0$, where $y_{k_j} \to \bar{y}$, if we evaluate the limit in the subsequence that converges to $\tilde{y}$ we get that $|\bar{y} - \tilde{y}| = 0$. □

Now, let us assume f is convex on (C, G^2) in the Riemannian sense, then we claim the dual function φ is convex in the classical sense, indeed, we know an expression for every geodesic on C and it is given by

$$\begin{cases} \gamma_{xy}(t) &= \nabla g^*(\nabla g(x) + t(\nabla g(y) - \nabla g(x))), \\ \gamma_{xy}(0) &= x \in C, \\ \gamma_{xy}(1) &= y \in C. \end{cases}$$

Then, since $f = \varphi \circ \nabla g$ and if we set $u = \nabla g(x)$, $v = \nabla g(y)$ we have

$$f \circ \gamma_{xy}(t) = \varphi(v + t(v - u)).$$

Moreover,

$$t f(y) + (1-t) f(x) = t\varphi(v) + (1-t)\varphi(u),$$

so φ is convex in classical sense on C^*.

Taking into account the above, we have our first convergence theorem:

Theorem 3.3.1 *Let us assume f convex on (C, G^2) in the Riemannian sense. Let $\{x_k\}$ the sequence induced by (3.3) starting from $x_0 \in C$. Assume f is bounded from below on $cl\,C$, $\nabla_G f$ is Lipschitz on the level sets of f, $t_k \in [0, h_{max}]$ with $0 < h_{max} < \frac{2}{L}$ and $\sum t_k = \infty$ where L is the Lipschitz constant on $[f \leq f(x_0)]$. Then:*

1. *$f(x_k) \to \inf_C f$ as long as $k \to \infty$.*
2. *If $\arg\min f \cap C \neq \emptyset$ then the algorithm converges to $x_\infty \in \arg\min f \cap C$.*

Proof : By the above φ is a $\mathcal{C}^1$ and convex function on C^*, moreover, φ is L-Lipschitz continuous on level sets of φ, indeed, we know that

$$dist(x, y) = |\nabla g(x) - \nabla g(y)|,$$

and so

$$\begin{aligned}|\nabla\varphi(u)-\nabla\varphi(v)| &= |\nabla_G f(\nabla g^*(u))-\nabla_G f(\nabla g^*(u))| \\ &\leq L dist(\nabla g^*(u),\nabla g^*(v)) \\ &\leq L|u-v|\end{aligned}$$

Therefore, by Proposition 3.3.1 we have $\varphi(y_k)\to\inf\varphi$ from where comes the first conclusions using the duality relations between f and φ. Besides, if $\arg\min f\cap C\neq\emptyset$ then $S:=\arg\min\varphi\neq\emptyset$, so there exists $y_\infty\in S$ such that $y_k\to y_\infty$, then x_k is a minimizing sequences for the problem $\inf\{f(x)\mid x\in C\}$ which converges to $x_\infty:=\nabla g^*(y_\infty)$. □

Let us consider the problem

$$(P_1)\quad \min\left\{\frac{1}{2}\log^2 x_1+\exp\left(\tan x_2\sec^2 x_2\right)\mid 0\leq x_1,-\frac{\pi}{2}\leq x_2\leq\frac{\pi}{2}\right\}$$

then $C=(0,+\infty)\times(-\frac{\pi}{2},\frac{\pi}{2})$. Thus, if we set $g(x_1,x_2)=x_1\log x_1-x_1+\frac{1}{2}\tan^2 x_2$, it is easy to see that this function satisfies $(H_0;C)$, then the dual function is given by $\varphi(y_1,y_2)=\frac{1}{2}y_1^2+\exp y_2$ because

$$\nabla g(x_1,x_2)=\left[\log x_1,\tan x_2\sec^2 x_2\right].$$

Therefore, the dual problem is written as

$$(D_1)\quad \min\left\{\frac{1}{2}y_1^2+\exp y_2\mid y_1,y_2\in\mathbb{R}\right\}.$$

Let us remark that (P_1) is not a convex problem, but (D_1) it is. Also, to have an idea of the relation between the solution of the primal and dual problem, let us note that the solution of (P_1) is $(1,-\frac{\pi}{2})\in\partial C$ and the solution of (D_1) is not attained, but $\{(0,-k)\}_k$ is a minimizing sequences for the dual problem and consequently, $val(P_1)=val(D_1)$.

3.4 Asymptotic behavior under Euclidean convexity

Auslender and Teboulle proved in [2] that our algorithm is valued-convergent in case when f is convex in the classical sense and d_k is the Euclidean gradient of the f. Here, with another set of hypothesis we ensure that the algorithm is value-convergent.

Proposition 3.4.1 *Let us assume f is convex on clC and $f+\chi_{clC}$ is coercive. Suppose $\nabla\varphi(y_k)\to 0$ and $dist(x_k,\partial C)>\alpha>0\ \forall k\geq 0$, then $f(x_k)\to\inf f$ as $k\to\infty$.*

Proof : Let us note that $\{\varphi(y_k)\}$ is decreasing and bounded below, so $\ell := \lim \varphi(y_k)$ is well-defined. Now, since f is convex on $\mathrm{cl}\, C$ and differentiable we know

$$f(x_k) + \langle \nabla f(x_k), u - x_k \rangle \leq f(u) \qquad \forall u \in C$$

and using the transformation ∇g we can write the gradient of f in terms of φ and y_k as

$$\varphi(y_k) + \langle \nabla^2 g(x_k) \nabla \varphi(y_k), \nabla g^*(v) - x_k \rangle \leq \varphi(v) \qquad \forall v \in C^*.$$

Hence, as $\{x_k\}$ is away from ∂C, we can deduce that $\{\nabla^2 g(x_k)\}$ is bounded

$$\langle \nabla^2 g(x_k) \nabla \varphi(y_k), \nabla g^*(v) - x_k \rangle \to 0 \qquad k \to \infty$$

and

$$\lim \varphi(y_k) = \ell \leq \varphi(v) \qquad \forall v \in C^*.$$

Therefore, $\{y_k\}$ is a minimizing sequence for φ and then by proposition 3.2.1 we conclude. □

Now we analyze the behavior of the algorithm and the relationship with the dual sequences $\{y_k\}$. First we state some simple facts

Proposition 3.4.2 *Let us assume f is convex on $\mathrm{cl}\, C$, S is bounded and $t_k \geq \alpha > 0$. Let $\{y_k\}$ be the dual sequences, recall that $y_k = \nabla g(x_k)$ for each $k \in \mathbb{N}$. Then:*

1. *If $\{y_k\}$ is such that $|y_{k+1} - y_k| \to 0$, then $d_k \to 0$.*
2. *If $\{y_k\}$ converges to some y_∞, then $y_\infty \in C^*$ and $\{x_k\}$ converges to $x_\infty \in S \cap C$.*
3. *If $\{x_k\}$ converges to $\bar{x} \in C$, then $\bar{x} \in S$.*

Proof :

1. From (3.3) follows that
$$y_{k+1} = y_k - t_k d_k \qquad \forall k \geq 0 \tag{3.8}$$
therefore,
$$|d_k| = \frac{1}{t_k}|y_{k+1} - y_k| \leq \frac{1}{\alpha}|y_{k+1} - y_k| \to 0 \text{ as } k \to \infty \tag{3.9}$$

2. Now, if the sequence is convergent, in particular, we have $d_k \to 0$. Let us assume $y_\infty \in C^*$, since g is of class $\mathcal{C}^2$ we have $x_k = \nabla g^*(y_k) \to \nabla g^*(y_\infty)$. If we take $x_\infty = \nabla g^*(y_\infty)$, by the above we have $\nabla_G f(x_\infty) = 0$ then $\nabla f(x_\infty) = 0$, and so $x_k \to x_\infty \in S \cap C$.

   Assuming the absurd, let us suppose $y_\infty \in \partial C^*$ it means that $|x_k| = |\nabla g^*(y_k)| \to \infty$ because g^* is a Legendre function. Moreover, by definition $x_k \in [f \leq f(x_0)]$. Hence, since S is bounded and f is convex on $\mathrm{cl}\,C$, every level set is also bounded, then we get a contradiction.

3. Finally, if $\{x_k\}$ converges to $\bar{x} \in C$, then $\{y_k\}$ converges to $y_\infty = \nabla g(\bar{x})$. Using the previous part we conclude the proof.

□

We can split the behavior of the dual sequence $\{y_k\}$, when it is bounded and when it is not. If the dual sequence is bounded then, using the same arguments of point 2 of Proposition 3.4.2 we find that the clusters points of $\{y_k\}$ are in C^*, and so, some of the clusters points of $\{x_k\}$ are in $S \cap C$. So the question is what happens if the dual sequence is unbounded? The next proposition provides us an answer.

Proposition 3.4.3 *Let us assume f is convex on $cl\,C$ and S is bounded. Let us suppose $|y_k| \to \infty$. Then $dist(x_k, \partial C) \to 0$.*

Proof : Let us assume $\exists \varepsilon > 0$ such that $dist(x_k, \partial C) \geq \varepsilon$, for all $k \geq 0$. By construction, we know that $\{x_k\}$ is bounded, then there exists a compact set $K \subseteq C$ such that $x_k \in K$ for each $k \geq 0$, therefore $\nabla g(x_k)$ is also bounded, but this is a contradiction because $y_k = \nabla g(x_k)$. □

Capítulo 4

Restriction to affine subspaces

4.1 Framework

Let C be an open and convex subset in $\mathbb{R}^n$, A be an $m \times n$ real matrix with $m \leq n$ and b be a real vector in $\mathbb{R}^n$. Let us define

$$\mathcal{F} = C \cap \mathcal{A},$$

where $\mathcal{A} = \{x \in \mathbb{R}^n : Ax = b\}$. We also denote by $\mathcal{A}_0 = \ker A$ and $\mathcal{F}^* = \Pi_{\mathcal{A}_0} C^*$. Hereinafter we will put all our attention to the submanifold $\mathcal{F}$ in order to propose an algorithm for solving problems of the form

$$(P_1) \quad \min\{f(x) : x \in \operatorname{cl} C, \, Ax = b\}.$$

Let $g \in \Gamma_0(\mathbb{R}^n)$ be a Legendre function that satisfies the basic condition $(H_0; C)$. We recall from [1, Section 5.2] the Legendre transform coordinates mapping on $\mathcal{F}$ associated with g is denoted by ϕ_g and is given by

$$\begin{array}{rccl} \phi_g : & \mathcal{F} & \to & \mathcal{F}^* \\ & x & \mapsto & \phi_g(x) = \Pi_{\mathcal{A}_0} \nabla g(x). \end{array}$$

We also recall that ϕ_g is a diffeomorphism from $\mathcal{F}$ to $\mathcal{F}^*$ and, that $d\phi_g(x) = \Pi_{\mathcal{A}_0} G(x)$ and $d\phi_g(x)^{-1} = \sqrt{G(x)^{-1}} \Pi_{\sqrt{G(x)}\mathcal{A}_0} \sqrt{G(x)^{-1}}$.Also, we have a formula for the inverse of the map ϕ_g and which is given by

$$\phi_g^{-1}(y) = \nabla[g^* \Box (\chi_{\mathcal{A}_0^\perp} + \langle \cdot, \tilde{x} \rangle)](y), \quad \text{for any } \tilde{x} \in \mathcal{A},$$

with $g_1 \Box g_2$ the *epigraphical sum* of two function. For more details see [1, Section 5].

Now, let us consider on $\mathcal{F}$ the Riemannian metric $(\cdot,\cdot)_x^{G^2}$ given by

$$(u,v)_x^{G^2} = \langle \Pi_{\mathcal{A}_0} G(x)u, \Pi_{\mathcal{A}_0} G(x)v \rangle \quad \forall x \in \mathcal{F},\ u,v \in T_x\mathcal{F} \simeq \mathcal{A}_0.$$

Then, as in Lemma 2.2.1, it is easy to see that ϕ_g is an isometry between $\mathcal{F}$ endowed with the Riemannian metric $(\cdot,\cdot)^{G^2}$ and $\mathcal{F}^*$ endowed with the Euclidean metric, and so we have a formula for the geodesic of $(\mathcal{F}, G^2)$ and it follows the same logic than the geodesic of (C, G^2), namely, the geodesic starting from x_0 with initial velocity $-d\phi_g(x_0)\Pi_{\mathcal{A}_0} d_0$ is given by

$$x(t) = \phi_g^{-1}(\phi_g(x_0) - t\Pi_{\mathcal{A}_0} d_0). \tag{4.1}$$

4.2 Properties of geodesics

It is natural to wonder if the same kind of results are stated for this case. Therefore, here we are going to specify the analogous properties of the geodesics on $\mathcal{F}$, we skip its proof and we give just a scheme of proof because they are essentially the same.

Corollary 4.2.1 *The geodesics on $(\mathcal{F}, G^2)$ with initial velocity $-d\phi_g(x_0)^{-1}\Pi_{\mathcal{A}_0} c$ solves asymptotically the linear optimization problem* $\inf\{\langle c,x\rangle :\ x \in C,\ Ax = b\}$.

The reason is the same as in section 5, that is, because the unique solution of

$$\text{(H-SD)} \qquad \begin{cases} \dot{x}(t) = -\nabla_G f|_{\mathcal{F}}(x(t)) & t \in (0, T_{max}) \\ x(0) = x_0 \end{cases}$$

where $f(x) = \langle c,x\rangle$ is the unique geodesic passing through x_0 with initial velocity $-d\phi_g(x_0)^{-1}\Pi_{\mathcal{A}_0} c$.

The question of whether or not Riemannian manifold is complete is also easy to respond in this new context, moreover, the arguments to prove the following corollary are exactly the same.

Corollary 4.2.2 *The Riemannian submanifod $\mathcal{F}$ endowed with the Riemannian metric given by G^2 is complete if and only if $C^* = \mathbb{R}^n$, or well, $\mathcal{F}^* = \mathcal{A}_0$.*

As in section 5, we just need that $\phi_g(x_0) - t\Pi_{\mathcal{A}_0} c = \Pi_{\mathcal{A}_0}(\nabla g(x_0) - tc) \in \mathcal{F}^*$ for all $t \in \mathbb{R}$. Now, since $\mathcal{F}^* = \Pi_{\mathcal{A}_0} C^*$ and $\mathcal{A}_0$ is a vectorial subspace, we are done.

Let $x_0, x_1 \in \mathcal{F}$ and set $y_0 = \phi_g(x_0)$ and $y_1 = \phi_g(x_1)$. Then as in section 5, it follows that $\rho(x_0,x_1) = \ell_0^1(y)$ where $\rho(x_0,x_1)$ denotes the geodesic distance between x_0 and x_1 in the Riemannian submanifold $\mathcal{F}$, $y(t) = \phi_g(x_0) - t\Pi_{\mathcal{A}_0} c$ with $c = \phi_g(x_0) - \phi_g(x_1)$ and $\ell_0^1(\cdot)$ is the length of a curve defined on $[0,1]$, inasmuch as $\phi_g(x_0), \phi_g(x_1) \in \mathcal{A}_0$ we have $y(t) = (1-t)\phi_g(x_0) + t\phi_g(x_1)$, and so

$$\rho(x_0,x_1) = |\phi_g(x_0) - \phi_g(x_1)|_2. \tag{4.2}$$

From corollary 4.2.1 we know that the geodesic starting from $x_0 \in \mathcal{F}$ with initial velocity $-d\phi_g(x_0)^{-1}\Pi_{\mathcal{A}_0} c$ is the unique solution of

$$\text{(H-SD)} \quad \begin{cases} \dot{x}(t) = -\nabla_G f|_{\mathcal{F}}(x(t)) \\ x(0) = x_0 \end{cases}$$

where $f(x) = \langle c, x \rangle$. Therefore, it is easy to see that $x(\cdot)$ also satisfies the following differential inclusion

$$\begin{cases} \frac{d}{dt}\nabla g(x(t)) + \nabla f(x(t)) \in \mathcal{A}_0^{\perp} \\ x(t) \in \mathcal{F} \\ x(0) = x_0 \end{cases}$$

Then integrating between 0 and t we get:

$$\nabla g(x(t)) - \nabla g(x_0) + ct \in \mathcal{A}_0^{\perp}, \tag{4.3}$$

and so the same conclusion follows, that is,

$$0 \in \partial(\langle c, \cdot \rangle + \frac{1}{t} D_g(\cdot, x_0) + \chi_{\mathcal{A}_0})(x(t))$$

4.3 Algorithm for solving constrained optimization problems

In section 3 we have proposed an algorithm in order to solve problems of the form:

$$(P_0) \quad \min\{f(x) : \ x \in \operatorname{cl} C\}.$$

The key in the good behavior of that algorithm lies in the dual approach we proposed. We recall that the previous algorithm can be seen as an Euclidean descent method in the space C^*. Therefore, in order to solve a more general kind of problems than (P_0) we follow the same logic, that is, we follow the geodesics of $\mathcal{F}$ in a certain direction for a while, then we change the direction and do the same.

Formally, let us consider the problem

$$(P) \quad \min\{f(x) : \ x \in \operatorname{cl} C, \ Ax = b\},$$

with f a differentiable function on C such that $S := \arg\min\{f(x) : x \in \mathrm{cl}\, C,\ Ax = b\} \neq \emptyset$. Assume g is a Legendre functions as before. Now, for a given $x_0 \in \mathcal{F}$, let us consider the linear optimization problem

$$(P_L) \quad \min\{\langle \nabla_G f(x_0), x\rangle : x \in \mathrm{cl}\, C,\ Ax = b\}.$$

Here we have chosen the Riemannian gradient of f with respect G just for sake of simplicity. Then the general scheme for this algorithm is the following:

1. choose $x_0 \in \mathcal{F}$.
2. **if** x_0 solves (P) $\rightarrow$ **stop**,
   else $\rightarrow$ compute $x(t) = \phi_g^{-1}(\phi_g(x_0) - t\Pi_{\mathcal{A}_0}\nabla_G f(x_0))$.
3. choose step $t_0 > 0$ such that $f(x_0) > f(x(t_0))$ and $x(t_0) \in \mathcal{F}$.
4. set $x_0 = x(t_0)$, return 2.

Let us consider the function defined by $\varphi(y) := f \circ \phi_g^{-1}(y)$ for each $y \in \mathcal{F}^*$. It easy to see that

$$\nabla f(x) = \nabla g^2(x)\Pi_{\mathcal{A}_0}\nabla\varphi(y), \quad \text{where } y = \phi_g(x).$$

Thus, $\nabla_G f(x_0) = \Pi_{\mathcal{A}_0}\nabla\varphi(y_0)$ where $y_0 = \phi_g(x_0)$, and so $\nabla_G f(x_0) = 0$ for $x_0 \in \mathcal{F}$ if and only of $\nabla\varphi(y_0) \in \mathcal{A}_0^{\perp}$ and that if and only if $y_0 = \phi_g(x_0) \in \mathcal{F}$ is an interior (relative) solution of

$$(D) \quad \min\{\varphi(y) : y \in \mathcal{F}^*\}.$$

Moreover, if y_0 is not an interior (relative) solution of (D), the mapping

$$t \in \mathbb{R}_+ \mapsto \alpha(t) := \varphi(y_0 - t\Pi_{\mathcal{A}_0}\nabla\varphi(y_0))$$

is not increasing, that is because

$$\frac{d}{dt}\alpha(t) = -\langle \nabla\varphi(y_0 - t\Pi_{\mathcal{A}_0}\nabla\varphi(y_0)), \Pi_{\mathcal{A}_0}\nabla\varphi(y_0)\rangle < 0. \quad \forall t \sim 0$$

So, the dual sequence $y_k = \phi_g(x_k)$, defined as in section 3.2, is a descent sequence for the function φ in the Euclidean space $\mathcal{F}^*$ and as in Proposition 3.2.1, the sequences $\{x_k\}$ and $\{y_k\}$ have the same minimizing properties for the function f and φ, respectively. Therefore, the algorithm generates a descent sequence $\{x_k\}$.

Besides, without losing generality, we can assume $\mathcal{A}_0 = \mathbb{R}^{n-m}$, therefore, the algorithm can be seen as a steepest descent method in the space $\mathbb{R}^{m-n}$ and the same convergences properties of section 3 are valid.

Finally, let us note two important points that are exactly than in the non restrictive case, the convexity in the Riemannian sense and the interior convergence. The convexity in the Riemannian sense of a function f on $\mathcal{F}$ also gives the convexity in the Euclidean dual space $\mathcal{F}^*$ of the dual function φ, and the same properties of convergence are valid. Also, it is easy to adapt the arguments of section 3 to this case, the reason is that the key of the proof is the relation between the primal and the dual functions.

Capítulo 5

Legendre square root functions

5.1 Framework

This section is focused on question when the squared Hessian Riemannian metric coincides with a Hessian Riemannian metric induced by another Legendre function. The motivation is to know how much our theory differs from the theory developed before by Alvarez et al in [1].

Definition 5.1.1 (Legendre square root) *Given $h \in \Gamma_0(\mathbb{R}^n)$ satisfying $(H_0;C)$, we say that g is a* Legendre square root *of h whenever g is Legendre,* $\operatorname{int}\operatorname{dom} g = C$*, $g|_C \in \mathcal{C}^2$ and*

$$(LSR) \qquad \forall x \in C,\ \nabla^2 h(x) = (\nabla^2 g(x))^2.$$

In such a case, we write $g \in LSR(h)$.

For instance, for the log-barrier $h(x) = -\log x$ if $x > 0$ and $+\infty$ otherwise, it is direct to verify that the Boltzmann-Shannon entropy, which is given by $g(x) = x\log x - x$ if $x \geq 0$ and $+\infty$ otherwise, satisfies (LSR) with $C = (0, +\infty)$. On the other hand, it is not difficult to verify that the Boltzmann-Shannon entropy do not have a Legendre square root (see Section 5.2 for further details).

Let us remark three simple facts:

- If h has a Legendre square root, then for all $\lambda > 0$, λh has a Legendre square root too.

- It is clear that if g_1 and g_2 are Legendre square root functions of h, then $g_1 - g_2$ is an affine function. This motivates us to define the following equivalence relation: $g_1 \circeq g_2 \quad \Leftrightarrow \quad g_1 - g_2$ is an affine function.
- If h has a Legendre square root and $h \circeq h_0$, then h_0 has a Legendre square root.

Therefore, $LSR(h)$ corresponds to the equivalence class of the Legendre square root functions of h. Hereinafter we denote by g any Legendre square root function of h.

5.2 Specialization to separable functions

From now on, we focus our attention on the case when the functions are separable: if $h : C \subseteq \mathbb{R}^n \to \mathbb{R}$, there are n functions $h_i : C_i \subseteq \mathbb{R} \to \mathbb{R}$ such that $C = C_1 \times ... \times C_n$ and

$$h(x_1, \ldots, x_n) = h_1(x_1) + \ldots + h_n(x_n).$$

We will study the existence of Legendre Legendre-square-root pairs in this case and we will apply the previous analysis in order to get a simple formula for the associated geodesic distance. We will also give some useful examples with their respective formulas such as the geodesics and its Fenchel conjugate.

5.2.1 Existence

The main purpose of this section is to characterize (necessary and sufficient conditions) those Legendre functions h for which there exists a Legendre square root in the sense of Definition 5.1.1. Let us start with a simple fact, the reader can check that there is a unique pair (h, g) whose domain is all $\mathbb{R}$, that is because $h(x) = \frac{1}{2}x^2$ is the unique solution (except by constants) to

$$y'' = [y'']^2 \qquad y'' > 0, \tag{5.1}$$

and therefore, in the separable case, there exists only one pair (h, g) whose domain is all $\mathbb{R}^n$.

The first existence theorem comes.

Theorem 5.2.1 (One-dimensional case) *Let $h \in \Gamma_0(\mathbb{R})$ be a Legendre function with $C = (a_0, b_0)$ with $-\infty \leq a_0 < b_0 \leq +\infty$, but at least one of them is finite. Then $\exists g \in LSR(h)$ if and only if $\sqrt{h''(\cdot)}$ is not integrable around the boundary of C.*

Proof : Assume without loss of generality that $a_0 \in \mathbb{R}$ and let $g \in \Gamma_0(\mathbb{R})$ be a Legendre function that satisfies (LSR) and take $x, x_0 \in C$, then we have

$$g'(x) = \int_{x_0}^{x} \sqrt{h''(s)}ds + g'(x_0), \tag{5.2}$$

then, if $\sqrt{h''(\cdot)}$ is integrable around a_0, since h is smooth enough, we have that it is integrable in any interval $[a_0, x_0]$, and we will have

$$\left| \int_{a_0}^{x_0} \sqrt{h''(s)}ds \right| < +\infty,$$

therefore, fixing $x_0 > a_0$ and letting $x \to a_0^+$ we will have a contradiction because g is essentially smooth.

Suppose conversely that $\sqrt{h''(\cdot)}$ is not integrable around a_0, let us consider the following differential equation: given $x_0 \in C$ and $g_0 \in \mathbb{R}$

$$\varphi'(x) = \int_{x_0}^{x} \sqrt{h''(s)}ds \quad \forall x \in C, \qquad \varphi(x_0) = g_0, \tag{5.3}$$

let $g(\cdot)$ be the unique solution of (5.3), which directly satisfy (LSR). Since $h''(x) > 0$, $\forall x \in C$, we have that $g''(x) = \sqrt{h''(x)} > 0$, $\forall x \in C$, therefore $g(\cdot)$ is strictly convex on C, and since $\sqrt{h''(\cdot)}$ is not integrable around a_0, we have that $|g'(x)| \longrightarrow +\infty$ as $x \to a_0$.

The same argument is applicable to the case $b_0 \in \mathbb{R}$, so the proof is completed. □

In view of the previous theorem we can obtain a simple formula for the Legendre square root, indeed, let us suppose that h has a Legendre square root g, let us fix $x_0 \leq x \leq u$ in C, then integrating (5.2) between x_0 y u with respect to x, we have

$$\begin{aligned} g(u) - g(x_0) &= \int_{x_0}^{u} \int_{x_0}^{x} \sqrt{h''(s)}dsdx + g'(x_0)(u - x_0) \\ &= \int_{x_0}^{u} \int_{s}^{u} \sqrt{h''(s)}dxds + g'(x_0)(u - x_0) \\ &= \int_{x_0}^{u} \sqrt{h''(s)}(u - s)ds + g'(x_0)(u - x_0). \end{aligned}$$

Since $u \mapsto g'(x_0)(u - x_0) + g(x_0)$ is an affine map, we can represent a Legendre square root of h by

$$g(x) = \int_{x_0}^{x} \sqrt{h''(s)}(x - s)ds, \qquad \text{for some } x_0 \in C. \tag{5.4}$$

Corollary 5.2.1 (Separable case) *Let $h \in \Gamma_0(\mathbb{R}^n)$ and $h_i \in \Gamma_0(\mathbb{R})$ be such that $h(x_1,...,x_n) = \sum_{i=1}^n h_i(x_i)$, where each h_i is a Legendre function with $C_i = \operatorname{int}\operatorname{dom} h_i = (a_i,b_i)$ and $-\infty \le a_i < b_i \le +\infty$, but at least one of them is finite for each i. Then $\exists g \in \Gamma_0(\mathbb{R}^n)$, a separable function satisfying (LSR) if only if each $\sqrt{h_i''}$ is not integrable around the boundary of each C_i.*

Proof : Clearly $\operatorname{int}\operatorname{dom} h = C_1 \times ... \times C_n$. First, we prove the existence of the Legendre square root. Since, each h_i satisfies the conditions of Theorem 5.2.1, we have the existence of a Legendre square root g_i, for each h_i, then we define $g(x_1,...,x_n) = \sum_{i=1}^n g_i(x_i)$. It is easy to see that g is a strictly convex function with $\operatorname{int}\operatorname{dom} g = \prod_{i=1}^n C_i$ and that $\nabla^2 g(x_1,...,x_n) = \operatorname{diag}(g''(x_1),...,g''(x_n))$, so the condition $\nabla^2 h = [\nabla^2 g]^2$ follows. Since we are in a finite dimensional context, we have

$$|\nabla g(x_1,...,x_n)| \ge \max_{i=1,...,n} \{|g_i'(x_i)|\} \tag{5.5}$$

if we take $x^k \longrightarrow \bar{x} \in \partial C$. Without losing generality, let us suppose that $x_i^k \longrightarrow a_i^+$ for some $i \in \{1,...,n\}$, therefore $|g_i'(x_i^k)| \longrightarrow +\infty$. Hence g is a Legendre function. To prove the converse, we just have to use the same argument as Theorem 5.2.1 for each $\partial_{x_i} g$. □

5.2.2 Geodesics distances

In this section we give some simple formulas for the geodesic distance in the case when the functions are separable.

Proposition 5.2.1 (One-dimensional case) *The geodesic distance between $x_0, x_1 \in C$ is given by*

$$dist(x_0,x_1) = \left| \int_{x_0}^{x_1} \sqrt{h''(s)} ds \right| \tag{5.6}$$

Proof : Just have to replace (5.2) in the formula of the geodesic distance we have given in section 5. □

The following corollary is a direct consequence of proposition 5.2.1 and gives a formula for the geodesics distance of a curve in direct products as the reader could find in [13, Lemma 4.1].

Corollary 5.2.2 (Separable case) *The geodesic distance between $u, v \in C$ is given by*

$$dist(u,v) = \sqrt{\sum_{i=1}^n dist_i(u_i,v_i)^2} \tag{5.7}$$

where $dist_i(u_i, v_i)$ *denotes the geodesic distance in the submanifold* C_i.

Proof : Thanks to corollary 5.2.1 we have

$$\partial_{x_i} g(u) - \partial_{x_i} g(v) = \int_{u_i}^{v_i} \sqrt{h_i''(s)} ds \tag{5.8}$$

and using Proposition 5.2.1 we are done. □

These results give us a simple formula to calculate the geodesic distance for a geodesic in C without the need knowing explicitly a square, it is enough knowing about its existence.

5.3 Existence of Legendre square root for general functions

In what follows, we turn to the question of the existence in a more general case. Indeed, our goal is to show that the problem is a tough one. In order to do this we will state a necessary condition for the existence in the two-dimensional case and then we will apply this to polyhedral sets.

Theorem 5.3.1 (Necessary condition, two-dimensional case) *Let* $h \in \Gamma_0(\mathbb{R}^2)$ *be a* $\mathcal{C}^3(C)$ *Legendre function. Let us assume* $\exists g \in LSR(h)$ *and define*

$$u(x_1, x_2) = \partial_{x_1 x_2} h(x_1, x_2) \text{ and } v(x_1, x_2) = \partial_{x_1 x_1} h(x_1, x_2) - \partial_{x_2 x_2} h(x_1, x_2)$$

then for all $(x_1, x_2) \in C$ *we have* $v(x_1, x_2) \nabla u(x_1, x_2) = u(x_1, x_2) \nabla v(x_1, x_2)$.

Proof : It is easy to check that h and g satisfies the following relations:

$$\partial_{x_1 x_1} g = \partial_{x_1 x_1} h^2 + \partial_{x_1 x_2} h^2 \tag{5.9}$$
$$\partial_{x_1 x_2} g = \partial_{x_1 x_2} h \, [\partial_{x_1 x_1} h + \partial_{x_2 x_2} h] \tag{5.10}$$
$$\partial_{x_2 x_2} g = \partial_{x_2 x_2} h^2 + \partial_{x_1 x_2} h^2 \tag{5.11}$$

Now, if we take the derivative of (5.9) and (5.10) with respect to x_2 and the derivative of (5.10) and (5.11) with respect to x_1 we get

$$\partial_{x_1 x_1 x_2} g = 2 \, [\partial_{x_1 x_1} h \, \partial_{x_1 x_1 x_2} h + \partial_{x_1 x_2} h \, \partial_{x_1 x_2 x_2} h]$$
$$\partial_{x_1 x_2 x_2} g = \partial_{x_1 x_2 x_2} h \, [\partial_{x_1 x_1} h + \partial_{x_2 x_2} h] + \partial_{x_1 x_2} h \, [\partial_{x_1 x_1 x_2} h + \partial_{x_2 x_2 x_2} h]$$
$$\partial_{x_1 x_1 x_2} g = \partial_{x_1 x_1 x_2} h \, [\partial_{x_1 x_1} h + \partial_{x_2 x_2} h] + \partial_{x_1 x_2} h \, [\partial_{x_1 x_1 x_1} h + \partial_{x_1 x_2 x_2} h]$$
$$\partial_{x_1 x_2 x_2} g = 2 \, [\partial_{x_2 x_2} h \, \partial_{x_1 x_2 x_2} h + \partial_{x_1 x_2} h \, \partial_{x_1 x_1 x_2} h]$$

The latest give step to the following relations

$$2\,[\partial_{x_1x_1}h\,\partial_{x_1x_1x_2}h+\partial_{x_1x_2}h\,\partial_{x_1x_2x_2}h]=\partial_{x_1x_1x_2}h\,[\partial_{x_1x_1}h+\partial_{x_2x_2}h]+\partial_{x_1x_2}h\,[\partial_{x_1x_1x_1}h+\partial_{x_1x_2x_2}h]$$
$$2\,[\partial_{x_2x_2}h\,\partial_{x_1x_2x_2}h+\partial_{x_1x_2}h\,\partial_{x_1x_1x_2}h]=\partial_{x_1x_2x_2}h\,[\partial_{x_1x_1}h+\partial_{x_2x_2}h]+\partial_{x_1x_2}h\,[\partial_{x_1x_1x_2}h+\partial_{x_2x_2x_2}h]$$

Written in another way we obtain

$$\partial_{x_1x_1}h\,\partial_{x_1x_1x_2}h+\partial_{x_1x_2}h\,\partial_{x_1x_2x_2}h=\partial_{x_1x_1x_2}h\,\partial_{x_2x_2}h+\partial_{x_1x_2}h\,\partial_{x_1x_1x_1}h$$
$$\partial_{x_2x_2}h\,\partial_{x_1x_2x_2}h+\partial_{x_1x_2}h\,\partial_{x_1x_1x_2}h=\partial_{x_1x_2x_2}h\,\partial_{x_1x_1}h+\partial_{x_1x_2}h\,\partial_{x_2x_2x_2}h$$

Thus

$$\partial_{x_1x_1x_2}h\,[\partial_{x_1x_1}h-\partial_{x_2x_2}h]=\partial_{x_1x_2}h\,[\,\partial_{x_1x_1x_1}h-\partial_{x_1x_2x_2}h]$$
$$\partial_{x_1x_2x_2}h\,[\partial_{x_1x_1}h-\partial_{x_2x_2}h]=\partial_{x_1x_2}h\,[\,\partial_{x_1x_1x_2}h-\partial_{x_2x_2x_2}h]$$

Finally,

$$\partial_{x_1}(\partial_{x_1x_2}h)\cdot[\partial_{x_1x_1}h-\partial_{x_2x_2}h]=\partial_{x_1x_2}h\cdot\partial_{x_1}[\partial_{x_1x_1}h-\partial_{x_2x_2}h]$$
$$\partial_{x_2}(\partial_{x_1x_2}h)\cdot[\partial_{x_1x_1}h-\partial_{x_2x_2}h]=\partial_{x_1x_2}h\cdot\partial_{x_2}[\partial_{x_1x_1}h-\partial_{x_2x_2}h].$$

□

5.3.1 Polyhedral sets

As we said before, we want to illustrate that the question of finding a non-separable function that has a Legendre square root is a tough one. Suppose that the open convex set C is given by

$$C=\{x\in\mathbb{R}^n:\ \langle a_i,x\rangle<b_i,\ i=1,...,p\} \tag{5.12}$$

By [1, section 4.4], we know that if we have Legendre function $\theta\in\Gamma_0(\mathbb{R})$ such that

$$\begin{cases}\text{(i) } (0,\infty)\subseteq\operatorname{dom}\theta\subseteq[0,\infty),\\ \text{(ii) } \theta\in\mathcal{C}^3(0,\infty)\text{ and }\lim_{s\to0^+}\theta'(s)=-\infty,\\ \text{(iii) } \forall s>0,\ \theta''(s)>0,\end{cases}$$

then $h(x):=\sum_{i=1}^{p}\theta(b_i-\langle a_i,x\rangle)$ is a Legendre function which satisfies all our previous assumptions. We will show that this function has a Legendre square root only in a very especial case when the geometry of C is rectangular. Let us suppose $n=2$ and $a_i=(p_i,q_i)$.

According to Theorem 5.3.1, if we compute the terms $u(x_1,x_2)\cdot\nabla v(x_1,x_2) - u(x_1,x_2)\cdot\nabla v(x_1,x_2)$ we get the following equations for all $(x_1,x_2)\in C$

$$\sum_{i,j=1}^{p} \theta''(b_i - p_i x_1 - q_i x_2)\theta'''(b_j - p_j x_1 - q_j x_2)\underbrace{[p_i^2 q_i(p_j^2 - q_j^2) - p_j q_j(p_i^2 - q_i^2)p_i]}_{c_{ij}} = 0 \quad (5.13)$$

$$\sum_{i,j=1}^{p} \theta''(b_i - p_i x_1 - q_i x_2)\theta'''(b_j - p_j x_1 - q_j x_2)\underbrace{[p_i q_i^2(p_j^2 - q_j^2) - p_j q_j(p_i^2 - q_i^2)q_i]}_{d_{ij}} = 0. \quad (5.14)$$

Additionally, as a consequence (ii) and the mean value theorem, we can assume $\theta''(s)\to-\infty$ as $s\to 0^+$. Moreover, for every $i\in 1,\dots,p$ we can always find a sequence $(x_{1,n},x_{2,n})$ such that

$$\lim_{n\to\infty}\theta''(b_i - p_i x_{1,n} - q_i x_{2,n}) = -\infty \quad (5.15)$$

$$\sup_{n\in\mathbb{N}}|\theta'''(b_j - p_j x_{1,n} - q_j x_{2,n})| < \infty, \quad \forall j\in\{1,\dots,p\}\setminus\{i\}. \quad (5.16)$$

Thus, the equations (5.13) and (5.14) are satisfied if and only if $c_{ij} = d_{ij} = 0$ for every i and j. Let us look with more detail the terms $c_{i,j}$ and $d_{i,j}$

$$\begin{aligned} c_{ij} &= p_i[p_i q_i p_j^2 - p_j q_j p_i^2 + p_j q_j q_i^2 - p_i q_i q_j^2] \\ &= p_i[p_i p_j(p_j q_i - p_i q_j) + q_i q_j(p_j q_i - p_i q_j)] \\ &= p_i(p_i p_j + q_i q_j)(p_j q_i - p_i q_j) \\ &= p_i\cdot\langle a_i, a_j\rangle\ (a_i\times a_j) \end{aligned}$$

Similarly, we get that $d_{ij} = q_i\cdot\langle a_i,a_j\rangle\ (a_i\times a_j)$. Now, to make sense of the set C is straightforward that $a_i\neq 0$ for every $i = 1,\dots,p$. This means that $c_{ij} = 0$ and $d_{ij} = 0$ if and only if the vector $a_1,\dots,a_p$ are perpendicular or parallel between each other. In fact, we can prove the existence of a function with Legendre square root in the case when the vector are perpendicular or parallel between each other and this is a direct consequence of the next proposition, because in this case we know the set C is rotation and translation of a set D whose faces are parallel to the canonical axes, in other words, there exists P a rotation matrix and d a translation vector such that $C = PD + b$

Proposition 5.3.1 *Let P be a rotation matrix and d a translation vector. Suppose also that φ and ψ are a Legendre Legendre-square-root pair with $D = \operatorname{int}\operatorname{dom} h$, then there exists h and g a Legendre-Legendre square root pair with $\operatorname{int}\operatorname{dom} h = PD + b$.*

Proof : Let us consider $h(x) := \varphi(P^t x - P^t b)$, then is clear that h is a Legendre function because $x\mapsto P^t x - P^t b$ is an affine transformation; see [1, Proposition 5.3]. Moreover, it is also clear that $\operatorname{int}\operatorname{dom} h = PD + b$. Then for every $x\in C$ we have

$$\nabla h(x) = P^t\nabla\varphi(P^t x - P^t b) \quad \text{and} \quad \nabla^2 h(x) = P\nabla^2\varphi(P^t x - P^t b)P^t.$$

But $\nabla^2\varphi(y) = [\nabla^2\psi(y)]^2$ for every $y \in D$. Thus

$$\begin{aligned}\nabla^2 h(x) &= P[\nabla^2\psi(P^t x - P^t b)]^2 P^t \\ &= P\nabla^2\psi(P^t x - P^t b)P^t P\nabla^2\psi(P^t x - P^t b)P^t \\ &= [P\nabla^2\psi(P^t x - P^t b)P^t]^2\end{aligned}$$

And, since $g(x) := \psi(P^t x - P^t b)$ is also a Legendre function with $\operatorname{int}\operatorname{dom} g = PD + b$, we conclude the proof. $\square$

Capítulo 6

Examples

6.1 Separable case

In this section we explore some simple but useful examples of the one-dimesional case, also, we use the previous theorem and give some calculations related to each function, in order to this, we denote by $dist_h$ the geodesic distance associated to h and D_h the Bregman distance related to h, that is,

$$D_h(x,y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle \qquad \forall x, y \in C. \tag{6.1}$$

Recall that, since g is a Legendre type function, $\nabla g^{-1} = \nabla g^*$ and thereby the geodesic starting from x_0 with initial velocity $-\nabla^2 g(x_0)^{-1} c$ is given by

$$x(t) = \nabla g^*(\nabla g(x_0) - ct) \qquad \forall t \geq 0$$

We start by showing that not all the barrier functions satisfies the previous criterion, for instances, let h be the Boltzmann-Shannon entropy, that is,

$$h(x) = x \log x - x \quad \text{ with } C = (0, +\infty),$$

we have that $\sqrt{h''(x)} = \dfrac{1}{\sqrt{x}}$, then theorem 5.2.1 tells us that there is no Legendre square root for h, in fact, all the functions that satisfy the condition (LSR) are equivalent to $g(x) = \dfrac{4}{3}x^{3/2}$, which is not a essentially smooth function on $(0, +\infty)$.

Now, we show some examples where the condition is fulfilled, and we give the corresponding square root, the geodesic distance and the Bregman distance associates to each Legendre function and the expression of the geodesics. We also talk about the completeness of the manifold. It is worth noting that the geodesic distances of Example 6.1.2 was previously obtained by Nesterov and Todd in [13], with different tools and a more extensive and complicated proof. Also the geodesic distance of Example 6.1.4 was previously obtained by Papa Quiroz and Oliveira in [16].

Exemple 6.1.1 *Let h be the inverse barrier, that is,*

$$h(x) = \frac{1}{2x} \quad \text{with } C = (0, +\infty),$$

we have that $\sqrt{h''(x)} = \frac{1}{\sqrt{x^3}}$*, then by Theorem 5.2.1, there is a Legendre square root for h since it is not integrable around 0, in fact,* $g(x) = -4\sqrt{x}$*, is a Legendre square root for h. Moreover, , the Fenchel conjugate of g is given by* $g^*(x^*) = -4\frac{1}{x^*}$*. The following resume all this.*

h	C	h^*	g	g^*	$dist_h(x,y)$	$D_h(x,y)$
$\frac{1}{2x}$	$(0,+\infty)$	$\sqrt{-2x^*}$	$-4\sqrt{x}$	$-\frac{4}{x^*}$	$2\left\lvert\frac{\sqrt{x}-\sqrt{y}}{\sqrt{xy}}\right\rvert$	$\frac{(x-y)^2}{2xy^2}$

Table 6.1: Summary table of Example 6.1.1

Therefore, the geodesics of (C, G^2) are given by

$$x(t) = \frac{4x_0}{(2 + ct\sqrt{x_0})^2}, \quad x_0 > 0 \tag{6.2}$$

Let us note that in this case, thank to corollary 2.2.2 we can ensure that (C, G^2) is not a complete manifold because $C^* = (-\infty, 0)$.

Exemple 6.1.2 *Let h be the logarithmic barrier, that is,*

$$h(x) = -\log x \quad \text{with } C = (0, +\infty),$$

we have that $\sqrt{h''(x)} = \frac{1}{x}$*, and again it may not be integrable around 0, then by Theorem 5.2.1, there is a Legendre square root for h , we can compute it and come up with that* $g(x) = x\log x - x$*, and so,* $g^*(x^*) = e^{x^*}$*.*

h	C	h^*	g	g^*	$dist_h(x,y)$	$D_h(x,y)$
$-\log x$	$(0,+\infty)$	$-(1+\log(-x^*))$	$x\log x - x$	e^{x^*}	$\left\lvert\log\left(\frac{x}{y}\right)\right\rvert$	$\frac{x}{y} - \log\left(\frac{x}{y}\right) - 1$

Table 6.2: Summary table of Example 6.1.2

Therefore, the geodesics of (C,G^2) are given by

$$x(t) = x_0 e^{-ct}, \quad x_0 > 0 \tag{6.3}$$

On the contrary of previous example, in this case the manifold (C,G^2) is complete.

Exemple 6.1.3 *Let $h(x) = \frac{1}{6}\sec^2 x - \frac{2}{3}\log\cos x$ with $C = (-\frac{\pi}{2}, \frac{\pi}{2})$, we have that $h'(x) = \frac{1}{3}\sec^2 x \tan x + \frac{2}{3}\tan x$, therefore h is essentially smooth, moreover $h''(x) = \sec^4 x$ which is positive on C, it follows that h is a Legendre function.*

On the other hand, $\sqrt{h''(x)} = \sec^2 x$ which is not integrable around $\{-\frac{\pi}{2}, \frac{\pi}{2}\}$ because $\tan x$ is a primitive for this function in C. Then by Theorem 5.2.1, there is a Legendre square root for h , and so, integrating once again, we have that $g(x) = -\log\cos x$, is a Legendre square root for h. Also, it is easy verify that $g^*(x^*) = x^* \arctan x^* - \frac{1}{2}\log(1+(x^*)^2)$.

h	C	g	g^*	$dist_h(x,y)$
$\frac{1}{6}\sec^2 x - \frac{2}{3}\log\cos t$	$(-\frac{\pi}{2}, \frac{\pi}{2})$	$-\log\cos x$	$x^* \arctan x^* - \frac{1}{2}\log(1+(x^*)^2)$	$\lvert\tan x - \tan y\rvert$

Table 6.3: Summary table of Example 6.1.3

Therefore, the geodesics of (C,G^2) are given by

$$x(t) = \arctan(\tan x_0 - ct), \quad x_0 \in (-\frac{\pi}{2}, \frac{\pi}{2}) \tag{6.4}$$

Let us note that in this case, thanks to Corollary 2.2.2 we can ensure that (C,G^2) is a complete manifold because $C^* = \mathbb{R}$.

Exemple 6.1.4 *Let $h(x) = (2x-1)[\log x - log(1-x)]$ with $C = (0,1)$, a straightforward computation give us that $\sqrt{h''(x)} = \frac{1}{x(1-x)}$, therefore, by Theorem 5.2.1, there is a Legendre square root for h , and it is given by $g(x) = -x\log x - (1-x)\log(1-x)$. The reader can verify that the Fenchel conjugate of g is given by $g^*(x^*) = \log(1+e^{x^*})$.*

h	C	g	g^*	$dist_h(x,y)$
$(2x-1)[\log x - \log(1-x)]$	$(0,1)$	$x\log x + (1-x)\log(1-x)$	$\log(1+e^{x^*})$	$\left\lvert \log\left(\frac{x(1-y)}{y(1-x)}\right)\right\rvert$

Table 6.4: Summary table of Example 6.1.4

Therefore, the geodesics of (C,G^2) are given by

$$x(t) = \frac{\exp\left(\log\left(\frac{x_0}{1-x_0}\right) - ct\right)}{1+\exp\left(\log\left(\frac{x_0}{1-x_0}\right) - ct\right)}, \quad x_0 \in (0,1) \tag{6.5}$$

As in previous example, the manifold (C,G^2) is complete, that is because $C^* = \mathbb{R}$.

6.2 Radial case

We finish this article by showing an example of a problem where $C \subseteq \mathbb{R}^n$ is the unitary ball. Let us recall that if $F \in \Gamma_0(\mathbb{R})$ is even and increasing on $\operatorname{dom} F \cap \mathbb{R}^+$ then if we set $g(x) = F(|x|)$ then $g \in \Gamma_0(\mathbb{R}^n)$ and $g^*(y) = F^*(|y|)$.

Let us set $F(r) = -\frac{1}{2}\log(1-r^2)$, then $g(x) = -\frac{1}{2}\log(1-|x|^2)$ and therefore $\nabla g(x) = \frac{|x|}{1-|x|^2}x$. It is easy to see that g satisfies $(H_0;C)$. Moreover, by the above, $g^*(0) = 0$ and g^* is given by the map:

$$y \longmapsto \frac{1}{2}\left(\sqrt{1+4|y|^2} - \log(|y|) + arctanh\left(\frac{1}{\sqrt{1+4|y|^2}}\right)\right),$$

and so $\nabla g^*(y) = \frac{\sqrt{1+4|y|^2}-1}{2|y|}y$. Thus, by Proposition 2.2.1 the geodesic of (C,G^2) are given by

$$x(t) = \frac{\sqrt{1+4\left|\frac{|x_0|}{1-|x_0|^2}x_0 - td_0\right|^2} - 1}{2\left|\frac{|x_0|}{1-|x_0|^2}x_0 - td_0\right|}\left[\frac{|x_0|}{1-|x_0|^2}x_0 - td_0\right], \quad x_0 \in B(0,1) \tag{6.6}$$

Besides, the manifold (C,G^2) is complete because $C^* = \mathbb{R}^n$. Moreover, thanks to Theorem 5.3.1 and it generalization to higher dimensions, it is easy to see that this barrier does not have a Legendre square root.

Bibliografía

[1] F. Álvarez, J. Bolte, and O. Brahic. Hessian riemannian gradient flows in convex programming. *SIAM J. Control and Optimization*, 43(2):477–501, 2004.

[2] A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM Journal on Optimization*, 16(3):697–725, 2006.

[3] H. Bauschke, J. Borwein, and P. Combettes. Essential smoothness, essential strict convexity, and legendre fucntions in banach spaces. *Communications in Contemporary Mathematics*, 3(4):615–647, 2001.

[4] D. A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming. I. Affine and projective scaling trajectories. *Trans. Amer. Math. Soc.*, 314(2):499–526, 1989.

[5] R. Burachik, L. M. G. Drummond, A. N. Iusem, and B. F. Svaiter. Full convergence of the steepest descent method with inexact line searches. *Optimization*, 32(2):137–146, 1995.

[6] M. P. do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, 1976.

[7] M. P. do Carmo. *Riemannian Geometry*. Birkhäuser Boston, 1992.

[8] A. V. Fiacco. Perturbed variations of penalty function methods. Example: projective SUMT. *Ann. Oper. Res.*, 27(1-4):371–380, 1990.

[9] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer Berlin Heidelberg New York, 2004.

[10] J. C. Gilbert. éléments d'optimisation différentiable, théorie et algorithmes. Syllabus de cours à l'ENSTA, Paris. `http://www-rocq.inria.fr/~gilbert/ensta/optim.html`, 2008.

[11] A. N. Iusem, B. F. Svaiter, and J. X. Da Cruz Neto. Central paths, generalized proximal point methods, and Cauchy trajectories in Riemannian manifolds. *SIAM J. Control Optim.*, 37(2):566–588 (electronic), 1999.

[12] G. P. McCormick. The projective SUMT method for convex programming. *Math. Oper. Res.*, 14(2):203–223, 1989.

[13] Y. Nesterov and M. Todd. On the riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math*, 2(4):333–361, 2002.

[14] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 2000.

[15] E. P. Quiroz and P. Oliveira. New results on linear optimization through diagonal metrics and riemannian geometry tools. Technical report es-645/04, Federal University of Rio de Janeiro, 2004.

[16] E. P. Quiroz and P. Oliveira. New self-concordant barrier for the hypercube. *Journal of Optimization Theory and Applications*, 135(3):475–490, 2007.

[17] T. Rapcsák. *Smooth Nonlinear Optimization in* $\mathbb{R}^n$. Kluwer Academic Publishers, 1997.

[18] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*. Society for Industrial and Applied Mathematics, 2001.

[19] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[20] C. Udriste. *Convex Function and Optimization Methods on Riemannian Manifolds*. Kluwer Academic Publishers, 1994.