



UNIVERSIDAD NACIONAL DE LA PLATA

Facultad de Ciencias Exactas

Departamento de Matemática

**Teoría y métodos para problemas de optimización
multiobjetivo**

Trabajo de Tesis Doctoral

NADIA SOLEDAD FAZZIO

Directora: María Laura Schuverdt

Codirectora: Nélica Echebest

Año 2018

Índice general

1. Introducción	V
2. Preliminares	1
2.1. Problema escalar de programación no lineal	1
2.2. Problema de optimización multiobjetivo	4
2.2.1. Pareto optimalidad	5
2.2.2. Condiciones necesarias de Fritz-John y clases de regularidad asociadas al problema de optimización multiobjetivo	8
2.2.3. Escalarización	9
2.3. Condiciones calificadoras	10
2.3.1. Definiciones de condiciones calificadoras utilizando el concepto de cono	12
2.3.2. Condiciones calificadoras y diferenciabilidad	13
3. Convergencia del método Lagrangiano Aumentado usando Quasi- normalidad y aplicación al problema multiobjetivo	19
3.1. Algoritmo y convergencia global	20
3.2. Condición <i>KKT</i> Aproximada para el <i>MOP</i>	26
3.3. Aplicación del método Lagrangiano Aumentado para el <i>MOP</i>	29
4. Método del gradiente proyectado no monótono para el problema	

multiobjetivo en un conjunto cerrado y convexo	35
4.1. Método del gradiente proyectado no monótono para el <i>MOP</i>	37
4.1.1. Búsqueda lineal no monótona	39
4.1.2. Algoritmo del gradiente proyectado multiobjetivo	40
4.2. Análisis de convergencia	41
4.3. Caso convexo	48
5. El método Lagrangiano Aumentado	55
5.1. Problema escalar con restricciones de cotas en las variables en el nivel inferior	56
5.2. Problema escalar con restricciones que satisfacen la condición de Abadie en el nivel inferior	62
5.3. Problema multiobjetivo con restricciones que satisfacen la condición de Abadie en el nivel inferior	68
6. Conclusiones	77
Bibliografía	79

Capítulo 1

Introducción

En las ciencias como en la ingeniería se dan, en muchas ocasiones, problemas que requieren encontrar el óptimo simultáneo de dos o más funciones objetivo. Problemas con estas características se denominan problemas de optimización multiobjetivo (*MOP*) o de multicriterio. El problema está en que normalmente no existe un elemento del conjunto factible que produzca un óptimo de forma simultánea para cada uno de los r objetivos que componen la función objetivo vectorial. Esto se debe a la existencia de conflictos entre los objetivos, que hace que la mejora de uno de ellos pueda dar lugar a un empeoramiento de algún otro. Pensemos por ejemplo, en el caso de un avión con dos objetivos que fueran su velocidad y el ahorro de combustible: un aumento de la velocidad traerá consigo una disminución del ahorro. Habrá que llegar por tanto a una situación de compromiso en la que todos los objetivos sean satisfechos en un grado aceptable, desde el punto de vista de diseño. Otro ejemplo puede verse en la fabricación de un cierto producto para el cual se requiera minimizar el costo de su producción, reducir la cantidad de horas extras y reducir al mínimo la merma de fabricación de dicho producto. Este ejemplo evidencia que el problema posee más de un objetivo.

Problemas multiobjetivo se pueden encontrar en áreas aplicadas como puede ser Ingeniería [30], Estadística [25], Finanzas [31], Análisis de Medio ambiente [47], entre otros.

A diferencia de los problemas de optimización con un único objetivo llamados problemas escalares, el concepto de óptimo es ahora relativo y será necesario decidir de alguna forma cuál es la mejor solución (o cuáles son las mejores soluciones) al problema presente.

La idea es considerar que ninguna de las componentes del vector que representa las funciones objetivo puede ser mejorada, sin deteriorar alguna otra componente. El concepto que se utiliza es el de punto eficiente o Pareto optimal, y es un concepto

introducido por Pareto [54] en Economía que tiene innumerables aplicaciones en Ingeniería, Neurociencia y Ciencias Sociales, entre otros.

La búsqueda de soluciones Pareto óptimas es un proceso caro, que consume mucho tiempo, porque generalmente hay un número muy grande (o a veces infinito) de éstas soluciones.

En los últimos años se han extendido conceptos y resultados de programación escalar no lineal al caso de problemas multiobjetivo. Las condiciones necesarias de optimalidad son tópicos de suma importancia en optimización, no sólo son útiles para identificar soluciones, sino que también son cruciales en el diseño de métodos para aproximar soluciones.

Sin embargo, para que las condiciones de optimalidad sean útiles, los problemas deben cumplir ciertas hipótesis. Estas hipótesis son llamadas condiciones de calidad o calificadoras cuando deben ser verificadas por las restricciones del problema y son llamadas condiciones de regularidad cuando deben ser verificadas por las funciones objetivo y por las restricciones del problema. Estas condiciones son utilizadas para asegurar que las soluciones del problema escalar cumplen las condiciones de Karush-Kuhn-Tucker [12]. En el caso multiobjetivo, a diferencia del caso escalar, existen multiplicadores para las diferentes funciones objetivo. La positividad de un multiplicador correspondiente a una función objetivo indica un rol activo de esa función en la identificación de las soluciones del problema. Considerando diferentes posibilidades en el signo de los multiplicadores asociados a las funciones objetivo se obtienen diferentes condiciones de regularidad [15]. Por ejemplo, cuando al menos un multiplicador es positivo se dice que la solución es regular débil o débilmente Kuhn-Tucker y esta situación será analizada en este trabajo.

Existe una gran variedad de métodos para resolver problemas multiobjetivo. Cada método posee características y aplicaciones diferentes. La elección de un método debe ser realizada con cuidado ya que no existe un método mejor que todos los otros. Cada problema posee sus características propias, por eso, un método puede ser muy bueno para un tipo de problema e ineficiente para otros. Ninguno de ellos puede decirse que sea superior a todos los otros.

Desde el punto de vista práctico, en los últimos años, procedimientos de optimización clásicos en el caso escalar han sido extendidos para resolver problemas multiobjetivos con buenos resultados teóricos. Por ejemplo, en el trabajo [42] se extiende el método de máximo descenso, en los trabajos [34, 35, 41] podemos encontrar varias versiones del método del gradiente proyectado, en [22] se definen métodos del tipo punto proximal, en [24] presentan un método basado en región de confianza, en [33] proponen un método de Newton para el caso multiobjetivo, en [37] podemos encontrar un algoritmo de penalidad vectorial para resolver el problema en forma exacta y en [36] puede encontrarse una completa revisión sobre métodos de descenso. En todos los trabajos se demuestra convergencia global de la sucesión que generan

los algoritmos utilizando hipótesis razonables.

En esta tesis estudiamos la posibilidad de extender el método Lagrangiano Aumentado clásico de optimización escalar, para resolver problemas con objetivos múltiples. El método Lagrangiano Aumentado es una técnica popular para resolver problemas de optimización con restricciones [2, 19, 20, 21]. Consideramos dos posibles extensiones:

- mediante el uso de escalarizaciones. Basados en el trabajo [37] consideramos el uso de funciones débilmente crecientes para analizar la convergencia global de un método Lagrangiano Aumentado para resolver el problema multiobjetivo con restricciones de igualdad y de desigualdad.
- mediante el uso de una función Lagrangiana Aumentada vectorial. En este caso el subproblema en el método Lagrangiano Aumentado tiene la particularidad de ser vectorial y planteamos su resolución mediante el uso de un método del tipo gradiente proyectado no monótono.

En las extensiones que presentamos en la tesis se analizan las hipótesis más débiles bajo las cuales es posible demostrar convergencia a un punto estacionario del problema multiobjetivo. El trabajo está organizado de la siguiente manera. En el Capítulo 2, definimos el problema de optimización no lineal escalar, el concepto de mínimo y de punto estacionario. Además definimos el problema de optimización multiobjetivo, el concepto de Pareto optimalidad y de punto estacionario. Definimos algunas condiciones de calidad que vamos a utilizar a lo largo de este trabajo y las relaciones entre estas.

En el Capítulo 3, proponemos una aplicación del método Lagrangiano Aumentado del caso escalar para problemas multiobjetivo con restricciones. El objetivo principal de este capítulo, es analizar la convergencia global del método utilizando la condición de calidad Quasinormalidad. Específicamente, se demuestra que los puntos límite que satisfacen Quasinormalidad o la Propiedad del Cono Continuo (*CCP*) son puntos regulares débiles del problema multiobjetivo.

En el Capítulo 4 consideramos una extensión del método del gradiente proyectado (*PGM*) a problemas de optimización multiobjetivo en un conjunto cerrado y convexo. El método del gradiente proyectado para la optimización multiobjetivo propuesto en [34, 35, 36, 41] se amplía para incluir una búsqueda lineal no monótona basada en el promedio de los valores funcionales anteriores en lugar de las reglas clásicas del tipo Armijo. Bajo hipótesis apropiadas se obtiene que los puntos de acumulación de la sucesión generada por el algoritmo propuesto son estacionarios. Además, utilizando hipótesis de convexidad, demostramos convergencia total a soluciones Pareto óptimas débiles de cualquier sucesión generada por el algoritmo no monótono propuesto.

En el Capítulo 5 se define una función Lagrangiana Aumentada vectorial y se define un método del tipo Lagrangiano Aumentado para resolver problemas multiobjetivo con dos tipos de restricciones: restricciones del nivel inferior que no son penalizadas y restricciones del nivel superior que son penalizadas. El algoritmo propuesto puede utilizar el algoritmo *PGM* del Capítulo 4 en la resolución del subproblema. La principal diferencia de la función Lagrangiana Aumentada de este capítulo con la del Capítulo 3 es que en el Capítulo 3 la función Lagrangiana Aumentada es escalar definida mediante una escalarización adecuada mientras que en este capítulo es vectorial. Se demuestra convergencia global del algoritmo propuesto utilizando la condición de Quasinormalidad para el conjunto factible junto con la condición de Abadie para el conjunto de restricciones del nivel inferior.

Capítulo 2

Preliminares

2.1. Problema escalar de programación no lineal

En este capítulo vamos a revisar algunos conceptos que usaremos a lo largo de esta tesis.

Presentamos el problema general de programación no lineal (*PNL*):

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeto a } x \in \Omega = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\} \end{aligned} \quad (2.1.1)$$

donde consideramos a las funciones $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ y $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ con derivadas primeras continuas en un conjunto abierto que contiene a Ω .

La función f se denomina función objetivo y el conjunto Ω se llama conjunto factible. Para $x^* \in \Omega$, denotaremos $A(x^*) = \{j \in \{1, \dots, p\} : g_j(x^*) = 0\}$ al conjunto de índices de las restricciones de desigualdad activas en x^* .

Una solución del problema (2.1.1), llamado minimizador global, es un punto $x^* \in \Omega$ tal que $f(x^*) \leq f(x)$ para todo $x \in \Omega$. Sin embargo, encontrar soluciones globales a estos problemas es en muchos casos muy difícil y hasta a veces imposible. La mayoría de los algoritmos populares se conforman con encontrar candidatos a minimizadores, puntos que satisfacen las condiciones de Karush-Kuhn-Tucker.

Definición 2.1.2. [12] Decimos que un punto $x^* \in \Omega$ cumple las condiciones de Karush-Kuhn-Tucker (*KKT*) para el problema (2.1.1) si existen vectores $\lambda^* \in \mathbb{R}^m$ y $\mu^* \in \mathbb{R}^p$ tales que

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla g_j(x^*) &= 0 \\ \mu_j^* g_j(x^*) &= 0, \quad \mu_j^* \geq 0 \quad j = 1, \dots, p \end{aligned} \quad (2.1.3)$$

Los escalares λ_i^* , μ_j^* con $i = 1, \dots, m$ y $j = 1, \dots, p$ se denominan multiplicadores de Lagrange.

Las condiciones necesarias de optimalidad son uno de los tópicos claves en optimización, ya que permiten identificar soluciones. Sin embargo, para que sean óptimas, los problemas deben cumplir ciertas hipótesis, entre ellas la satisfacción de condiciones calificadoras y/o condiciones de regularidad.

Para el problema *PNL* (2.1.1), las condiciones calificadoras son propiedades de los puntos factibles que, cuando se satisfacen en un minimizador, aseguran el cumplimiento de las condiciones de Karush-Kuhn-Tucker en ese punto [12].

Entre las condiciones calificadoras o de calidad más conocidas y que consideraremos en este trabajo podemos mencionar Abadie [1], Quasinormalidad [46], Mangasarian Fromovitz [49], la Condición de Dependencia Lineal Positiva Constante (*CPLD*), introducida por Qi y Wei [55] y posteriormente analizada por Andreani, Martínez y Schuverdt [8] y la Propiedad del Cono Continuo (*CCP*) [6].

Estas condiciones permiten asegurar que si x^* es un minimizador del problema, entonces x^* cumple las condiciones *KKT* de la Definición 2.1.2. Es decir, x^* es un punto estacionario de la función de Lagrange asociada al problema *PNL*, donde la función de Lagrange para el problema (2.1.1) está definida de la siguiente manera:

$$l(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x).$$

Luego, cuando se satisface una condición calificadora, es posible pensar en términos de los multiplicadores de Lagrange asociados y, consecuentemente, diseñar algoritmos eficientes basados en ideas de dualidad. Sin una condición calificadora, se podrían obtener sólo condiciones de tipo Fritz-John, [12] y no del tipo Karush-Kuhn-Tucker para el *PNL*.

Definición 2.1.4. [12] Decimos que un punto $x^* \in \Omega$ cumple las condiciones de Fritz-John para el problema (2.1.1), si existe un escalar $\mu_0^* \geq 0$ y multiplicadores $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}^p$ tales que

$$\begin{aligned} \mu_0^* \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla g_j(x^*) &= 0 \\ \mu_j^* g_j(x^*) &= 0, \quad \mu_j^* \geq 0 \quad j = 1, \dots, p \\ \mu_0^*, \lambda_1^*, \dots, \lambda_m^*, \mu_1^*, \dots, \mu_p^* &\text{ no todos nulos.} \end{aligned} \quad (2.1.5)$$

En muchas situaciones prácticas los problemas *PNL* del tipo (2.1.1) poseen además restricciones de cotas en las variables o cajas. Debido a la particularidad que poseen las restricciones de cajas, los problemas son tratados independientemente de los problemas *PNL* con restricciones no lineales generales.

Consideremos el siguiente problema

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeto a } x \in \Omega = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\} \\ & \quad x \in X = \{x \in \mathbb{R}^n : l \leq x \leq u\} \end{aligned} \quad (2.1.6)$$

donde f , h y g son funciones como en (2.1.1) y $l, u \in \mathbb{R}^n$, $-\infty < l_i < u_i < \infty$ para todo $i = 1, \dots, n$.

El siguiente lema establece condiciones necesarias y suficientes para las condiciones *KKT* de un problema de minimizar una función f en el conjunto X .

Lema 2.1.7. *Un punto $x^* \in X = \{x \in \mathbb{R}^n : l \leq x \leq u\}$ es *KKT* del problema $\min_{x \in X} f(x)$ si y solo si*

$$P_X(x^* - \nabla f(x^*)) - x^* = 0,$$

donde $P_X(v)$ denota la proyección de v sobre el conjunto X .

Demostración.

\implies) Supongamos que x^* es un punto *KKT* de $\min_{x \in X} f(x)$ entonces existen $\nu \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$, $\nu \geq 0$, $\mu \geq 0$, con $(x_i^* - u_i)\nu_i = 0$, $(l_i - x_i^*)\mu_i = 0$, para todo $i = 1, \dots, n$ tales que

$$\nabla f(x^*) + \sum_{i=1}^n \mu_i (-e_i) + \sum_{i=1}^n \nu_i e_i = 0. \quad (2.1.8)$$

Sea $y^* = P_X(x^* - \nabla f(x^*))$. Veamos que $y^* = x^*$. Por definición de proyección, y^* es solución del problema convexo

$$\min_{y \in X} \frac{1}{2} \|y - (x^* - \nabla f(x^*))\|^2.$$

Pero por (2.1.8), x^* satisface las condiciones *KKT* para el problema anterior y como tiene solución única debe ser $y^* = x^*$.

\impliedby) Ahora supongamos que

$$x^* = P_X(x^* - \nabla f(x^*)).$$

Luego, x^* es solución del problema convexo

$$\min_{y \in X} \frac{1}{2} \|y - (x^* - \nabla f(x^*))\|^2.$$

Entonces existen vectores $\nu \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$, $\nu \geq 0$, $\mu \geq 0$, con $(x_i^* - u_i)\nu_i = 0$, $(l_i - x_i^*)\mu_i = 0$, para todo $i = 1, \dots, n$ tales que

$$\nabla f(x^*) + \sum_{i=1}^n \mu_i(-e_i) + \sum_{i=1}^n \nu_i e_i = 0.$$

Por lo tanto x^* es un punto *KKT* del problema $\min_{x \in X} f(x)$.

□

Podemos observar que $P_X(x^* - \nabla f(x^*)) - x^* = 0$ si y solo si

$$\begin{cases} \frac{\partial f(x^*)}{\partial x_i} \geq 0 & \text{si } x_i^* = l_i, \text{ (Porque } x_i^* - \frac{\partial f(x^*)}{\partial x_i} \leq l_i) \\ \frac{\partial f(x^*)}{\partial x_i} = 0 & \text{si } l_i < x_i^* < u_i, \text{ (Porque } x_i^* - \frac{\partial f(x^*)}{\partial x_i} = x_i^*) \\ \frac{\partial f(x^*)}{\partial x_i} \leq 0 & \text{si } x_i^* = u_i, \text{ (Porque } x_i^* - \frac{\partial f(x^*)}{\partial x_i} \geq u_i). \end{cases} \quad (2.1.9)$$

Luego, $P_X(x^* - \nabla f(x^*)) - x^* = 0$ si y solo si $\nabla f(x^*) + \sum_{i=1}^n \mu_i(-e_i) + \sum_{i=1}^n \nu_i e_i = 0$

ya que, por (2.1.9):

si $l_i < x_i^* < u_i$, entonces $\frac{\partial f(x^*)}{\partial x_i} = 0$ y $\mu_i = 0$, $\nu_i = 0$;

si $x_i^* = l_i$, $\frac{\partial f(x^*)}{\partial x_i} - \mu_i = 0$, entonces $\frac{\partial f(x^*)}{\partial x_i} = \mu_i > 0$;

si $x_i^* = u_i$, $\frac{\partial f(x^*)}{\partial x_i} + \nu_i = 0$, entonces $\frac{\partial f(x^*)}{\partial x_i} = -\nu_i < 0$.

Observación 2.1.10. $P_X(x^* - \nabla f(x^*)) - x^* = 0$ si y solo si $\nabla f(x^*) + \sum_{i=1}^n \mu_i(-e_i) +$

$\sum_{i=1}^n \nu_i e_i = 0$, con $\mu_i = \frac{\partial f(x^*)}{\partial x_i}$, si $x_i^* = l_i$, y $\mu_i = 0$ en caso contrario; y $\nu_i = -\frac{\partial f(x^*)}{\partial x_i}$ si $x_i^* = u_i$ y $\nu_i = 0$ en caso contrario.

2.2. Problema de optimización multiobjetivo

El problema de optimización multiobjetivo, al que nos referiremos por sus siglas en inglés, *MOP*, y que vamos a considerar en este trabajo, está limitado a espacios Euclidianos de dimensión finita. Su forma general es la siguiente

$$\begin{aligned} & \text{Minimizar } F(x) = (F_1(x), F_2(x), \dots, F_r(x))^T \\ & \text{sujeto a } x \in \Omega = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\} \end{aligned} \quad (2.2.1)$$

donde consideramos las funciones $F_l : \mathbb{R}^n \rightarrow \mathbb{R}$, $l = 1, \dots, r$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ y $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ con derivadas primeras continuas en un conjunto abierto que contiene a Ω .

El conjunto $Y = F(\Omega) = \{F(x) : x \in \Omega\}$ se denomina espacio objetivo.

En el caso escalar el significado de optimalidad es claro, por tratarse de la minimización de una función objetivo bajo las restricciones impuestas. En el caso del problema multiobjetivo, el significado no se mantiene. Si existe una solución factible que minimiza todos los objetivos simultáneamente, obviamente será la solución óptima. Sin embargo, raramente podemos esperar que ocurra esta situación, ya que por lo general los objetivos están en conflicto unos con otros. El concepto de optimalidad que vamos a considerar es el de solución Pareto optimal [54].

2.2.1. Pareto optimalidad

Definición 2.2.2. Un punto $x^* \in \Omega$ es Pareto optimal (o eficiente) si no existe otro punto $x \in \Omega$ tal que $F_l(x) \leq F_l(x^*)$ para todo $l = 1, \dots, r$ con al menos $F_l(x) < F_l(x^*)$ para algún $l = 1, \dots, r$.

Debido a que es frecuente que haya infinitas soluciones Pareto optimales, se denomina conjunto Pareto optimal o frontera Pareto al conjunto de tales soluciones.

Definición 2.2.3. Un punto $x^* \in \Omega$ es Pareto optimal débil si no existe otro punto $x \in \Omega$ tal que $F_l(x) < F_l(x^*)$ para todo $l = 1, \dots, r$.

Veamos un ejemplo de puntos Pareto optimal para el problema multiobjetivo sin restricciones ($\Omega = \mathbb{R}^n$).

Ejemplo 2.2.4. Sean $F_1(x) = x^2$ y $F_2(x) = (x - 2)^2$.

El problema multiobjetivo sin restricciones es el siguiente

$$\text{Minimizar } (F_1(x), F_2(x))^T$$

y en la Figura 2.1 se observa la gráfica de ambas funciones.

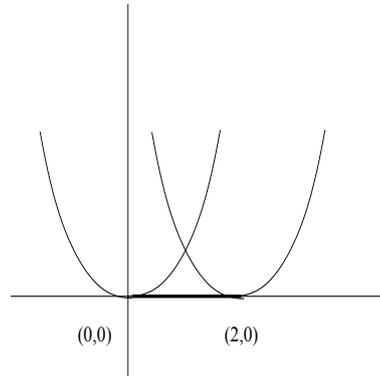


Figura 2.1: Gráfica de las funciones F_1 y F_2 .

Por ejemplo, el punto $x = 0$ es Pareto optimal, pues no existe otro x distinto de $x = 0$ tal que $F_1(x) \leq F_1(0) = 0$ y $F_2(x) \leq F_2(0) = 4$.

De manera análoga, el punto $x = 1$ es Pareto optimal, pues no existe otro x distinto de $x = 1$ tal que $F_1(x) \leq F_1(1) = 1$ y $F_2(x) \leq F_2(1) = 1$: si queremos que $F_1(x) \leq 1$ entonces debería ser $x \in [-1, 1]$ y $F_2(x) \leq 1$ entonces debería ser $x \in [1, 3]$. Los puntos Pareto optimales son los $x \in \mathbb{R}$ tales que están sobre la línea en oscuro en la Figura 2.1, esto se corresponde con $x \in [0, 2]$.

Veamos otro ejemplo de puntos Pareto optimales y Pareto optimales débiles para un problema lineal con restricciones.

Ejemplo 2.2.5. Sean $F_1(x_1, x_2) = 5x_1 - 2x_2$ y $F_2(x_1, x_2) = -x_1 + 4x_2$.

El problema multiobjetivo,

$$\begin{aligned}
 & \text{Minimizar} && (F_1(x_1, x_2), F_2(x_1, x_2))^T \\
 & \text{sujeto a} && -x_1 + x_2 \leq 3 \\
 & && x_1 + x_2 \leq 8 \\
 & && x_2 \leq 4 \\
 & && x_1 \leq 6 \\
 & && x_1 \geq 0, x_2 \geq 0.
 \end{aligned} \tag{2.2.6}$$

En la Figura 2.2 se observa el gráfico de la región factible.

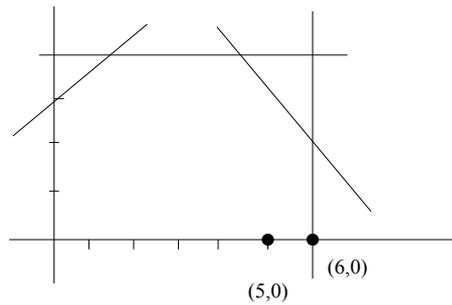


Figura 2.2: Región factible del ejemplo 2.2.5

Consideremos $\tilde{x} = (6, 0)$, su imagen es $(F_1(\tilde{x}), F_2(\tilde{x})) = (30, -6)$, \tilde{x} es un punto Pareto optimal débil ya que no existe x factible distinto de \tilde{x} tal que $F_1(x) < F_1(\tilde{x}) = 30$ y $F_2(x) < F_2(\tilde{x}) = -6$ a la vez.

Si consideramos $\tilde{\tilde{x}} = (5, 0)$, su imagen es $(F_1(\tilde{\tilde{x}}), F_2(\tilde{\tilde{x}})) = (25, -5)$, $\tilde{\tilde{x}}$ no es un punto Pareto optimal débil ya que existe $\tilde{x} = (6, 0)$ para el cual $-6 = F_2(\tilde{x}) < F_2(\tilde{\tilde{x}}) = -5$. Con un razonamiento similar al anterior se puede demostrar que \tilde{x} es el único punto Pareto débil para este problema.

A continuación definiremos las versiones locales de los puntos Pareto optimales.

Definición 2.2.7. Un punto $x^* \in \Omega$ se denomina localmente Pareto optimal si existe un entorno $B(x^*, \epsilon)$ tal que no existe otro vector $x \in \Omega \cap B(x^*, \epsilon)$ tal que $F_l(x) \leq F_l(x^*)$ para todo $l = 1, \dots, r$, con al menos $F_l(x) < F_l(x^*)$ para algún $l = 1, \dots, r$.

Definición 2.2.8. Un punto $x^* \in \Omega$ se denomina localmente Pareto optimal débil si existe un entorno $B(x^*, \epsilon)$ tal que no existe otro vector $x \in \Omega \cap B(x^*, \epsilon)$ tal que $F_l(x) < F_l(x^*)$ para todo $l = 1, \dots, r$.

A modo de ejemplo podemos ver en las Figuras 2.3 y 2.4 algunos de estos puntos.

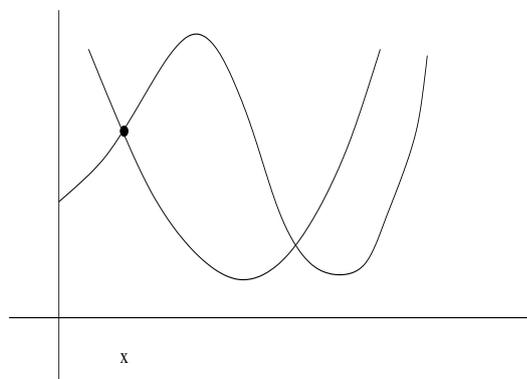


Figura 2.3: Ejemplo de un punto localmente Pareto optimal.

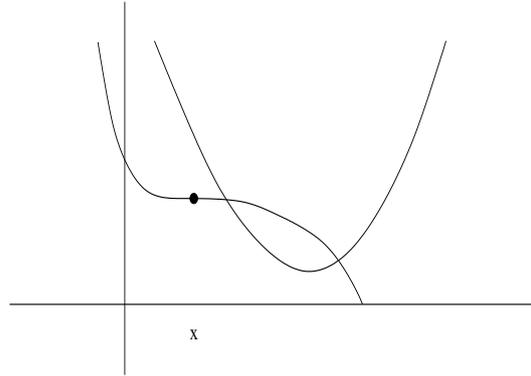


Figura 2.4: Ejemplo de un punto localmente Pareto óptimo débil.

2.2.2. Condiciones necesarias de Fritz-John y clases de regularidad asociadas al problema de optimización multiobjetivo

En esta sección presentamos una generalización de la bien conocida regla de multiplicadores de Fritz-John publicada en 1948. Originalmente, Fritz-John formuló su regla para la optimización de una función a valores reales sujeta a restricciones de igualdad.

La siguiente definición es una extensión de las condiciones (2.1.5) para el problema multiobjetivo.

Definición 2.2.9. [26] Consideremos el problema multiobjetivo (2.2.1) y supongamos que las funciones involucradas son continuamente diferenciables en $x^* \in \Omega$. Un vector no nulo (θ, λ, μ) , $\theta \in \mathbb{R}^r$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ con $\theta \geq 0$, $\mu \geq 0$ que satisface

$$\sum_{l=1}^r \theta_l \nabla F_l(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0, \quad (2.2.10)$$

$$\mu_j g_j(x^*) = 0, \quad j = 1, \dots, p, \quad (2.2.11)$$

$$(\theta, \lambda, \mu) \text{ no todos nulos} \quad (2.2.12)$$

se denomina vector de multiplicadores de Fritz-John. Denotamos con $M(x^*)$ al conjunto de tales vectores (θ, λ, μ) asociados a x^* .

Observación 2.2.13. Si construimos la función de Fritz-John asociada al MOP

$$F(x)_{MOP} = \sum_{l=1}^r \theta_l F_l(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x)$$

podemos interpretar que la condición (2.2.10) equivale a decir que x^* es un punto estacionario de $F(x)_{MOP}$.

Desafortunadamente, las condiciones de Fritz-John no garantizan por sí mismas que los multiplicadores θ_i asociados a las funciones objetivo no sean todos nulos. Esto sucede también en el caso escalar. Cuando el vector de multiplicadores θ es nulo, las condiciones de Fritz-John no hacen uso de la información que proveen los gradientes de las funciones objetivo. Sin imponer condiciones adicionales, uno no puede siquiera asegurar al menos un $\theta_l > 0$. Para obtener un resultado más fuerte es necesario tener alguna hipótesis de regularidad sobre el problema. Siguiendo la clasificación realizada por Bigi y Pappalardo [15], establecemos las nociones de regularidad que usaremos en este trabajo. Estos conceptos están relacionados a los puntos factibles para los cuales $M(x) \neq \emptyset$.

Definición 2.2.14. Dado $x \in \Omega$ tal que $M(x) \neq \emptyset$, decimos que

1. x es un punto regular débil si existe $(\theta, \lambda, \mu) \in M(x)$ con $\theta \neq 0$.
2. x es un punto totalmente regular débil si para todo $(\theta, \lambda, \mu) \in M(x)$ existe $l \in \{1, \dots, r\}$ tal que $\theta_l \neq 0$.
3. x es un punto regular si existe $(\theta, \lambda, \mu) \in M(x)$ con $\theta_l > 0$ para todo $l \in \{1, \dots, r\}$.
4. x es un punto totalmente regular si para todo $(\theta, \lambda, \mu) \in M(x)$, se tiene $\theta_l > 0$ para todo $l \in \{1, \dots, r\}$.

2.2.3. Escalarización

Una de las principales estrategias para encontrar soluciones de problemas multiobjetivo es la escalarización [38, 39]. Utilizando esta estrategia se obtienen soluciones óptimas de un problema vectorial resolviendo uno o varios problemas de optimización escalar parametrizados. Entre las técnicas de escalarización se encuentra el método de los pesos en el cual uno minimiza una combinación lineal no negativa de las funciones objetivo [50]. Una de las dificultades principales de esta técnica es que los parámetros en la combinación no son conocidos de antemano y la dependencia de parámetros auxiliares puede generar la aparición de inestabilidad y dificultades numéricas si el problema de optimización tiene soluciones factibles para una cantidad pequeña de parámetros o bien no tiene solución para ningún rango de parámetros, [33].

Las funciones débilmente creciente y fuertemente creciente son funciones auxiliares que pertenecen a una clase de funciones vectoriales monótonas y son funciones que fueron ampliamente estudiadas en la literatura.

Definición 2.2.15. [48] Una función $\Phi : \mathbb{R}^r \rightarrow \mathbb{R}$ es llamada débilmente creciente (en inglés, weakly-increasing) si para todo $u, v \in \mathbb{R}^r$

$$u < v \implies \Phi(u) < \Phi(v).$$

Φ es llamada fuertemente creciente (en inglés, strongly-increasing) si para todo $u, v \in \mathbb{R}^r$

$$u \leq v \text{ y } u_{j_0} < v_{j_0} \text{ para algún } j_0 \Rightarrow \Phi(u) < \Phi(v).$$

Un ejemplo de una función débilmente creciente que no es fuertemente creciente es $\Phi(u) = \max_{i=1, \dots, r} \{u_i\}$.

Un ejemplo de una función fuertemente creciente es $\Phi(u) = \sum_{i=1}^r \Phi_i(u_i)$ donde $\Phi_i : \mathbb{R} \rightarrow \mathbb{R}$ y Φ_i es estrictamente creciente para todo $i = 1, \dots, r$.

Lema 2.2.16. 1. Si Φ es una función débilmente creciente y $x^* \in \underset{x \in \Omega}{\operatorname{argmin}} \Phi(F(x))$, entonces x^* es un punto Pareto optimal débil del problema (2.2.1).

2. Si Φ es una función fuertemente creciente y $x^* \in \underset{x \in \Omega}{\operatorname{argmin}} \Phi(F(x))$, entonces x^* es un punto Pareto optimal del problema (2.2.1).

Demostración. 1. Si x^* no es un punto Pareto optimal débil del problema (2.2.1), entonces existe $x \in \Omega$ tal que $F(x) < F(x^*)$.

Más aún, como Φ es una función débilmente creciente, $\Phi(F(x)) < \Phi(F(x^*))$ y contradice que $x^* \in \Omega$ sea minimizador de $\underset{x \in \Omega}{\operatorname{argmin}} \Phi(F(x))$.

2. Si x^* no es un punto Pareto optimal del problema, entonces existen $x \in \Omega$ y $j_0 \in \{1, \dots, r\}$ tal que $F(x) \leq F(x^*)$ y $F_{j_0}(x) < F_{j_0}(x^*)$. Más aún, como Φ es una función fuertemente creciente, $\Phi(F(x)) < \Phi(F(x^*))$ y contradice que $x^* \in \Omega$ sea minimizador de $\underset{x \in \Omega}{\operatorname{argmin}} \Phi(F(x))$. \square

Como mencionamos previamente, existe una gran cantidad de condiciones calificadoras tanto para el problema de optimización multiobjetivo *MOP* (2.2.1) como para el problema de programación no lineal *NLP* que exigen el cumplimiento de diferentes propiedades del conjunto de restricciones que definen al conjunto factible Ω . Desde el punto de vista teórico, es interesante debilitar las condiciones calificadoras y por otra parte manejar aquellas que sean más fáciles de verificar.

2.3. Condiciones calificadoras

Para presentar las distintas condiciones calificadoras o de calidad es necesario definir diferentes aproximaciones concernientes al conjunto factible Ω definido en

(2.2.1). Para presentar las definiciones y los resultados de esta sección seguimos el tratamiento realizado en el libro [10].

Definición 2.3.1. [12] Un conjunto C incluido en \mathbb{R}^n se dice cono si para cada x en C , αx pertenece a C para todo α mayor o igual a 0.

La primer aproximación que definimos es la que se denomina cono de primer orden de las variaciones factibles o cono tangente linealizado y se establece como sigue:

Definición 2.3.2. El cono de primer orden de las variaciones factibles en $x \in \Omega$ está dado por

$$L_{\Omega}(x) \equiv \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0 \text{ si } i = 1, \dots, m \text{ y } \nabla g_j(x)^T d \leq 0 \text{ si } j \in A(x)\}.$$

Para encontrar las direcciones que son tangentes al conjunto factible Ω adoptamos la siguiente definición.

Definición 2.3.3. El cono de tangentes de Ω en $x \in \Omega$ se define como

$$T_{\Omega}(x) \equiv \left\{ d \in \mathbb{R}^n : d = \lim_{k \rightarrow \infty} \alpha_k (x^k - x) \text{ donde } \alpha_k > 0, \{x^k\} \subset \Omega \text{ y } \lim_{k \rightarrow \infty} x^k = x \right\}.$$

Teorema 2.3.4. (Proposición 9.2.1 [32]) $T_{\Omega}(x) \subset L_{\Omega}(x)$.

En algunos casos es útil considerar el cono dual o polar para lo cual damos su definición a continuación.

Definición 2.3.5. El cono dual o polar de un cono dado C se define como

$$C^* = \{u \in \mathbb{R}^n : \langle t, u \rangle \leq 0 \text{ para todo } t \in C\}.$$

Definición 2.3.6. [13] El cono normal de Ω en $x \in \Omega$ se define como

$$N_{\Omega}(x) = T_{\Omega}^*(x).$$

Ahora, para $x \in \Omega$ consideremos el conjunto $G(x)$ definido por los vectores que son combinaciones lineales de los gradientes de las restricciones que componen el conjunto factible Ω utilizando escalares no negativos asociados a las restricciones de desigualdad activas en x :

$$G(x) = \left\{ z \in \mathbb{R}^n : z = \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_j \nabla g_j(x), \mu_j \geq 0 \text{ para todo } j \in A(x) \right\}.$$

Enunciamos el siguiente lema que se utilizará para la demostración del Teorema 2.3.8.

Lema 2.3.7. (*Versión geométrica del lema de Farkas*) Sea $C \subset \mathbb{R}^n$ un cono convexo cerrado. Entonces

$$C^{**} = C.$$

El siguiente teorema establece la relación que existe entre $L_\Omega(x)$ y $G(x)$ y sirve para describir al cono polar del tangente linealizado $L_\Omega(x)$ en términos de combinaciones lineales de los gradientes de las restricciones del problema (2.2.1).

Teorema 2.3.8. $L_\Omega^*(x) = G(x)$.

Demostración. Consideremos $d \in L_\Omega(x)$ y $s \in G(x)$. Tenemos que

$$d^T s = \sum_{i=1}^m \lambda_i d^T \nabla h_i(x) + \sum_{j \in A(x)} \mu_j d^T \nabla g_j(x).$$

Veamos que $L_\Omega(x) = G^*(x)$. Como $d \in L_\Omega(x)$, tenemos que $d^T \nabla h_i(x) = 0$ para todo $i = 1, \dots, m$ y $d^T \nabla g_j(x) \leq 0$ para todo $j \in A(x)$. Como $\mu_j \geq 0$, se tiene que $d^T s \leq 0$ para todo $s \in G(x)$. Luego $d \in G^*(x)$.

Supongamos ahora que $d \in G^*(x)$, entonces $d^T s \leq 0$ para todo $s \in G(x)$. Como $\nabla h_i(x)$ y $-\nabla h_i(x)$ son elementos de $G(x)$, tenemos que $d^T \nabla h_i(x) \leq 0$ y $-d^T \nabla h_i(x) \leq 0$ para todo $i \in \{1, \dots, m\}$. Por lo tanto $d^T \nabla h_i(x) = 0$ para todo $i = 1, \dots, m$.

Además, como $\nabla g_j(x) \in G(x)$, para todo $j \in A(x)$, tenemos que $d^T \nabla g_j(x) \leq 0$, para todo $j \in A(x)$, y esto implica que $d \in L_\Omega(x)$.

Por lo tanto se tiene $L_\Omega(x) = G^*(x)$ y usando el Lema 2.3.7 tenemos que $L_\Omega^*(x) = G(x)$. \square

En la siguiente subsección definimos algunas de las condiciones calificadoras que existen en la literatura y que mencionamos al inicio de este capítulo.

Dividimos la presentación de las condiciones de calidad siguiendo dos líneas: condiciones de calidad que se definen mediante el abordaje de conos y aquellas que se definen mediante el concepto de dependencia e independencia lineal.

2.3.1. Definiciones de condiciones calificadoras utilizando el concepto de cono

En esta subsección definimos las condiciones que serán utilizadas para el análisis de convergencia global de los algoritmos propuestos en los capítulos siguientes.

Definición 2.3.9. Decimos que la condición calificadora de Abadie (*ACQ*) [1] se satisface en $x \in \Omega$ si $L_\Omega(x) \subset T_\Omega(x)$.

Observación 2.3.10. Si $x \in \Omega$ cumple Abadie se tiene que $L_\Omega(x) = T_\Omega(x)$, ya que por el Teorema 2.3.4 la contención $T_\Omega(x) \subset L_\Omega(x)$ vale siempre. Luego $L_\Omega^*(x) = T_\Omega^*(x)$ y se tiene por la Definición 2.3.6 y el Teorema 2.3.8 que

$$N_\Omega(x) = G(x),$$

lo que nos proporciona una caracterización del cono normal $N_\Omega(x)$ en términos de las combinaciones lineales positivas de los gradientes de las restricciones definidas en el conjunto Ω .

Dado $x^* \in \Omega$ el conjunto $G(x^*)$ es un cono convexo y cerrado.

Dada una multifunción $F : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$, el límite exterior de $F(z)$ cuando $z \rightarrow z^*$ [59] es denotado por

$$\limsup_{z \rightarrow z^*} F(z) = \{w^* \in \mathbb{R}^d : \exists (z^k, w^k) \rightarrow (z^*, w^*) \text{ con } w^k \in F(z^k)\}.$$

Luego, la multifunción F se dice que es semicontinua exterior en z^* si

$$\limsup_{z \rightarrow z^*} F(z) \subset F(z^*).$$

Definición 2.3.11. [6] Decimos que $x^* \in \Omega$ satisface la Propiedad del Cono Continuo *CCP* si la multifunción $x \rightrightarrows G(x)$ es semicontinua exterior en x^* , es decir,

$$\limsup_{x \rightarrow x^*} G(x) \subset G(x^*).$$

2.3.2. Condiciones calificadoras y diferenciabilidad

Por completitud presentamos la definición de las condiciones calificadoras de Mangasarian-Fromovitz y de Dependencia Lineal Positiva Constante.

Definición 2.3.12. La condición de Mangasarian-Fromovitz (*MFCQ*) [49] se satisface en $x^* \in \Omega$ si

1. $\{\nabla h_i(x^*)\}_{i \in \{1, \dots, m\}}$ son linealmente independientes,
2. $L_\Omega(x^*) \neq \emptyset$.

Definición 2.3.13. [55] Un punto $x^* \in \Omega$ cumple la condición de Dependencia Lineal Positiva Constante (*CPLD*) si para cualquier $I_0 \subset \{1, \dots, m\}$ y $J_0 \subset A(x^*)$, siempre que

$$\{\nabla g_j(x^*)\}_{j \in J_0} \cup \{\nabla h_i(x^*)\}_{i \in I_0}$$

es positivo linealmente dependiente, hay un entorno $B(x^*, \epsilon)$ de x^* tal que para cualquier $y \in B(x^*, \epsilon)$,

$$\{\nabla g_j(y)\}_{j \in J_0} \cup \{\nabla h_i(y)\}_{i \in I_0}$$

es linealmente dependiente.

En [55] se demuestra que la condición de calidad de Dependencia Lineal Positiva Constante es más débil que Mangasarian-Fromovitz.

Definición 2.3.14. Un punto $x^* \in \Omega$ es Quasinormal [46] si no existen escalares $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$, tales que

- $\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0$.
- $\mu_j \geq 0, j = 1, \dots, p$.
- $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$, no todos nulos.
- En cada entorno $B(x^*, \epsilon)$ de x^* existe $x \in B(x^*, \epsilon)$ tal que $\lambda_i h_i(x) > 0$ para todo i con $\lambda_i \neq 0$, y $\mu_j g_j(x) > 0$ para todo j con $\mu_j \neq 0$.

En [8] se demuestra que la Condición de Dependencia Lineal Positiva Constante es más fuerte que la Quasinormalidad.

En [6] se demuestra que la condición de calidad de Abadie es más débil que la Propiedad del Cono Continuo. Y que esta última es más fuerte que la condición de Dependencia Lineal Positiva Constante pero que la Propiedad del Cono Continuo no tiene relación con la Quasinormalidad.

En la Figura 2.5 se observa la relación entre las condiciones mencionadas.

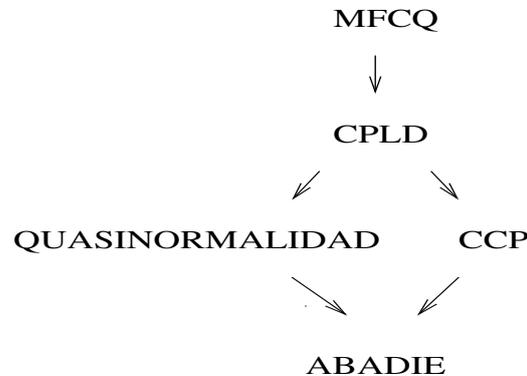


Figura 2.5: Relación entre las distintas condiciones de calidad.

La condición de calidad llamada Quasinormalidad fue introducida por Hestenes para el conjunto factible Ω . La extensión de Quasinormalidad al caso en el que se tiene el conjunto factible Ω junto con un conjunto abstracto de restricciones ($\bar{\Omega} \neq \mathbb{R}^n$) se investigó en [13].

Consideremos el siguiente problema

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeto a } x \in \Omega \cap \bar{\Omega} \end{aligned} \tag{2.3.15}$$

siendo $\bar{\Omega}$ un conjunto abstracto de restricciones. La Definición 2.3.14 fue extendida al caso en el cual el conjunto factible se define como $\Omega \cap \bar{\Omega}$ siendo $\bar{\Omega}$ un conjunto abstracto de restricciones.

Definición 2.3.16. [13] Decimos que un punto factible x^* del problema (2.3.15) es Quasinormal para $\Omega \cap \bar{\Omega}$ si no existen escalares $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ y una sucesión $\{x^k\} \subset \bar{\Omega}$ tales que

$$(i) \quad - \left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) \in N_{\bar{\Omega}}(x^*).$$

$$(ii) \quad \mu_j \geq 0, \quad j = 1, \dots, p.$$

$$(iii) \quad \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p \text{ no todos nulos.}$$

$$(iv) \quad \{x^k\} \text{ converge a } x^*, \text{ y para todo } k, \lambda_i h_i(x^k) > 0 \text{ para todo } i \text{ con } \lambda_i \neq 0, \text{ y } \mu_j g_j(x^k) > 0 \text{ para todo } j \text{ con } \mu_j \neq 0.$$

En el Capítulo 5 consideramos dos conjuntos de restricciones: Ω que será denotado el conjunto de restricciones del nivel superior y $\bar{\Omega}$ el conjunto de restricciones del nivel inferior. Para el análisis de convergencia utilizamos la condición que las restricciones que definen el conjunto $\bar{\Omega}$ deben cumplir la condición de Abadie.

Por ejemplo, conjuntos $\bar{\Omega}$ que satisfacen la condición de Abadie son los siguientes:

- $\bar{\Omega}_1 = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$ donde g es cóncava y h lineal;
- $\bar{\Omega}_2 = X = \{x \in \mathbb{R}^n : l \leq x \leq u\}$.

En el siguiente lema demostramos que $\bar{\Omega}_1$ cumple la condición de Abadie.

Lema 2.3.17. *Supongamos que $x^* \in \bar{\Omega}_1$, donde $\bar{\Omega}_1 = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$ siendo h lineal y g cóncava. Luego x^* satisface la condición de Abadie.*

Demostración. Veamos que x^* es Quasinormal, luego como Quasinormal implica Abadie, obtenemos que x^* cumple la condición de Abadie. Supongamos por el absurdo que x^* no es Quasinormal, entonces existen $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ tales que

$$(i) \quad \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0$$

$$(ii) \quad \mu \geq 0$$

$$(iii) \quad \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p \text{ no todos nulos}$$

$$(iv) \quad \text{En cada entorno de } x^* \text{ existe } x \text{ tal que } \lambda_i h_i(x) > 0 \text{ para todo } i = 1, \dots, m \text{ si } \lambda_i \neq 0, \text{ y } \mu_j g_j(x) > 0 \text{ para todo } j = 1, \dots, p \text{ si } \mu_j \neq 0.$$

Como h lineal y g cóncava

$$h_i(x) = h_i(x^*) + \nabla h_i^T(x^*)(x - x^*), i = 1, \dots, m \quad (2.3.18)$$

$$g_j(x) \leq g_j(x^*) + \nabla g_j^T(x^*)(x - x^*), j = 1, \dots, p \quad (2.3.19)$$

Multiplicando por λ_i en (2.3.18) y por μ_j en (2.3.19) y sumando, por (ii) se obtiene

$$\begin{aligned} \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x) &\leq \sum_{i=1}^m \lambda_i h_i(x^*) + \sum_{j=1}^p \mu_j g_j(x^*) + \\ &\left(\sum_{i=1}^m \lambda_i \nabla h_i^T(x^*) + \sum_{j=1}^p \mu_j \nabla g_j^T(x^*) \right)^T (x - x^*) \end{aligned} \quad (2.3.20)$$

Por otro lado, como x^* es factible, por (ii)

$$\sum_{i=1}^m \lambda_i h_i(x^*) + \sum_{j=1}^p \mu_j g_j(x^*) \leq 0. \quad (2.3.21)$$

Luego, por (i), (2.3.20) y (2.3.21) se tiene que

$$\sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x) \leq 0. \quad (2.3.22)$$

Por otro lado, por (iii) existe $\lambda_i \neq 0$ o $\mu_i \neq 0$. Si $\lambda_i \neq 0$, $\lambda_i h_i(x) > 0$ para todo i ; y si $\mu_j \neq 0$ $\mu_j g_j(x) > 0$ para todo j . Luego

$$\sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x) > 0.$$

Es absurdo puesto que contradice (2.3.22).

Por lo tanto x^* satisface la condición de Abadie. \square

Por el lema anterior el conjunto $\bar{\Omega}_2$ también cumple la condición de Abadie ya que está formado por restricciones lineales de desigualdad que son funciones cóncavas.

Observación 2.3.23. Para $X = [l, u]$ se tiene:

- Utilizando el Teorema 2.3.8 y la Observación 2.3.10, como X cumple la condición de Abadie, se obtiene para todo $x^* \in X$,

$$N_X(x^*) = \left\{ z \in \mathbb{R}^n : z = \sum_{i:x_i^*=l_i} \xi_i(-e_i) + \sum_{i:x_i^*=u_i} \nu_i(e_i), \xi_i \geq 0, \nu_i \geq 0 \right\}$$

- Para todo $x^* \in X = [l, u]$, por la Observación 2.1.10 y el resultado anterior. $P_X(x^* - \nabla f(x^*)) - x^* = 0$ si y solo si $-\nabla f(x^*) \in G(x^*) = N_X(x^*)$.

Capítulo 3

Convergencia del método Lagrangiano Aumentado usando Quasinormalidad y aplicación al problema multiobjetivo

En este capítulo consideramos el problema de optimización multiobjetivo con restricciones (*MOP*) de la siguiente forma:

$$\begin{aligned} & \text{Minimizar } F(x) \\ & \text{suje}to \text{ a } x \in \Omega \end{aligned} \tag{3.0.1}$$

donde $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$, $F(x) = (F_1(x), \dots, F_r(x))^T$, es una función vectorial diferenciable y Ω es un conjunto factible formado por restricciones de igualdad y de desigualdad de la forma

$$\Omega = \{x : h(x) = 0, g(x) \leq 0\} \tag{3.0.2}$$

para $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ y $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$. Supongamos que todas las funciones involucradas son continuamente diferenciables en \mathbb{R}^n .

Nos proponemos resolver el problema original usando una técnica popular en el caso escalar: el método Lagrangiano Aumentado. Este método ha sido ampliamente estudiado y utilizado en las últimas décadas en el caso $r = 1$. Puede encontrarse en el libro publicado recientemente [21] y sus referencias. El método Lagrangiano Aumentado es un método iterativo que consiste en agregar cierto término de penalidad a la función objetivo que tiene un costo alto para los puntos que no son factibles. En [2, 3] se demuestra la convergencia global de un algoritmo Lagrangiano Aumentado a puntos *KKT* suponiendo la condición de calidad Dependencia Lineal Positiva Constante [8, 55]. Estos resultados se ampliaron usando la condición de calidad Generador Po-

sitivo Constante (*CPG*) definida en [5]. Recientemente, se demostró la convergencia global a los puntos Karush-Kuhn-Tucker de un algoritmo Lagrangiano Aumentado usando la Propiedad del Cono Continuo (*CCP*) [6]. En [6] los autores muestran que la condición de *CCP* es más débil que *CPG* [5], más fuerte que Abadie [1] y que no tiene relación con Quasinormalidad [46]. La condición de calidad Quasinormalidad (Definición 2.3.14) es una condición muy general que se introdujo en [46] asociada a problemas escalares con restricciones, extendida a problemas con un conjunto abstracto de restricciones (Definición 2.3.16) en [13, 52] y utilizada en el contexto de condiciones de optimalidad sucesivas para desigualdades variacionales en [44].

En la Sección 3.1 demostramos que el método clásico Lagrangiano Aumentado [21] para problemas escalares converge utilizando la condición de calidad Quasinormalidad. Este resultado no había sido demostrado con anterioridad.

Además, en este capítulo extendemos el método Lagrangiano Aumentado [21] para resolver el problema de optimización multiobjetivo (3.0.1) y analizamos la convergencia global del mismo.

El método Lagrangiano Aumentado clásico utiliza una sucesión de subproblemas que son considerados más fáciles de resolver. En la sección 3.2, consideramos la condición aproximada Karush-Kuhn-Tucker para el *MOP* (*AKKT* para el *MOP*) definida recientemente en [40] para proporcionar el sentido en que se resuelve el subproblema en el caso vectorial. Mostramos que la propiedad *CCP* es la condición más débil que garantiza que *AKKT* para el *MOP* implica regularidad débil. En la Sección 3.3 se demuestra la convergencia global a puntos regulares débiles del algoritmo Lagrangiano Aumentado propuesto cuando el punto límite satisface la condición de Quasinormalidad o la Propiedad del Cono Continuo.

3.1. Algoritmo y convergencia global

En esta sección consideramos el problema escalar no lineal con restricciones de la forma:

$$\begin{aligned} & \textit{Minimizar} \quad f(x) \\ & \textit{sujeto a} \quad x \in \Omega \end{aligned} \tag{3.1.1}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es continuamente diferenciable en \mathbb{R}^n y Ω esta dado por (3.0.2). Dado $x^* \in \Omega$ denotamos $A(x^*) = \{j \in \{1, \dots, p\} : g_j(x^*) = 0\}$ al conjunto de restricciones de desigualdad activas en x^* .

Para resolver (3.1.1), el método clásico Lagrangiano Aumentado usa una sucesión de subproblemas más fáciles de resolver. En cada subproblema, dados el parámetro

de penalidad $\rho > 0$ fijo y estimados de los multiplicadores de Lagrange $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p, \mu \geq 0$, una función Lagrangiana Aumentada es minimizada aproximadamente. Una vez que se encuentra la solución aproximada, se definen nuevos parámetros y se inicia una nueva iteración. Si bien existen varias funciones de penalidad [29], la función de penalidad externa utilizada en este trabajo es la función cuadrática. Por su forma simple, la función de penalidad cuadrática es la más utilizada en la práctica, aunque de vez en cuando el uso de otras funciones de penalidad puede tener ventajas, véase por ejemplo [11, 29] y sus referencias.

Dados $x \in \mathbb{R}^n, \rho > 0, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p, \mu \geq 0$ la forma general de la función Lagrangiana Aumentada asociada al problema escalar (3.1.1) considerando la función de penalidad cuadrática es

$$L(x, \lambda, \mu, \rho) = f(x) + \frac{\rho}{2} \left[\sum_{i=1}^m \left(h_i(x) + \frac{\lambda_i}{\rho} \right)^2 + \sum_{j=1}^p \left(\max \left\{ 0, g_j(x) + \frac{\mu_j}{\rho} \right\}^2 \right) \right].$$

Presentamos a continuación el algoritmo inexacto del método Lagrangiano Aumentado propuesto para resolver (3.1.1), [21]:

Algoritmo 3.1 Los parámetros utilizados por el algoritmo son los siguientes:

$$\begin{aligned} \tau &\in [0, 1), \gamma > 1, \rho_0 \in \mathbb{R}_+, \\ -\infty &< \bar{\lambda}_i^{min} < \bar{\lambda}_i^{max} < \infty \quad \forall i = 1, \dots, m, \\ 0 &\leq \bar{\mu}_i^{min} < \bar{\mu}_i^{max} < \infty \quad \forall i = 1, \dots, p, \\ \bar{\lambda}_i^0 &\in [\bar{\lambda}_i^{min}, \bar{\lambda}_i^{max}] \quad \forall i = 1, \dots, m, \\ \bar{\mu}_i^0 &\in [0, \bar{\mu}_i^{max}] \quad \forall i = 1, \dots, p, \\ \sigma_j^0 &= \max \left\{ g_j(x^0), -\frac{\bar{\mu}_j^0}{\rho_0} \right\} \quad \forall j = 1, \dots, p. \end{aligned}$$

Hacer $k \leftarrow 0$.

Paso 1. Resolver el subproblema.

Calcular (si es posible) $x^k \in \mathbb{R}^n$ una solución aproximada de

$$\begin{aligned} \text{Minimizar } & L(x, \bar{\lambda}^k, \bar{\mu}^k, \rho_k) \\ \text{sujeto a } & x \in \mathbb{R}^n. \end{aligned} \tag{3.1.2}$$

De no ser posible calcular x^k , detener el algoritmo.

Paso 2. Estimar los multiplicadores de Lagrange.

Para $i = 1, \dots, m$ calcular

$$\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(x^k) \quad (3.1.3)$$

y

$$\bar{\lambda}_i^{k+1} \in [\bar{\lambda}_i^{min}, \bar{\lambda}_i^{max}] \quad (3.1.4)$$

Para todo $j = 1, \dots, p$ calcular

$$\mu_j^{k+1} = \text{máx} \{0, \bar{\mu}_j^k + \rho_k g_j(x^k)\} \quad (3.1.5)$$

$$\sigma_j^k = \text{máx} \left\{ g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k} \right\} \quad (3.1.6)$$

$$\bar{\mu}_j^{k+1} \in [0, \bar{\mu}_j^{max}] \quad (3.1.7)$$

Paso 3. Adaptar el parámetro de penalidad.

Si $k = 0$ o

$$\text{máx} \{ \|h(x^k)\|_\infty, \|\sigma^k\|_\infty \} \leq \tau \text{máx} \{ \|h(x^{k-1})\|_\infty, \|\sigma^{k-1}\|_\infty \}$$

definir $\rho_{k+1} = \rho_k$.

Sino, definir $\rho_{k+1} = \gamma \rho_k$.

Hacer $k \leftarrow k + 1$. Ir al Paso 1.

Luego vamos a definir el sentido en el que se debe considerar una minimización aproximada en el Paso 1 para obtener el resultado de convergencia global.

El parámetro σ^k mide la infactibilidad y la no complementariedad con respecto a las restricciones de desigualdad.

Para obtener (3.1.6), se reemplaza la restricción de desigualdad $g_j(x) \leq 0$ por la restricción de igualdad utilizando variables de holgura $g_j(x) + z_j^2 = 0$ y se calcula z_j optimizando la función Lagrangiano Aumentado cuadrática obteniendo

$$z_j = \sqrt{\text{máx} \left\{ 0, -\frac{\mu_j}{\rho} - g_j(x) \right\}}.$$

Luego, se define $\sigma_j^k = g_j(x^k) + z_j^2 = \text{máx} \left\{ g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k} \right\}$.

Como se estableció en [21] en el caso cuadrático y en [60] en el caso no cuadrático, se puede demostrar que las estimaciones de los multiplicadores correspondientes a restricciones inactivas son nulos en el límite:

Teorema 3.1.8. *Supongamos que la sucesión generada por el Algoritmo 3.1 es tal que $\lim_{k \in K} x^k = x^*$ para $K \subset \mathbb{N}$. Entonces si $g_i(x^*) < 0$ se tiene que $\lim_{k \in K} \mu_i^{k+1} = 0$.*

Demostración. Por (3.1.5) sabemos que

$$\mu_i^{k+1} = \max \{0, \bar{\mu}_i^k + \rho_k g_i(x^k)\}.$$

Supongamos que $g_i(x^*) < 0$ y sean $c > 0$ y k_1 tales que $g_i(x^k) < c < 0$ para $k \geq k_1, k \in K$.

Luego consideremos dos casos.

1. Si la sucesión $\{\rho_k\}$ tiende a infinito, como $\{\bar{\mu}_i^k\}$ es acotada, existe $k_2 \geq k_1$, tal que, para todo $k \geq k_2, k \in K, \bar{\mu}_i^k + \rho_k g_i(x^k) < 0$. Esto implica que $\mu_i^{k+1} = 0$ para todo $k \geq k_2, k \in K$.

2. Si la sucesión $\{\rho_k\}$ es acotada, entonces vamos a demostrar que

$$\lim_{k \rightarrow \infty} \sigma_i^k = \lim_{k \rightarrow \infty} \max \left\{ g_i(x^k), -\frac{\bar{\mu}_i^k}{\rho_k} \right\} = 0.$$

Esto es, para $k \geq \bar{k}$, ρ_k no aumenta. Luego,

$$\|\sigma^k\|_\infty \leq \max \{ \|h(x^k)\|_\infty, \|\sigma^k\|_\infty \} \leq \tau \max \{ \|h(x^{k-1})\|_\infty, \|\sigma^{k-1}\|_\infty \}$$

Es por ello que

$$\begin{aligned} \|\sigma^{k+1}\|_\infty &\leq \tau \max \{ \|h(x^k)\|_\infty, \|\sigma^k\|_\infty \} \\ \|\sigma^{k+2}\|_\infty &\leq \tau^2 \max \{ \|h(x^k)\|_\infty, \|\sigma^k\|_\infty \} \end{aligned}$$

y así sucesivamente.

Como $g_i(x^k) < c < 0$ para $k \geq k_1, k \in K$, se tiene que $\lim_{k \in K} \frac{\bar{\mu}_i^k}{\rho_k} = 0$. Entonces, como la sucesión $\{\rho_k\}$ es acotada, $\lim_{k \in K} \bar{\mu}_i^k = 0$. Luego para $k \in K$ grande $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0$. Esto implica que $\lim_{k \in K} \mu_i^{k+1} = 0$. Es lo que queríamos probar. \square

Corolario 3.1.9. *Si x^* es un punto factible y $\lim_{k \in K} x^k = x^*$ entonces para todo $i = 1, \dots, p$*

$$\lim_{k \in K} \min \{ \mu_i^{k+1}, -g_i(x^k) \} = 0.$$

Demostración. Como x^* es factible, $h_i(x^*) = 0$ para todo $i = 1, \dots, m$ y $g_i(x^*) \leq 0$ para todo $i = 1, \dots, p$.

Si $g_i(x^*) < 0$ entonces por el Teorema 3.1.8 $\lim_{k \rightarrow \infty} \mu_i^{k+1} = 0$, entonces

$$\lim_{k \in K} \min \{ \mu_i^{k+1}, -g_i(x^k) \} = 0.$$

Si $g_i(x^*) = 0$, $\lim_{k \in K} g_i(x^k) = 0$. Luego, como $\mu_i^{k+1} \geq 0$,

$$\lim_{k \in K} \min \{ \mu_i^{k+1}, -g_i(x^k) \} = 0.$$

□

Vamos a demostrar que el Algoritmo 3.1 converge a un punto *KKT* usando la condición de Quasinormalidad. Por lo que sabemos es la primera vez que la condición de Quasinormalidad se utiliza para analizar la convergencia del método Lagrangiano Aumentado. Establecemos nuevamente la definición de Quasinormalidad por completitud.

Definición 3.1.10. [46] Se dice que $x^* \in \Omega$ satisface la condición de calidad Quasinormalidad si no existen escalares $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ tales que

(i)
$$\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0;$$

(ii) $\mu_j \geq 0, j = 1, \dots, p;$

(iii) $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ no todos nulos;

(iv) en cada entorno $B(x^*, \epsilon)$ de x^* existe $y \in B(x^*, \epsilon)$ tal que $\lambda_i h_i(y) > 0$ para todo i con $\lambda_i \neq 0$ y $\mu_j g_j(y) > 0$ para todo j con $\mu_j > 0$.

Teorema 3.1.11. Sea $\{\epsilon_k\}$ una sucesión acotada tal que $\lim_{k \rightarrow \infty} \epsilon_k = 0$. Sea $\{x^k\}$ una sucesión generada por el Algoritmo 3.1 satisfaciendo, para cada $k \in \mathbb{N}$:

$$\|\nabla_x L(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)\| \leq \epsilon_k. \tag{3.1.12}$$

Supongamos que x^* es un punto límite factible de $\{x^k\}$ tal que la condición de calidad Quasinormal se satisface. Entonces, x^* es un punto *KKT* del problema original (3.1.1).

Demostración. Sea K un subconjunto infinito de \mathbb{N} tal que $\lim_{k \in K} x^k = x^*$.

Por (3.1.12), (3.1.3) y (3.1.5) se tiene que existe una sucesión de parámetros de tolerancia $\{\epsilon^k\} \subset \mathbb{R}^n$ tal que $\epsilon^k \geq 0$, $\|\epsilon^k\| \leq \varepsilon_k$ y

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^{k+1} \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^{k+1} \nabla g_j(x^k) = \epsilon^k. \quad (3.1.13)$$

Definimos $\delta_k = \sqrt{1 + \sum_{i=1}^m (\lambda_i^{k+1})^2 + \sum_{j=1}^p (\mu_j^{k+1})^2}$. Luego $\delta_k > 0$. Dividiendo (3.1.13) por δ_k obtenemos que:

$$\frac{1}{\delta_k} \nabla f(x^k) + \sum_{i=1}^m \left(\frac{\lambda_i^{k+1}}{\delta_k} \right) \nabla h_i(x^k) + \sum_{j=1}^p \left(\frac{\mu_j^{k+1}}{\delta_k} \right) \nabla g_j(x^k) = \frac{\epsilon^k}{\delta_k}. \quad (3.1.14)$$

Ahora, como la sucesión $\left\{ \left(\frac{1}{\delta_k}, \frac{\lambda^{k+1}}{\delta_k}, \frac{\mu^{k+1}}{\delta_k} \right) \right\}_{k \in K}$ es acotada, existen un subconjunto $\bar{K} \subset K$ y escalares $\theta \in \mathbb{R}$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ tales que $\theta \geq 0$, $\mu \geq 0$,

$$\lim_{k \in \bar{K}} \left(\frac{1}{\delta_k}, \frac{\lambda^{k+1}}{\delta_k}, \frac{\mu^{k+1}}{\delta_k} \right) = (\theta, \lambda, \mu) \quad (3.1.15)$$

$$\|(\theta, \lambda, \mu)\| = 1. \quad (3.1.16)$$

Tomando límite cuando $k \in \bar{K}$ tiende a infinito en (3.1.14) se obtiene que

$$\theta \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0, \text{ con } \theta \geq 0 \text{ y } \mu \geq 0. \quad (3.1.17)$$

Supongamos que $\theta = 0$.

Entonces, por (3.1.16) y (3.1.17) existen escalares $\lambda_1, \dots, \lambda_m$, μ_1, \dots, μ_p no todos nulos tales que $\mu \geq 0$ y

$$\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0. \quad (3.1.18)$$

Por lo tanto por el Corolario 3.1.9 se tiene que $\mu_j = 0$ si $j \notin A(x^*)$.

Observar que, por construcción, las sucesiones $\{\bar{\lambda}^k\}_{k \in \bar{K}}$ y $\{\bar{\mu}^k\}_{k \in \bar{K}}$ son acotadas y como $\lim_{k \in \bar{K}} \frac{1}{\delta_k} = \theta = 0$, existe un subconjunto $K_1 \subset \bar{K}$ tal que

$\lim_{k \in K_1} \frac{\bar{\lambda}^k}{\delta_k} = 0$, $\lim_{k \in K_1} \frac{\bar{\mu}^k}{\delta_k} = 0$. Entonces, por (3.1.3), (3.1.5) y (3.1.15) se obtiene que

$$\lim_{k \in K_1} \frac{\rho_k h(x^k)}{\delta_k} = \lambda \text{ y } \lim_{k \in K_1} \frac{\rho_k \sigma^k}{\delta_k} = \mu.$$

Entonces, para cada $\lambda_i \neq 0$ existe $k_i \in K_1$ tal que si $k \geq k_i$, $k \in K_1$,

$$\lambda_i h_i(x^k) > 0. \quad (3.1.19)$$

Además, para cada $\mu_j > 0$ existe $\tilde{k}_j \in K_1$ tal que si $k \geq \tilde{k}_j$, $k \in K_1$, $\mu_j \sigma_j^k > 0$. Entonces, como $\sigma_j^k = \max \left\{ -\frac{\tilde{\mu}_j^k}{\rho_k}, g_j(x^k) \right\}$, se satisface

$$\mu_j g_j(x^k) > 0. \quad (3.1.20)$$

Por lo tanto, existe $\bar{k} \in K_1$, $\bar{k} \geq \max \left\{ \max_{\{i:\lambda_i \neq 0\}} \{k_i\}, \max_{\{j:\mu_j > 0\}} \{\tilde{k}_j\} \right\}$ tal que (3.1.19) y (3.1.20) valen para todo $k \geq \bar{k}$, $k \in K_1$.

Entonces, si $\theta = 0$, por (3.1.18)-(3.1.20) se contradice la definición de Quasinormalidad. Por lo tanto $\theta \neq 0$ y por (3.1.17) y el Corolario 3.1.9 x^* satisface la condición *KKT*. \square

3.2. Condición *KKT* Aproximada para el *MOP*

Como ya hemos mencionado en la introducción, la condición de calidad más débil que había sido utilizada para establecer convergencia global del algoritmo Lagrangiano Aumentado a los puntos *KKT* era la propiedad recientemente definida *CCP*, [6] que establecemos nuevamente por completitud.

Definición 3.2.1. [6] Se dice que $x^* \in \Omega$ satisface la Propiedad del Cono Continuo (*CCP*) si la multifunción $x \Rightarrow K(x)$ es semicontinua exterior en x^* , es decir, $\limsup_{x \rightarrow x^*} K(x) \subset K(x^*)$ donde

$$K(x) = \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x) : \lambda_i \in \mathbb{R}, \mu_j \geq 0 \right\} \quad (3.2.2)$$

y $\limsup_{x \rightarrow x^*} K(x) = \{w^* \in \mathbb{R}^n : \exists (x^k, w^k) \rightarrow (x^*, w^*) \text{ con } w^k \in K(x^k)\}$.

En [6] los autores prueban que *CCP* es más débil que *CPG* [5] y más fuerte que Abadie [1] y demuestran que *CCP* y Quasinormalidad [46] son condiciones de calidad independientes.

Observación 3.2.3. Consideremos el *MOP* (3.0.1) y la Definición 2.2.14. Usando el cono (3.2.2) dado en la definición de *CCP*, la condición de regularidad débil para (3.0.1) ((2) de la Definición 2.2.14) en $x^* \in \Omega$ se puede escribir como: existe $\theta \in \mathbb{R}^r, \theta \geq 0$ con $\theta \neq 0$ tal que

$$-\sum_{l=1}^r \theta_l \nabla F_l(x^*) \in K(x^*).$$

En [40] los autores, inspirados en los trabajos sobre condiciones de optimalidad secuenciales en el caso escalar, como [4, 7, 44], introducen la condición de Karush-Kuhn-Tucker Aproximada para el *MOP*.

Definición 3.2.4. [40] Se dice que un punto $x^* \in \Omega$ satisface la condición de Karush-Kuhn-Tucker Aproximada (*AKKT*) para el *MOP* si existen sucesiones $\{x^k\} \subset \mathbb{R}^n$, $\{\theta^k\} \subset \mathbb{R}^r$, $\theta^k \geq 0$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p$, $\mu^k \geq 0$ tales que $\lim_{k \rightarrow \infty} x^k = x^*$,

$$\lim_{k \rightarrow \infty} \left(\sum_{l=1}^r \theta_l^k \nabla F_l(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \right) = 0, \quad (3.2.5)$$

$$\sum_{l=1}^r \theta_l^k = 1, \quad (3.2.6)$$

$$g_j(x^*) < 0 \implies \mu_j^k = 0 \text{ para } k \text{ suficientemente grande, } j = 1, \dots, p. \quad (3.2.7)$$

En el Teorema 3.1 del trabajo [40] los autores prueban que *AKKT* para el *MOP* es una condición de optimalidad para un punto Pareto optimal débil local. Por completitud enunciamos el teorema.

Teorema 3.2.8. Si $x^* \in \Omega$ es un punto Pareto débil local del problema (3.0.1), entonces x^* satisface la condición *AKKT* para el *MOP* con sucesiones $\{x^k\}$ y $\{(\theta^k, \lambda^k, \mu^k)\}$. Además se tiene que

1. $\mu^k = b_k \max\{0, g(x^k)\}$ y $\lambda^k = c_k h(x^k)$, donde $b_k, c_k > 0$, para todo k , y para $l = 1, \dots, r$

2. $F_l(x^k) - F_l(x^*) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \leq 0$ para todo k .

En el siguiente teorema demostramos que, como ocurre en el caso escalar (Teorema 3.1 en [6]) la condición de calidad *CCP* es la más débil tal que *AKKT* para el *MOP* implica regularidad débil.

Teorema 3.2.9. *La condición de calidad CCP es la más débil bajo la cual AKKT para el MOP implica regularidad débil, independientemente de la función objetivo vectorial.*

Demostración. Veamos primero que si CCP vale, la condición secuencial AKKT para el MOP implica regularidad débil, independientemente de la función objetivo vectorial.

Sea $F(x) = (F_1(x), \dots, F_r(x))^T$ una función objetivo vectorial tal que la condición secuencial AKKT para el MOP se satisface en un punto factible x^* . Luego existen sucesiones $\{x^k\} \subset \mathbb{R}^n$, $\{\theta^k\} \subset \mathbb{R}^r$, $\theta^k \geq 0$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p$, $\mu^k \geq 0$ tales que

$$\lim_{k \rightarrow \infty} x^k = x^*, \mu_j^k = 0 \text{ para } j \notin A(x^*) \text{ para } k \text{ suficientemente grande, } \sum_{l=1}^r \theta_l^k = 1,$$

$$\zeta^k = \sum_{l=1}^r \theta_l^k \nabla F_l(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \text{ y } \lim_{k \rightarrow \infty} \zeta^k = 0.$$

Definimos $\omega^k = \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k)$. Entonces, $\omega^k \in K(x^k)$ y

$$\omega^k = \zeta^k - \sum_{l=1}^r \theta_l^k \nabla F_l(x^k). \quad (3.2.10)$$

Por (3.2.6), la sucesión $\{\theta^k\}_{k \in \mathbb{N}}$ es acotada y entonces existen un subconjunto $\bar{K} \subset \mathbb{N}$ y un vector $\theta \in \mathbb{R}^r$, $\theta \geq 0$ tales que $\lim_{k \in \bar{K}} \theta^k = \theta$ y $\sum_{l=1}^r \theta_l = 1$. Por lo tanto $\theta \neq 0$. Tomando límite en (3.2.10) con $k \in \bar{K}$ cuando tiende a infinito y usando la continuidad de los gradientes de F_1, \dots, F_r se tiene que

$$-\sum_{l=1}^r \theta_l \nabla F_l(x^*) = \lim_{k \rightarrow \bar{K}} \omega^k \in \limsup_{k \rightarrow \infty} K(x^k) \subset \limsup_{x \rightarrow x^*} K(x) \subset K(x^*),$$

donde la última inclusión se deriva de la definición de CCP. Por lo tanto, dado que $\theta \neq 0$, por la Observación 3.2.3 x^* es débil regular como queríamos probar.

Ahora, demostraremos que, si la condición secuencial AKKT para el MOP implica regularidad débil para cualquier función objetivo vectorial, entonces la condición CCP se satisface.

Sea $\omega^* \in \limsup_{x \rightarrow x^*} K(x)$, por definición de límite exterior, existen sucesiones $\{x^k\}$, $\{\omega^k\}$ tales que $\lim_{k \rightarrow \infty} x^k = x^*$, $\lim_{k \rightarrow \infty} \omega^k = \omega^*$ y $\omega^k \in K(x^k)$.

Definimos $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$ como $F(x) = (-(w^*)^T x, \dots, -(w^*)^T x)^T$ para todo $x \in \mathbb{R}^n$. Entonces la condición *AKKT* para el MOP se satisface en x^* para esta función vectorial ya que existe $\{\theta^k\} \subset \mathbb{R}^r$, $\theta^k \geq 0$ tal que $\lim_{k \rightarrow \infty} x^k = x^*$, $\sum_{l=1}^r \theta_l^k = 1$ y

$$\sum_{l=1}^r \theta_l^k \nabla F_l(x^k) + \omega^k = -\omega^* + \omega^k \rightarrow 0.$$

Entonces por hipótesis la condición de regularidad débil se satisface en x^* , esto es, existe $\theta \in \mathbb{R}^r$, $\theta \geq 0$, $\theta \neq 0$ tal que $-\sum_{l=1}^r \theta_l \nabla F_l(x^*) = -\omega^* \in K(x^*)$. \square

3.3. Aplicación del método Lagrangiano Aumentado para el MOP

Una de las propiedades más importantes del método Lagrangiano Aumentado es que, por su filosofía simple, se puede adaptar fácilmente para resolver diferentes clases de problemas: problemas de optimización sin derivadas [28], problemas de optimización global [18, 19] y problemas portfolio [20], para mencionar algunos de ellos.

Dados $x \in \mathbb{R}^n$, $\rho > 0$, $\theta \in \mathbb{R}^r$, $\theta \geq 0$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$ definimos la función de Fritz-John Aumentada asociada al problema (3.0.1) como:

$$(FJ)(x, \theta, \lambda, \mu, \rho) = \sum_{l=1}^r \theta_l F_l(x) + \frac{\rho}{2} \left[\sum_{i=1}^m \left(h_i(x) + \frac{\lambda_i}{\rho} \right)^2 + \sum_{j=1}^p \left(\max \left\{ 0, g_j(x) + \frac{\mu_j}{\rho} \right\} \right)^2 \right].$$

Dado que el algoritmo clásico Lagrangiano Aumentado procede minimizando aproximadamente la función Lagrangiana Aumentada en cada iteración, necesitamos definir el sentido en el cual esta minimización se interpreta en el caso vectorial. Para esto se aplica el Algoritmo 3.1 para resolver el problema (3.1.1) para $f(x) = \sum_{l=1}^r \theta_l F_l(x)$. La siguiente suposición establece un criterio bajo el cual x^k es un minimizador aproximado del subproblema

$$\begin{aligned} & \text{Minimizar } (FJ)(x, \theta^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k) \\ & \text{sujeto a } x \in \mathbb{R}^n. \end{aligned}$$

Suposición 1 En el Paso 1 del Algoritmo 3.1, obtenemos $x^k \in \mathbb{R}^n$ tal que existe

$\theta^k \in \mathbb{R}^r, \theta^k \geq 0$ satisfaciendo

$$\|\nabla_x(FJ)(x^k, \theta^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)\| \leq \varepsilon_k, \quad (3.3.1)$$

$$\sum_{l=1}^r \theta_l^k = 1, \quad (3.3.2)$$

donde la sucesión $\{\varepsilon_k\}$ es acotada tal que $\varepsilon_k \geq 0$.

En el siguiente teorema establecemos que, cuando el algoritmo admite un punto factible, este punto satisface la condición de optimalidad *AKKT* para el *MOP* de la Definición 3.2.4.

Teorema 3.3.3. *Supongamos que la sucesión $\{x^k\}$ es generada por el Algoritmo 3.1 satisfaciendo la Suposición 1 y $K \subset \mathbb{N}$ es tal que $\lim_{k \in K} x^k = x^*$ y x^* es factible. Además supongamos que la sucesión acotada $\{\varepsilon_k\}$ en la Suposición 1 es tal que $\lim_{k \rightarrow \infty} \varepsilon_k = 0$. Entonces, x^* satisface la condición *AKKT* para el *MOP* (3.0.1).*

Demostración. Como $\lim_{k \rightarrow \infty} \varepsilon_k = 0$, por la Suposición 1, (3.1.3) y (3.1.5) se tiene que existen sucesiones $\{\theta^k\} \subset \mathbb{R}^r, \theta^k \geq 0, \{\lambda^{k+1}\} \subset \mathbb{R}^m, \{\mu^{k+1}\} \subset \mathbb{R}^p, \mu^k \geq 0$ tales que $\lim_{k \in K} x^k = x^*$ y las igualdades (3.2.5) y (3.2.6) valen. Del Teorema 3.1.8 se obtiene que (3.2.7) vale y por lo tanto x^* satisface la condición *AKKT* para el *MOP* (3.0.1). \square

La solución aproximada x^k satisfaciendo (3.3.1) y (3.3.2) se puede obtener usando, por ejemplo, funciones débilmente creciente definidas en la Sección 2.2.3.

Recordemos la definición de débilmente creciente.

Definición 3.3.4. [48] Una función $\Phi : \mathbb{R}^r \rightarrow \mathbb{R}$ es llamada débilmente creciente (en inglés, weakly-increasing) si para todo $u, v \in \mathbb{R}^r$

$$u < v \implies \Phi(u) < \Phi(v).$$

Algunos ejemplos de funciones débilmente creciente son $\Phi(u) := \sum_{i=1}^r a_i u_i$

y $\Phi(u) := \sum_{i=1}^r a_i e^{u_i}$ con $a_i > 0$. Observar que, en ambos ejemplos, la función Φ es diferenciable y $\nabla \Phi(u) > 0$.

Recordemos la definición de la condición de optimalidad *KKT* Aproximada para el problema escalar por completitud.

Definición 3.3.5. Se dice que un punto $x^* \in \Omega$ satisface la condición de optimalidad llamada *KKT Aproximada (AKKT)* [6, 21] para el problema de $\begin{array}{l} \text{Minimizar } f(x) \\ \text{sujeto a } x \in \Omega \end{array}$, si existen sucesiones $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p$, $\mu^k \geq 0$ tales que $\lim_{k \rightarrow \infty} x^k = x^*$ y

$$\lim_{k \rightarrow \infty} \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \right) = 0, \quad (3.3.6)$$

$$\lim_{k \rightarrow \infty} \min\{\mu_j^k, -g_j(x^k)\} = 0 \text{ para todo } j = 1, \dots, p. \quad (3.3.7)$$

En la siguiente Proposición mostramos que se puede obtener una solución aproximada que verifica (3.3.1) y (3.3.2) usando una clase de funciones diferenciables débilmente creciente.

Proposición 3.3.8. *Sea Φ una función débilmente creciente con primera derivada continua para la cual $\nabla \Phi(u) \neq 0$. Supongamos que x^* satisface la condición de optimalidad secuencial AKKT asociada al problema escalar*

$$\begin{array}{l} \text{Minimizar } \Phi(F(x)) \\ \text{sujeto a } x \in \Omega. \end{array} \quad (3.3.9)$$

Entonces x^ satisface la condición de optimalidad secuencial AKKT para el MOP asociada al problema vectorial (3.0.1).*

Demostración. Como x^* satisface la condición AKKT asociada a (3.3.9), entonces existen sucesiones $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p$, $\mu^k \geq 0$ tales que $\lim_{k \rightarrow \infty} x^k = x^*$, $\lim_{k \rightarrow \infty} \min\{\mu_j^k, -g_j(x^k)\} = 0$, para todo $j = 1, \dots, p$, y por (3.3.6) existe $\{\epsilon^k\} \subset \mathbb{R}^n$, $\epsilon^k \geq 0$ tal que $\lim_{k \rightarrow \infty} \epsilon^k = 0$ y

$$\sum_{l=1}^r \frac{\partial \Phi(F(x^k))}{\partial u_l} \nabla F_l(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) = \epsilon^k. \quad (3.3.10)$$

Como Φ es débilmente creciente y $\nabla \Phi(u) \neq 0$ existe $K \subset \mathbb{N}$ tal que para todo $k \in K$,

$$\beta_k = \sum_{l=1}^r \frac{\partial \Phi(F(x^k))}{\partial u_l} \geq c > 0.$$

Dividiendo (3.3.10) por β_k se obtiene:

$$\sum_{l=1}^r \frac{\frac{\partial \Phi(F(x^k))}{\partial u_l}}{\beta_k} \nabla F_l(x^k) + \sum_{i=1}^m \frac{\lambda_i^k}{\beta_k} \nabla h_i(x^k) + \sum_{j=1}^p \frac{\mu_j^k}{\beta_k} \nabla g_j(x^k) = \frac{\epsilon^k}{\beta_k}.$$

Entonces, usando el Corolario 3.1.9, existen sucesiones $\{\theta^k\}$, $\{\bar{\lambda}^k\}$ y $\{\bar{\mu}^k\}$ definidas como

$$\theta^k = \frac{\nabla\Phi(F(x^k))}{\beta_k}, \quad \bar{\lambda}^k = \frac{\lambda^k}{\beta_k}, \quad \bar{\mu}^k = \frac{\mu^k}{\beta_k}$$

tales que (3.2.5)-(3.2.7) se verifican. Entonces x^* satisface la condición de optimalidad secuencial *AKKT* para el *MOP* asociada a la función vectorial F . \square

Usando el Teorema 3.2.9 obtenemos el siguiente resultado de convergencia global.

Teorema 3.3.11. *Supongamos que bajo las hipótesis del Teorema 3.3.3 x^* satisface la propiedad CCP, luego x^* es un punto regular débil del problema vectorial (3.0.1).*

Demostración. Por el Teorema 3.3.3 se tiene que x^* satisface la condición *AKKT* para el *MOP*. Luego, por el Teorema 3.2.9 x^* es un punto regular débil como queríamos probar. \square

En el siguiente teorema demostramos que se puede obtener el mismo resultado de convergencia utilizando la condición de calidad de Quasinormalidad.

Teorema 3.3.12. *Supongamos que bajo las hipótesis del Teorema 3.3.3, x^* satisface la condición de Quasinormalidad, luego x^* es un punto regular débil del problema vectorial (3.0.1).*

Demostración. Por (3.3.1), (3.1.3) y (3.1.5) tenemos que existe una sucesión de parámetros de tolerancia $\{\epsilon^k\} \subset \mathbb{R}^n$ tal que $\|\epsilon^k\| \leq \epsilon_k$ y para cada $k \in \mathbb{N}$,

$$\sum_{l=1}^r \theta_l^k \nabla F_l(x^k) + \sum_{i=1}^m \lambda_i^{k+1} \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^{k+1} \nabla g_j(x^k) = \epsilon^k. \quad (3.3.13)$$

Definimos $\delta_k = \sqrt{\sum_{l=1}^r (\theta_l^k)^2 + \sum_{i=1}^m (\lambda_i^{k+1})^2 + \sum_{j=1}^p (\mu_j^{k+1})^2}$. Por (3.3.2) tenemos que $\delta_k > 0$. Dividiendo (3.3.13) por δ_k obtenemos:

$$\sum_{l=1}^r \left(\frac{\theta_l^k}{\delta_k} \right) \nabla F_l(x^k) + \sum_{i=1}^m \left(\frac{\lambda_i^{k+1}}{\delta_k} \right) \nabla h_i(x^k) + \sum_{j=1}^p \left(\frac{\mu_j^{k+1}}{\delta_k} \right) \nabla g_j(x^k) = \frac{\epsilon^k}{\delta_k}. \quad (3.3.14)$$

Ahora, la sucesión $\left\{ \left(\frac{\theta^k}{\delta_k}, \frac{\lambda^{k+1}}{\delta_k}, \frac{\mu^{k+1}}{\delta_k} \right) \right\}$ es acotada y existe un subconjunto $\bar{K} \subset K$ y vectores $\theta \in \mathbb{R}^r$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ tales que $\theta \geq 0$, $\mu \geq 0$,

$$\lim_{k \in \bar{K}} \left(\frac{\theta^k}{\delta_k}, \frac{\lambda^{k+1}}{\delta_k}, \frac{\mu^{k+1}}{\delta_k} \right) = (\theta, \lambda, \mu), \quad y \quad (3.3.15)$$

$$\|(\theta, \lambda, \mu)\| = 1. \quad (3.3.16)$$

Tomando límite para $k \in \bar{K}$ tendiendo a infinito en (3.3.14) obtenemos que

$$\sum_{l=1}^r \theta_l \nabla F_l(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0 \text{ con } \mu \geq 0 \text{ y } \theta \geq 0. \quad (3.3.17)$$

Supongamos que $\theta = 0$. Entonces,

$$\lim_{k \in \bar{K}} \frac{\theta^k}{\delta_k} = 0. \quad (3.3.18)$$

La sucesión $\{\theta^k\}_{k \in \bar{K}}$ verifica $\sum_{l=1}^r \theta_l^k = 1$, $\theta^k \geq 0$ entonces es una sucesión acotada y existe un subconjunto $\bar{\bar{K}} \subset \bar{K}$ y $\bar{\theta} \in \mathbb{R}^r$ tal que $\lim_{k \in \bar{\bar{K}}} \theta^k = \bar{\theta}$. Entonces, tenemos que $\sum_{l=1}^r \bar{\theta}_l = 1$ y $\bar{\theta} \geq 0$. Luego $\bar{\theta} \neq 0$ y, por (3.3.18), $\lim_{k \in \bar{\bar{K}}} \frac{1}{\delta_k} = 0$.

Observar que, por (3.3.16) y (3.3.17) existen escalares $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ no todos nulos tal que $\mu \geq 0$,

$$\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0 \quad (3.3.19)$$

y, usando el Corolario 3.1.9, tenemos que $\mu_j = 0$ si $j \notin A(x^*)$.

Por construcción las sucesiones $\{\bar{\lambda}^k\}_{k \in \bar{K}}$, $\{\bar{\mu}^k\}_{k \in \bar{K}}$ son acotadas y como $\lim_{k \in \bar{K}} \frac{1}{\delta_k} = 0$, existe un subconjunto $K_1 \subset \bar{K}$ tal que $\lim_{k \in K_1} \frac{\bar{\lambda}^k}{\delta_k} = 0$ y $\lim_{k \in K_1} \frac{\bar{\mu}^k}{\delta_k} = 0$. Entonces, por (3.3.15), (3.1.3) y (3.1.5), obtenemos que

$$\lim_{k \in K_1} \frac{\rho_k h(x^k)}{\delta_k} = \lambda \text{ y } \lim_{k \in K_1} \frac{\rho_k \max \left\{ -\frac{\bar{\mu}^k}{\rho_k}, g(x^k) \right\}}{\delta_k} = \mu.$$

Entonces, para cada $\lambda_i \neq 0$ existe $k_i \in K_1$ tal que si $k \geq k_i$, $k \in K_1$,

$$h_i(x^k) \lambda_i > 0. \quad (3.3.20)$$

Además, para cada $\mu_j > 0$ existe $\tilde{k}_j \in K_1$ tal que si $k \geq \tilde{k}_j$, $k \in K_1$,

$$\max \left\{ -\frac{\bar{\mu}_j^k}{\rho_k}, g_j(x^k) \right\} \mu_j > 0.$$

Entonces, como $\bar{\mu}^k > 0$, tenemos que

$$g_j(x^k)\mu_j > 0. \tag{3.3.21}$$

Entonces, existe $\bar{k} \in K_1$, $\bar{k} \geq \max\{\max_{\{i:\lambda_i \neq 0\}} \{k_i\}, \max_{\{j:\mu_j > 0\}} \{\tilde{k}_j\}\}$ tal que (3.3.20) y (3.3.21) se satisfacen para todo $k \geq \bar{k}, k \in K_1$.

Luego, si $\theta = 0$, por (3.3.19)-(3.3.21) obtenemos una contradicción con la definición de Quasinormalidad. Entonces $\theta \neq 0$ y por (3.3.17) y el Corolario 3.1.9 x^* es un punto débil regular del problema (3.0.1). \square

En este capítulo demostramos que el método Lagrangiano Aumentado clásico para resolver el problema escalar converge a puntos *KKT* utilizando la condición de calidad de Quasinormalidad.

Esta es la primera vez que la condición de calidad de Quasinormalidad es utilizada en el contexto del Lagrangiano Aumentado.

Además, utilizando una clase de funciones de escalarización y la condición secuencial *AKKT* para el *MOP* estudiamos una aplicación del método Lagrangiano Aumentado para el problema multiobjetivo.

Demostramos que si la clase de funciones de escalarización es débilmente creciente y diferenciable con gradiente no nulo hay convergencia a un punto regular débil del *MOP* utilizando las condiciones de calidad *CCP* o Quasinormalidad. En este estudio, por el uso de una función de escalarización, los subproblemas en el método Lagrangiano Aumentado son escalares.

Capítulo 4

Método del gradiente proyectado no monótono para el problema multiobjetivo en un conjunto cerrado y convexo

En este capítulo estamos interesados en el siguiente problema de optimización multiobjetivo:

$$\begin{array}{ll} \text{Minimizar} & F(x) \\ \text{sujeto a} & x \in C \end{array} \quad (4.0.1)$$

donde $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$, $F(x) = (F_1(x), \dots, F_r(x))$ es continuamente diferenciable y $C \subset \mathbb{R}^n$ es cerrado y convexo.

Para la resolución de (4.0.1) abordamos una técnica ampliamente estudiada que es el método del gradiente proyectado *PGM* estudiado en [34, 35, 36, 41]. En los trabajos mencionados, *PGM* comparte dos características esenciales:

- son todos métodos de descenso: el valor objetivo de la función vectorial F disminuye en cada iteración componente a componente; y
- la convergencia de las sucesiones generadas por los métodos propuestos se establece bajo suposiciones razonables.

En este capítulo proponemos combinar *PGM* clásico de optimización vectorial con dos ingredientes del caso escalar:

- (i) sugerimos asociar un paso variable para calcular la dirección de búsqueda que

puede ser actualizado usando las ideas del parámetro espectral presentado por Barzilai y Borwein [9], analizado por Raydan [57] y luego utilizado en [16]; y

- (ii) la búsqueda lineal clásica de Armijo es reemplazada por una estrategia no monótona desarrollada por Zhang y Hager en [62].

El método del gradiente proyectado para diferentes opciones del tamaño del paso ha sido muy estudiado en el caso escalar, ver por ejemplo [9, 16, 17, 43] y sus referencias. Desde el punto de vista práctico, la elección espectral de la longitud del paso utilizado para calcular la dirección de búsqueda requiere poco trabajo computacional situación que lo hace más eficiente especialmente para problemas a gran escala. Desde el punto de vista teórico, se ha demostrado que el esquema acelera enormemente la convergencia de métodos basados en ideas de gradientes. Como se menciona en [17] es natural combinar la idea del gradiente espectral con una búsqueda lineal no monótona.

Las técnicas de búsqueda lineal no monótona se han estudiadas con gran éxito en varios trabajos en el caso escalar, por ejemplo [27, 43, 45, 53, 58, 62]. Proponemos usar la técnica de búsqueda lineal no monótona definida en [62] basada en el promedio de valores funcionales sucesivos. Probamos que, sin hipótesis de convexidad, los puntos de acumulación de la sucesión generada por el algoritmo propuesto son estacionarios para el problema (4.0.1).

Definición 4.0.2. [48] Decimos que un punto factible $\bar{x} \in C$ es estacionario para F si y solo si

$$JF(\bar{x})(C - \{\bar{x}\}) \cap [-\mathbb{R}_{++}^n] = \emptyset \tag{4.0.3}$$

donde $JF(\bar{x})$ es la matriz Jacobiana de F en \bar{x} ,

$$-\mathbb{R}_{++}^n = \{z \in \mathbb{R}^n : z_i < 0 \text{ para todo } i = 1, \dots, n\},$$

$$C - \{\bar{x}\} := \{z \in \mathbb{R}^n : z = x - \bar{x} : x \in C\}.$$

Claramente, si $\bar{x} \in C$ es estacionario para F entonces para todo $v \in C - \{\bar{x}\}$ se tiene que

$$JF(\bar{x})v \not\leq 0,$$

es decir, no existe una dirección factible a partir de \bar{x} que sea de descenso para todas las funciones objetivo en \bar{x} simultáneamente. En el siguiente teorema se demuestra que la condición (4.0.3) es una condición necesaria para la optimalidad Pareto.

Teorema 4.0.4. *Un punto Pareto optimal para F en C es estacionario.*

Demostración. Supongamos que $\bar{x} \in C$ es un punto Pareto optimal y que \bar{x} no es un punto estacionario para F entonces existe $v \in C - \{\bar{x}\}$ tal que

$$JF(\bar{x})v < 0.$$

Luego como F es diferenciable, para cada $i = 1, \dots, r$

$$\lim_{t \rightarrow 0} \frac{F_i(\bar{x} + tv) - F_i(\bar{x})}{t} = \nabla F_i(\bar{x})^T v < 0.$$

Luego, existe $\bar{t} > 0$ tal que para todo $t \in (0, \bar{t}]$

$$F_i(\bar{x} + tv) < F_i(\bar{x}) \text{ para todo } i = 1, \dots, r.$$

Luego, para todo $t \in [0, \text{mín}\{1, \bar{t}\}]$ se contradice que $\bar{x} \in C$ es un punto Pareto optimal. \square

En [23] se demuestra que el método de máximo descenso para el problema de minimización convexa (escalar), con tamaño del paso obtenido utilizando la regla de Armijo, es totalmente convergente a una solución (esencialmente, bajo la suposición de existencia de óptimo). Utilizando una técnica similar ampliamos este resultado al caso multiobjetivo, es decir, mostramos que para el caso convexo cualquier sucesión generada por el algoritmo propuesto converge a un punto Pareto optimal débil de (4.0.1), independientemente de cual sea el iterado inicial. La convergencia total de *PGM* para el *MOP* se estableció en [34, 35, 36, 41] con búsqueda lineal monótona de Armijo.

En [51] los autores introducen un método de máximo descenso con búsqueda no monótonas para el *MOP*. En [56] los autores proponen dos algoritmos para optimización vectorial con una función objetivo convexa basada en la búsqueda lineal no monótona [43] en el contexto escalar.

Hasta donde sabemos, esta es la primera vez que una búsqueda lineal no monótona basada en [62] es utilizada para analizar la convergencia global de *PGM* para el *MOP*.

Este capítulo está organizado de la siguiente forma: en la sección 4.1 presentamos el método del gradiente proyectado no monótono para el *MOP*. En la sección 4.2 se demuestra la convergencia global del algoritmo a puntos estacionarios. En la sección 4.3 se analiza el caso convexo.

4.1. Método del gradiente proyectado no monótono para el *MOP*

El método del gradiente proyectado que consideramos utiliza la dirección de gradiente proyectado definida como $v_\beta(x)$, para cada punto $x \in C$, por

$$v_\beta(x) = \operatorname{argmin}_{v \in C - \{x\}} \beta \varphi_x(v) + \frac{1}{2} \|v\|^2 \quad (4.1.1)$$

donde $\beta > 0$ es un parámetro fijo y

$$\varphi_x(v) = \max_{i=1,\dots,r} \{\nabla F_i(x)^T v\}. \quad (4.1.2)$$

También consideramos la función de valor óptimo $\theta_\beta : C \rightarrow \mathbb{R}$ dada por

$$\theta_\beta(x) = \beta \varphi_x(v_\beta(x)) + \frac{1}{2} \|v_\beta(x)\|^2. \quad (4.1.3)$$

La demostración de la siguiente proposición se puede ver en [36].

Proposición 4.1.4. Sean $v_\beta : C \rightarrow \mathbb{R}^n$ y $\theta_\beta : C \rightarrow \mathbb{R}$, dados por

$$v_\beta(x) = \operatorname{argmin}_{v \in C - \{x\}} \beta \varphi_x(v) + \frac{1}{2} \|v\|^2$$

y

$$\theta_\beta(x) = \beta \varphi_x(v_\beta(x)) + \frac{1}{2} \|v_\beta(x)\|^2$$

donde $\beta > 0$ es un parámetro fijo y $\varphi_x(v) = \max_{i=1,\dots,r} \{\nabla F_i(x)^T v\}$. Entonces

- (i) $\theta_\beta(x) \leq 0$ para todo $x \in C$.
- (ii) La función $\theta_\beta(\cdot)$ es continua.
- (iii) Las siguientes condiciones son equivalentes:

1. El punto $x \in C$ no es un punto estacionario.
2. $\theta_\beta(x) < 0$.
3. $v_\beta(x) \neq 0$.

En particular, x es estacionario si y solo si $\theta_\beta(x) = 0$.

Observación 4.1.5. Si $x \in C$ no es un punto estacionario existe $v \in C - \{x\}$ tal que

$$JF(x)v < 0.$$

A continuación enunciamos algunas propiedades de la función φ_x que serán utilizadas a lo largo de este capítulo.

Proposición 4.1.6. [41] La función φ_x dada por $\varphi_x(v) = \max_{i=1,\dots,r} \{\nabla F_i(x)^T v\}$ tiene las siguientes propiedades:

- (i) $\varphi_x(v)$ es continua.
- (ii) $\varphi_x(u + v) \leq \varphi_x(u) + \varphi_x(v)$ para todo $u, v \in \mathbb{R}^n$.
- (iii) $|\varphi_x(u) - \varphi_x(v)| \leq \|JF(x)u - JF(y)v\|$.

4.1.1. Búsqueda lineal no monótona

En la literatura está demostrado que, en el caso escalar, la técnica de búsqueda lineal no monótona puede mejorar la eficiencia de los métodos de descenso, [17, 27, 53, 58, 62].

La búsqueda lineal no monótona desarrollada por Grippo, Lampariello y Lucidi en [43] se basa en el máximo valor funcional considerando algunas iteraciones anteriores: dado $\sigma \in (0, 1)$, $M \geq 1$ el paso α_k debe satisfacer

$$f(x^k + \alpha_k d^k) \leq M_k + \sigma \alpha_k \nabla f(x^k)^T d^k \quad (4.1.7)$$

donde $M_k = \max_{0 \leq j \leq m(k)} f(x_{k-j})$ y $m(k) = \min\{m(k-1) + 1, M\}$.

En nuestra opinión, la extensión de (4.1.7) a problemas multiobjetivo no es inmediata y teniendo en cuenta esta observación proponemos utilizar una búsqueda lineal no monótona diferente.

En [62] los autores proponen el siguiente esquema: dados $\sigma \in (0, 1)$, $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$, $\eta_k \in [\eta_{\min}, \eta_{\max}]$ el paso α_k deben satisfacer

$$f(x^k + \alpha_k d^k) \leq C_k + \sigma \alpha_k \nabla f(x^k)^T d^k$$

donde $C_{k+1} = \frac{\eta_k Q_k C_k + f(x^{k+1})}{Q_{k+1}}$ y $Q_{k+1} = \eta_k Q_k + 1$, $C_0 = f(x^0)$ y $Q_0 = 1$.

Como se menciona en [62], si $\eta_k = 0$ para cada k , la búsqueda lineal es la búsqueda monótona habitual de Armijo. Si $\eta_k = 1$ para cada k , luego $C_k = A_k$ donde $A_k = \frac{1}{k+1} \sum_{i=0}^k f(x^i)$ es el promedio de los valores funcionales sucesivos. Por lo tanto, C_k se puede ver como una combinación de los valores funcionales $f(x^0), \dots, f(x^k)$ utilizando un parámetro η_k que controla el grado de no monotonía.

Para utilizar una búsqueda lineal no monótona en el método de descenso para el problema de optimización multiobjetivo, proponemos lo siguiente: dados $\sigma \in (0, 1)$, $C^0 = F(x^0)$, $Q_0 = 1$, $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$, $\eta_k \in [\eta_{\min}, \eta_{\max}]$ y v^k tal que $JF(x^k)v^k < 0$ el paso α_k debe satisfacer

$$F(x^k + \alpha_k v^k) \leq C^k + \sigma \alpha_k JF(x^k)v^k \quad (4.1.8)$$

donde

$$C^{k+1} = \frac{\eta_k Q_k C^k + F(x^{k+1})}{Q_{k+1}} \text{ y } Q_{k+1} = \eta_k Q_k + 1.$$

Observar que se puede escribir

$$C^{k+1} = \frac{(\eta_k Q_k + 1) C^k + F(x^{k+1}) - C^k}{Q_{k+1}} = C^k + \frac{F(x^{k+1}) - C^k}{Q_{k+1}}$$

y entonces

$$C^k - C^{k+1} = \frac{C^k - F(x^{k+1})}{Q_{k+1}}. \quad (4.1.9)$$

4.1.2. Algoritmo del gradiente proyectado multiobjetivo

Ahora podemos definir una extensión del *PGM* utilizando la búsqueda lineal no monótona (4.1.8) para el *MOP* con restricciones (4.0.1).

Algoritmo 4.1. Dados $\sigma \in (0, 1)$, $Q_0 = 1$, $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$, $\eta_0 \in [\eta_{\min}, \eta_{\max}]$, $0 < \beta_{\min} < \beta_{\max} < \infty$ y $\beta_0 \in [\beta_{\min}, \beta_{\max}]$. Sea $x^0 \in C$ un punto arbitrario. Sea $C^0 = F(x^0)$.
 Hacer $k \leftarrow 0$.

Paso 1. Calcular la dirección de búsqueda.

$$v_{\beta_k}(x^k) := \operatorname{argmin}_{v \in C - \{x^k\}} \beta_k \varphi_{x^k}(v) + \frac{\|v\|^2}{2}$$

donde

$$\varphi_{x^k}(v) = \max_{i=1, \dots, r} \{\nabla F_i(x^k)^T v\}.$$

Paso 2. Criterio de parada. Calcular

$$\theta_{\beta_k}(x^k) = \beta_k \varphi_{x^k}(v_{\beta_k}(x^k)) + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2.$$

Si $\theta_{\beta_k}(x^k) = 0$, detener el algoritmo.

Paso 3. Calcular el paso máximo. Elegir α_k como el mayor $\alpha \in \{2^{-j} : j = 0, 1, \dots\}$ tal que

$$F(x^k + \alpha v_{\beta_k}(x^k)) \leq C^k + \sigma \alpha JF(x^k) v_{\beta_k}(x^k). \quad (4.1.10)$$

Paso 4. Sea $x^{k+1} = x^k + \alpha_k v_{\beta_k}(x^k)$. Definimos β_{k+1} tal que $\beta_{k+1} \in [\beta_{\min}, \beta_{\max}]$.

Elegimos $\eta_k \in [\eta_{\min}, \eta_{\max}]$ y sean $C^{k+1} = \frac{\eta_k Q_k C^k + F(x^{k+1})}{Q_{k+1}}$ y $Q_{k+1} = \eta_k Q_k + 1$.

Hacer $k \leftarrow k + 1$, e ir al Paso 1.

Observar que por (4.1.9) y (4.1.10)

$$C^{k+1} = \frac{\eta_k Q_k C^k + F(x^{k+1})}{Q_{k+1}} \leq C^k + \frac{\sigma \alpha_k JF(x^k) v_{\beta_k}(x^k)}{Q_{k+1}} \leq C^k$$

implica que $\{C^k\}$ es una sucesión decreciente en \mathbb{R}^r .

Las diferencias entre el Algoritmo 4.1 y una extensión del *PGM* para el *MOP* con restricciones (ver por ejemplo el Algoritmo 4.2 en [36]) son:

1. la presencia de un paso variable β_k en lugar de un parámetro fijo para calcular la dirección de búsqueda en el Paso 1; y
2. el uso de la búsqueda lineal no monótona en el Paso 3.

Luego, si el Algoritmo 4.1 se detiene con el criterio de parada en el Paso 2 entonces $x^k \in C$ es un punto estacionario para F .

4.2. Análisis de convergencia

Vamos a suponer que el Algoritmo 4.1 no tiene una terminación finita y por lo tanto genera sucesiones infinitas $\{x^k\}$, $\{v_{\beta_k}(x^k)\}$ y $\{\alpha_k\}$.

Por completitud enunciaremos el Lema 1.1 de [62] adaptado para el Algoritmo 4.1.

Lema 4.2.1. *Sea $\{x^k\}$ la sucesión generada por el Algoritmo 4.1. Entonces:*

1. Si $JF(x^k)v_{\beta_k}(x^k) \leq 0$ para cada k se tiene que $F(x^k) \leq C^k \leq A^k$ para cada k , donde $A^k = \frac{1}{k+1} \sum_{i=0}^k F(x^i)$.
2. Si $JF(x^k)v_{\beta_k}(x^k) < 0$ y $F(x)$ está acotada inferiormente entonces existe α_k que satisface las condiciones de búsqueda clásica de Wolfe o de Armijo (que se corresponde con la elección de $\eta_k = 0$ en (4.1.8)).

El siguiente lema se sigue del Lema 4.2.1.

Lema 4.2.2. *Si $v_{\beta_k}(x^k) \in C - \{x^k\}$ y $JF(x^k)v_{\beta_k}(x^k) < 0$ para cada k entonces $F(x^k) \leq C^k \leq A^k$ para cada k , donde $A^k = \frac{1}{k+1} \sum_{i=0}^k F(x^i)$.*

Demostración. Por definición $F(x^0) = C^0 = A^0$ por lo tanto valen las desigualdades. Para cada $k \geq 1$ definamos $D^k : \mathbb{R} \rightarrow \mathbb{R}^r$ como

$$D^k(t) = \frac{tC^{k-1} + F(x^k)}{t+1}.$$

Tenemos que $(D^k)' : \mathbb{R} \rightarrow \mathbb{R}^r$ se calcula como

$$(D^k)'(t) = \frac{C^{k-1} - F(x^k)}{(t+1)^2}.$$

De (4.1.10) como $JF(x^k)v_{\beta_k}(x^k) < 0$ se obtiene que $F(x^k) \leq C^{k-1}$ lo que implica que cada componente de $(D^k)'(t)$ es mayor o igual a 0 para todo $t > 0$. Luego $F(x^k) = D^k(0) \leq D^k(t)$ para todo $t > 0$. En particular tomando $t = \eta_{k-1}Q_{k-1}$,

$$F(x^k) = D^k(0) \leq D^k(\eta_{k-1}Q_{k-1}) = C^k.$$

Y queda así demostrado que para todo k , $F(x^k) \leq C^k$.

Probamos ahora la desigualdad $C^k \leq A^k$ por inducción.

Para $k = 0$, esto se cumple $C^0 = F(x^0)$. Ahora supongamos que $C^j \leq A^j$ para todos los $0 \leq j < k$. Como $Q_{j+1} = \eta_j Q_j + 1$, $Q_0 = 1$ y $\eta_k \in [0, 1]$ tenemos que

$$Q_{j+1} = 1 + \sum_{i=0}^j \prod_{m=0}^i \eta_{j-m} \leq j + 2. \quad (4.2.3)$$

Como cada componente de D^k es monótona y no decreciente, (4.2.3) implica que $C^k = D^k(\eta_{k-1}Q_{k-1}) = D^k(Q_k - 1) \leq D^k(k)$. Por el paso de inducción,

$$D^k(k) = \frac{kC^{k-1} + F(x^k)}{k+1} \leq \frac{kA^{k-1} + F(x^k)}{k+1} = A^k.$$

Por lo tanto $F(x^k) \leq C^k \leq A^k$ para cada k . □

Vamos a enunciar el Lema 2.1 de [36] que será usado para la demostración del Lema 4.2.5 en el cual probamos que el Algoritmo 4.1 está bien definido.

Lema 4.2.4. (Lema 2.1 de [36]) Sea $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$ una función continuamente diferenciable, $v \in \mathbb{R}^n$ tal que $JF(x)v < 0$ y $\sigma \in (0, 1)$. Entonces existe $\epsilon > 0$ (que puede depender de x, v y σ) tal que

$$F(x + tv) < F(x) + \sigma t JF(x)v$$

para cualquier $t \in (0, \epsilon)$. En particular, v es una dirección de descenso para F en x .

Lema 4.2.5. Sea $\{x^k\} \subset \mathbb{R}^n$ una sucesión generada por el Algoritmo 4.1. Entonces:

- (a) $\{x^k\} \subset C$.
- (b) El Algoritmo 4.1 está bien definido.

Demostración.

(a) Hacemos la demostración por inducción sobre k . Por el Algoritmo 4.1 sabemos que $x^0 \in C$. Supongamos que $x^k \in C$. Observemos que, por los Pasos 1 y 3 del

algoritmo, $v_{\beta_k}(x^k) \in C - \{x^k\}$ y $\alpha_k \in (0, 1]$. Entonces, para algún $z^k \in C$, $v_{\beta_k}(x^k) = z^k - x^k$ y por la convexidad de C se deduce que

$$x^{k+1} = x^k + \alpha_k v_{\beta_k}(x^k) = (1 - \alpha_k)x^k + \alpha_k z^k \in C$$

como queríamos demostrar.

(b) Por el Lema 4.2.4 existe $\epsilon > 0$ tal que para cualquier $\alpha \in (0, \epsilon)$

$$F(x^k + \alpha v_{\beta_k}(x^k)) < F(x^k) + \sigma \alpha JF(x^k) v_{\beta_k}(x^k).$$

Luego como $F(x^k) \leq C^k$ por el Lema 4.2.2 tenemos que

$$F(x^k + \alpha v_{\beta_k}(x^k)) < C^k + \sigma \alpha JF(x^k) v_{\beta_k}(x^k).$$

Por lo tanto existe $\alpha \in (0, \epsilon)$ satisfaciendo la condición (4.1.10). \square

Lema 4.2.6. *La sucesión $\{\|v_{\beta_k}(x^k)\|\}$ es acotada.*

Demostración. Por la Proposición 4.1.4 (i) y el Lema 4.2.5

$$\beta_k \varphi_{x^k}(v_{\beta_k}(x^k)) + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2 = \theta_{\beta_k}(x^k) \leq 0.$$

Entonces, por la desigualdad de Cauchy Schwarz, tenemos que para todo $i = 1, \dots, r$

$$-\beta_k \|\nabla F_i(x^k)\| \|v_{\beta_k}(x^k)\| + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2 \leq \theta_{\beta_k}(x^k) \leq 0.$$

Luego

$$\|v_{\beta_k}(x^k)\| \left(-\beta_k \|\nabla F_i(x^k)\| + \frac{1}{2} \|v_{\beta_k}(x^k)\| \right) \leq 0$$

y se tiene que

$$\|v_{\beta_k}(x^k)\| \leq 2\beta_k \|\nabla F_i(x^k)\|.$$

Por lo tanto, como F es continuamente diferenciable y $\{\beta_k\}$ es acotada, por el Lema 4.2.5 (a) obtenemos que $\{\|v_{\beta_k}(x^k)\|\}$ es una sucesión acotada. \square

Observar que, de forma análoga, para cada $x \in C$, la sucesión escalar $\{\|v_{\beta_k}(x)\|\}$ es acotada.

Proposición 4.2.7. *Sea $\{x^k\} \subset \mathbb{R}^n$ una sucesión generada por el Algoritmo 4.1. Entonces existen $\bar{K} \subset \mathbb{N}$, $\beta > 0$ y $\bar{x} \in C$ tales que*

$$\liminf_{k \in \bar{K}} \theta_{\beta_k}(x^k) = \theta_{\bar{\beta}}(\bar{x}) \quad (4.2.8)$$

donde $\lim_{k \in \bar{K}} x^k = \bar{x}$, $\lim_{k \in \bar{K}} \beta^k = \bar{\beta}$.

Demostración. Como $\{x^k\} \subset C$, C es cerrado y $\{\beta^k\}$ es acotada existen $K \subset \mathbb{N}$, $\bar{\beta} \in \mathbb{R}$ y $\bar{x} \in C$ tales que $\lim_{k \in K} x^k = \bar{x}$, $\lim_{k \in K} \beta^k = \bar{\beta}$. Además, $\bar{\beta} > 0$ ya que $\beta_k \in [\beta_{\min}, \beta_{\max}]$ con $\beta_{\min} > 0$.

Primero demostremos que existe $\bar{K} \subset K$ tal que

$$\lim_{k \in \bar{K}} \theta_{\beta_k}(\bar{x}) = \theta_{\bar{\beta}}(\bar{x}). \quad (4.2.9)$$

Como $v_{\beta_k}(\bar{x}) \in C - \{\bar{x}\}$, a partir de la definición de $v_{\bar{\beta}}(\bar{x})$ tenemos que

$$\theta_{\bar{\beta}}(\bar{x}) \leq \bar{\beta} \varphi_{\bar{x}}(v_{\beta_k}(\bar{x})) + \frac{1}{2} \|v_{\beta_k}(\bar{x})\|^2 = (\bar{\beta} - \beta_k) \varphi_{\bar{x}}(v_{\beta_k}(\bar{x})) + \theta_{\beta_k}(\bar{x}).$$

Luego, usando la observación que sigue al Lema 4.2.6 (la sucesión $\{\|v_{\beta_k}(\bar{x})\|\}$ es acotada) y la Proposición 4.1.6 (i), tomando límite inferior en una subsucesión apropiada $\bar{K} \subset K$ cuando $k \in \bar{K}$ tendiendo a infinito obtenemos que

$$\theta_{\bar{\beta}}(\bar{x}) \leq \liminf_{k \in \bar{K}} \theta_{\beta_k}(\bar{x}). \quad (4.2.10)$$

Ahora, de la definición de $v_{\beta_k}(\bar{x})$, ya que $v_{\bar{\beta}}(\bar{x}) \in C - \{\bar{x}\}$ tenemos que $\theta_{\beta_k}(\bar{x}) \leq \beta_k \varphi_{\bar{x}}(v_{\bar{\beta}}(\bar{x})) + \frac{1}{2} \|v_{\bar{\beta}}(\bar{x})\|^2$.

Luego, tomando límite superior para $k \in \bar{K}$ cuando k tiende a infinito obtenemos

$$\limsup_{k \in \bar{K}} \theta_{\beta_k}(\bar{x}) \leq \theta_{\bar{\beta}}(\bar{x})$$

que, junto con la desigualdad (4.2.10) demuestra (4.2.9).

En segundo lugar, como $v_{\bar{\beta}}(\bar{x}) \in C - \{\bar{x}\}$ tenemos que $v_{\bar{\beta}}(\bar{x}) + \bar{x} - x^k \in C - \{x^k\}$ y de la definición de $v_{\beta_k}(x^k)$

$$\theta_{\beta_k}(x^k) = \beta_k \varphi_{x^k}(v_{\beta_k}(x^k)) + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2 \leq \beta_k \varphi_{x^k}(v_{\bar{\beta}}(\bar{x}) + \bar{x} - x^k) + \frac{1}{2} \|v_{\bar{\beta}}(\bar{x}) + \bar{x} - x^k\|^2.$$

Por lo tanto, por la Proposición 4.1.6 (ii) obtenemos

$$\theta_{\beta_k}(x^k) \leq \beta_k \varphi_{x^k}(v_{\bar{\beta}}(\bar{x})) + \beta_k \varphi_{x^k}(\bar{x} - x^k) + \frac{1}{2} \|v_{\bar{\beta}}(\bar{x})\|^2 + \frac{1}{2} \|\bar{x} - x^k\|^2 + \langle v_{\bar{\beta}}(\bar{x}), \bar{x} - x^k \rangle.$$

Por la Proposición 4.1.6 (i) y la propiedad de que F es continuamente diferenciable, tomando límite superior para $k \in \bar{K}$ en la desigualdad anterior tenemos

$$\limsup_{k \in \bar{K}} \theta_{\beta_k}(x^k) \leq \bar{\beta} \varphi_{\bar{x}}(v_{\bar{\beta}}(\bar{x})) + \frac{1}{2} \|v_{\bar{\beta}}(\bar{x})\|^2 = \theta_{\bar{\beta}}(\bar{x}). \quad (4.2.11)$$

Como $v_{\beta_k}(x^k) \in C - \{x^k\}$ tenemos que $v_{\beta_k}(x^k) + x^k - \bar{x} \in C - \{\bar{x}\}$ y

$$\theta_{\beta_k}(\bar{x}) \leq \beta_k \varphi_{\bar{x}}(v_{\beta_k}(x^k) + x^k - \bar{x}) + \frac{1}{2} \|v_{\beta_k}(x^k) + x^k - \bar{x}\|^2 \leq$$

$$\beta_k \varphi_{\bar{x}}(v_{\beta_k}(x^k)) + \beta_k \varphi_{\bar{x}}(x^k - \bar{x}) + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2 + \frac{1}{2} \|x^k - \bar{x}\|^2 + \langle v_{\beta_k}(x^k), x^k - \bar{x} \rangle.$$

Entonces, por el Lema 4.2.6 y la Proposición 4.1.6 (i), tomando límite inferior para $k \in \bar{K}$ en ambos lados de la desigualdad, tenemos, de (4.2.9),

$$\begin{aligned} \theta_{\bar{\beta}}(\bar{x}) &\leq \liminf_{k \in \bar{K}} [\beta_k \varphi_{\bar{x}}(v_{\beta_k}(x^k)) + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2 + \beta_k \varphi_{x^k}(v_{\beta_k}(x^k)) - \beta_k \varphi_{x^k}(v_{\beta_k}(x^k))] \\ &= \liminf_{k \in \bar{K}} [\theta_{\beta_k}(x^k) + \beta_k (\varphi_{\bar{x}}(v_{\beta_k}(x^k)) - \varphi_{x^k}(v_{\beta_k}(x^k)))]. \end{aligned}$$

Ahora, usando la Proposición 4.1.6 (iii), obtenemos que

$$\theta_{\bar{\beta}}(\bar{x}) \leq \liminf_{k \in \bar{K}} [\theta_{\beta_k}(x^k) + \beta_k \|JF(\bar{x}) - JF(x^k)\| \|v_{\beta_k}(x^k)\|] \leq \liminf_{k \in \bar{K}} \theta_{\beta_k}(x^k) \quad (4.2.12)$$

donde usamos el Lema 4.2.6 y que F es continuamente diferenciable en la última desigualdad. Luego, combinando (4.2.11) con (4.2.12) se obtiene que

$$\limsup_{k \in \bar{K}} \theta_{\beta_k}(x^k) \leq \theta_{\bar{\beta}}(\bar{x}) \leq \liminf_{k \in \bar{K}} \theta_{\beta_k}(x^k)$$

que implica (4.2.8). \square

El siguiente teorema es el resultado principal de esta sección: establece la estacionariedad de los puntos de acumulación de las sucesiones generadas por el Algoritmo 4.1.

Teorema 4.2.13. *Supongamos que F es acotada inferiormente. Sea $\{x^k\} \subset \mathbb{R}^n$ una sucesión generada por el Algoritmo 4.1. Entonces cada punto de acumulación de $\{x^k\}$ es un punto estacionario factible de (4.0.1).*

Demostración. Sea \bar{x} un punto de acumulación de la sucesión $\{x^k\}$. Entonces existe $K \subset \mathbb{N}$ tal que $\lim_{k \in K} x^k = \bar{x}$. La factibilidad de \bar{x} se obtiene combinando que C es cerrado con el Lema 4.2.5.

La sucesión $\{\beta_k\}_{k \in K}$ es acotada, entonces existen $K_0 \subset K$ y $\bar{\beta} > 0$ tales que $\lim_{k \in K_0} \beta_k = \bar{\beta}$.

Considerando que $\alpha_k \in (0, 1]$ para todo $k \in K_0$ tenemos las siguientes dos posibilidades: (a) $\limsup_{k \in K_0} \alpha_k > 0$, o (b) $\limsup_{k \in K_0} \alpha_k = 0$.

Supongamos que vale (a). Entonces existen $K_1 \subset K_0$ y $\bar{\alpha} > 0$ tales que $\lim_{k \in K_1} \alpha_k = \bar{\alpha}$. Como $\bar{\alpha} > 0$ existe $K_2, K_2 \subset K_1$ tal que $\alpha_k \geq \epsilon > 0$ para todo $k \in K_2$. Además, para todo $k \in K_2$, por (4.1.10) y $JF(x^k)v_{\beta_k}(x^k) < 0$, tenemos que

$$F(x^{k+1}) \leq C^k + \sigma \alpha_k JF(x^k)v_{\beta_k}(x^k) \leq C^k + \sigma \epsilon JF(x^k)v_{\beta_k}(x^k). \quad (4.2.14)$$

Entonces

$$-\sigma\epsilon JF(x^k)v_{\beta_k}(x^k) \leq C^k - F(x^{k+1}). \quad (4.2.15)$$

Ahora, usando (4.1.9) y (4.2.15) obtenemos

$$C^k - C^{k+1} \geq \sigma\epsilon \frac{-JF(x^k)v_{\beta_k}(x^k)}{Q_{k+1}}. \quad (4.2.16)$$

Entonces, como $\sum_{k=0}^m C^k - C^{k+1} = C^0 - C^{m+1}$ y F es acotada inferiormente tenemos que, por el Lema 4.2.2, $R \leq F(x^k) \leq C^k$, para todo k y podemos concluir que C^k es acotada inferiormente. Por (4.2.16)

$$\sum_{k=0}^{\infty} \frac{\sigma\epsilon(-JF(x^k)v_{\beta_k}(x^k))}{Q_{k+1}} < \infty. \quad (4.2.17)$$

Si suponemos que $\liminf_{k \in K_2} (-JF(x^k)v_{\beta_k}(x^k)) \neq 0$, entonces

$$\liminf_{k \in K_2} (-JF(x^k)v_{\beta_k}(x^k)) \geq c > 0, \text{ para alguna constante } c.$$

Además, como $Q_{k+1} = \eta_k Q_k + 1$ tenemos que $Q_{k+1} = 1 + \sum_{i=0}^k \prod_{m=0}^i \eta_{k-m} \leq k + 2$ y por lo tanto se obtiene que

$$\sum_{k=0}^{\infty} \frac{(-JF(x^k)v_{\beta_k}(x^k))}{Q_{k+1}} \geq \sum_{k=0}^{\infty} \frac{c}{k+2},$$

en contradicción con (4.2.17).

Por lo tanto $\liminf_{k \in K_2} (-JF(x^k)v_{\beta_k}(x^k)) = 0$, con lo cual, componente a componente, podemos escribir

$$\liminf_{k \in K_2} (-\nabla F_i(x^k)^T v_{\beta_k}(x^k)) = 0, \text{ para todo } i = 1, \dots, r. \quad (4.2.18)$$

Luego, combinando (4.2.18) con (4.2.8) existe $\bar{K} \subset K_2$ tal que

$$0 = \liminf_{k \in \bar{K}} \beta_k \nabla F_i(x^k)^T v_{\beta_k}(x^k) \leq \liminf_{k \in \bar{K}} \beta_k \varphi_{x^k}(v_{\beta_k}(x^k)) + \frac{1}{2} \|v_{\beta_k}(x^k)\|^2 =$$

$$\liminf_{k \in \bar{K}} \theta_{\beta_k}(x^k) = \theta_{\bar{\beta}}(\bar{x})$$

y concluimos que $\theta_{\bar{\beta}}(\bar{x}) = 0$. Entonces, por la Proposición 4.1.4, \bar{x} es un punto estacionario de (4.0.1).

Supongamos que (b) se satisface. Debido al Lema 4.2.6, existen $K_1 \subset K_0$ y $\bar{v} \in \mathbb{R}^n$ tales que $\lim_{k \in K_1} v_{\beta_k}(x^k) = \bar{v}$ y $\lim_{k \in K_1} \alpha_k = 0$. Recordando que

$$\max_{i=1, \dots, r} \{\beta_k \nabla F_i(x^k)^T v_{\beta_k}(x^k)\} \leq \theta_{\beta_k}(x^k) \leq 0$$

entonces tomando límite cuando $k \in K_1$ tiende a infinito tenemos

$$\max_{i=1, \dots, r} \{\bar{\beta} \nabla F_i(\bar{x})^T \bar{v}\} \leq \theta_{\bar{\beta}}(\bar{x}) \leq 0. \quad (4.2.19)$$

Sea q un entero positivo fijo arbitrario. Como $\alpha_k \rightarrow 0$, para $k \in K_1$ existe un $k_1 \in K_1$ tal que para todo $k \in K_1$ $k \geq k_1$ tenemos $\alpha_k < 2^{-q}$, lo que significa que la condición no monótona en el Paso 3 del Algoritmo 4.1 no se satisface en x^k para $\alpha = 2^{-q}$:

$$F(x^k + 2^{-q} v_{\beta_k}(x^k)) \not\leq C^k + \sigma 2^{-q} JF(x^k) v_{\beta_k}(x^k).$$

Entonces para cada $k \in K_1$, $k \geq k_1$ existe $i = i(k) \in \{1, \dots, r\}$ tal que

$$F_i(x^k + 2^{-q} v_{\beta_k}(x^k)) > C_i^k + \sigma 2^{-q} \nabla F_i(x^k)^T v_{\beta_k}(x^k)$$

donde $C_i^k = \frac{\eta_{k-1} Q_{k-1} C_i^{k-1} + F_i(x^k)}{Q_k}$.

Como $\{i(k)\}_{k \in K_1} \subset \{1, \dots, r\}$, existe $K_2 \subset K_1$ y un índice i_0 tal que $i_0 = i(k)$ para todo $k \in K_2$, $k \geq k_1$

$$F_{i_0}(x^k + 2^{-q} v_{\beta_k}(x^k)) > C_{i_0}^k + \sigma 2^{-q} \nabla F_{i_0}(x^k)^T v_{\beta_k}(x^k)$$

y, por el Lema 4.2.2, $F(x^k) \leq C^k$ tenemos

$$F_{i_0}(x^k + 2^{-q} v_{\beta_k}(x^k)) > F_{i_0}(x^k) + \sigma 2^{-q} \nabla F_{i_0}(x^k)^T v_{\beta_k}(x^k).$$

Tomando límite cuando $k \in K_2$ en la desigualdad anterior, obtenemos

$$F_{i_0}(\bar{x} + 2^{-q} \bar{v}) \geq F_{i_0}(\bar{x}) + \sigma 2^{-q} \nabla F_{i_0}(\bar{x})^T \bar{v}.$$

Dado que esta desigualdad se cumple para cualquier entero positivo q y para i_0 (dependiendo de q), por el Lema 4.2.4 tenemos que $JF(\bar{x})\bar{v} \not\leq 0$. Entonces

$$\max_{i=1, \dots, r} \{\nabla F_i(\bar{x})^T \bar{v}\} \geq 0,$$

que, junto con (4.2.19) implica que $\theta_{\bar{\beta}}(\bar{x}) = 0$.

Luego concluimos que \bar{x} es un punto estacionario de (4.0.1). \square

4.3. Caso convexo

El objetivo de esta sección es demostrar, bajo hipótesis de convexidad sobre la función F , que en el caso que el Algoritmo 4.1 no tenga una terminación finita, la sucesión $\{x^k\}$ converge a un punto Pareto optimal débil.

Decimos que la función $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$ es \mathbb{R}_+^r -convexa si

$$F(\lambda x + (1 - \lambda)z) \leq \lambda F(x) + (1 - \lambda)F(z)$$

para todo $x, z \in \mathbb{R}^n$ y todo $\lambda \in [0, 1]$. Claramente, $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$ es \mathbb{R}_+^r -convexa si y solo si cada componente $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$ es convexa.

La prueba de la convergencia total a un punto Pareto optimal débil se desprende del análisis realizado en [36] y referencias en el mismo. De la teoría del análisis y optimización convexa, sabemos que de (4.1.1), existe $w(x) \in \mathbb{R}^r$ con

$$w(x) \geq 0, \quad \sum_{i=1}^r w_i(x) = 1 \quad (4.3.1)$$

tal que $v_\beta(x) = P_C(x - \beta JF(x)^T w(x)) - x$, donde $P_C(z)$ es la proyección ortogonal de z en C , satisfaciendo

$$\left\langle \beta \sum_{i=1}^r w_i(x) \nabla F_i(x) + v_\beta(x), v - v_\beta(x) \right\rangle \geq 0 \quad (4.3.2)$$

para todo $v \in C - \{x\}$. Ver [12, 36].

Usaremos el siguiente resultado técnico para probar la convergencia total de la sucesión $\{x^k\}$ generada por el Algoritmo 4.1.

Lema 4.3.3. *Supongamos que F es una función \mathbb{R}_+^r -convexa y sea $\{x^k\}$ una sucesión generada por el Algoritmo 4.1. Sea $w^k := w(x^k) \in \mathbb{R}^r$ tal que (4.3.1) se verifica y*

$$v_{\beta_k}^k := v_{\beta_k}(x^k) = P_C(x^k - \beta_k JF(x^k)^T w^k) - x^k.$$

Si para $x^* \in C$ y $k \geq 0$ tenemos que $F(x^*) \leq C^k$, entonces

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + 2\beta_{\max} \alpha_k |\langle w^k, JF(x^k) v_{\beta_k}^k \rangle| + 2\beta_{\max} \alpha_k \langle w^k, C^k - F(x^k) \rangle. \quad (4.3.4)$$

Demostración. Como $x^{k+1} = x^k + \alpha_k v_{\beta_k}^k$, tenemos que

$$\|x^{k+1} - x^*\|^2 = \|x^k - x^* + \alpha_k v_{\beta_k}^k\|^2 = \|x^k - x^*\|^2 + \alpha_k^2 \|v_{\beta_k}^k\|^2 - 2\alpha_k \langle v_{\beta_k}^k, x^* - x^k \rangle. \quad (4.3.5)$$

Usando (4.3.2) con $x = x^k$ y $\beta = \beta_k$ obtenemos

$$\langle \beta_k JF(x^k)^T w^k + v_{\beta_k}^k, v - v_{\beta_k}^k \rangle \geq 0 \text{ para todo } v \in C - \{x^k\}.$$

Tomando $v = x^* - x^k \in C - \{x^k\}$ en la desigualdad anterior, obtenemos

$$-\langle v_{\beta_k}^k, x^* - x^k \rangle \leq \beta_k \langle w^k, JF(x^k)(x^* - x^k) \rangle - \beta_k \langle w^k, JF(x^k)v_{\beta_k}^k \rangle - \|v_{\beta_k}^k\|^2. \quad (4.3.6)$$

Por la convexidad de F , $JF(x^k)(x^* - x^k) \leq F(x^*) - F(x^k)$. Luego, este hecho junto con $w^k \geq 0$ y $F(x^*) \leq C^k$, implica

$$\langle w^k, JF(x^k)(x^* - x^k) \rangle \leq \langle w^k, F(x^*) - F(x^k) \rangle \leq \langle w^k, C^k - F(x^k) \rangle.$$

Además, como x^k no es un punto estacionario, tenemos que $JF(x^k)v_{\beta_k}^k < 0$. Entonces, $\langle w^k, JF(x^k)v_{\beta_k}^k \rangle < 0$ y de (4.3.6) se tiene que

$$\begin{aligned} -\langle v_{\beta_k}^k, x^* - x^k \rangle &\leq \beta_k \langle w^k, C^k - F(x^k) \rangle + \beta_k |\langle w^k, JF(x^k)v_{\beta_k}^k \rangle| - \|v_{\beta_k}^k\|^2 \\ &\leq \beta_{\max} \langle w^k, C^k - F(x^k) \rangle + \beta_{\max} |\langle w^k, JF(x^k)v_{\beta_k}^k \rangle| - \|v_{\beta_k}^k\|^2. \end{aligned}$$

Entonces, de (4.3.5),

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 + \alpha_k^2 \|v_{\beta_k}^k\|^2 + 2\alpha_k \beta_k \langle w^k, JF(x^k)(x^* - x^k) \rangle + \\ &\quad 2\alpha_k \beta_k |\langle w^k, JF(x^k)v_{\beta_k}^k \rangle| - 2\alpha_k \|v_{\beta_k}^k\|^2 \\ &\leq \|x^k - x^*\|^2 + 2\alpha_k \beta_{\max} \langle w^k, C^k - F(x^k) \rangle + \\ &\quad 2\alpha_k \beta_{\max} |\langle w^k, JF(x^k)v_{\beta_k}^k \rangle| + (\alpha_k^2 - 2\alpha_k) \|v_{\beta_k}^k\|^2. \end{aligned}$$

Y el resultado se obtiene porque $\alpha_k^2 - 2\alpha_k \leq 0$ pues $0 < \alpha_k \leq 1$. \square

El siguiente teorema se establece con el fin de demostrar la convergencia total de la sucesión.

Teorema 4.3.7. *Supongamos que $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$ es una función \mathbb{R}_+^r -convexa. Supongamos que $\{x^k\}$ es una sucesión generada por el Algoritmo 4.1 considerando $\eta_{\max} < 1$, $v_{\beta_k}(x^k)$ definido como en el Lema 4.3.3 y*

$$L := \{x \in C : F(x) \leq C^k \text{ para todo } k\}. \quad (4.3.8)$$

Supongamos que $L \neq \emptyset$. Entonces $\{x^k\}$ es una sucesión Quasi-Féjer convergente a L , lo que significa que para cada $x^ \in L$ existe una sucesión $\{\varepsilon_k\} \subset \mathbb{R}$, $\varepsilon_k \geq 0$ tal que para todo k*

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + \varepsilon_k$$

con $\sum_{k=0}^{\infty} \varepsilon_k < \infty$.

Demostración. Por hipótesis existe $x^* \in L$. Como F es una función \mathbb{R}_+^r -convexa por el Lema 4.3.3, para todo k

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + 2\beta_{\max}\alpha_k |\langle w^k, JF(x^k)v_{\beta_k}^k \rangle| + 2\beta_{\max}\alpha_k \langle w^k, C^k - F(x^k) \rangle. \quad (4.3.9)$$

Considerando la base canónica de \mathbb{R}^r , $\{e^1, \dots, e^r\}$, podemos escribir para cada k , $w^k = \sum_{i=1}^r w_i^k e^i$.

Entonces, de (4.3.9) y $0 \leq w_i^k \leq 1$ para todo i y k y $C^k \geq F(x^k)$, tenemos que para todo k

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + 2\beta_{\max}\alpha_k \sum_{i=1}^r \left(|\langle e^i, JF(x^k)v_{\beta_k}^k \rangle| + \langle e^i, C^k - F(x^k) \rangle \right).$$

Definiendo $\varepsilon_k := 2\beta_{\max}\alpha_k \sum_{i=1}^r \left(|\langle e^i, JF(x^k)v_{\beta_k}^k \rangle| + \langle e^i, C^k - F(x^k) \rangle \right)$ tenemos que $\varepsilon_k \geq 0$ y

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + \varepsilon_k.$$

Demostremos ahora que $\sum_{k=0}^{\infty} \varepsilon_k < \infty$.

Definamos $\varepsilon_k^1 := 2\beta_{\max}\alpha_k \sum_{i=1}^r |\langle e^i, JF(x^k)v_{\beta_k}^k \rangle|$ y $\varepsilon_k^2 := 2\beta_{\max}\alpha_k \sum_{i=1}^r \langle e^i, C^k - F(x^k) \rangle$.

De la búsqueda lineal en el Paso 3, obtenemos para $i = 1, \dots, r$

$$-\alpha_k \langle e^i, JF(x^k)v_{\beta_k}^k \rangle \leq \frac{1}{\sigma} (C_i^k - F_i(x^{k+1})).$$

Como x^k no es estacionario, también tenemos que para todo i y k

$$\langle e^i, JF(x^k)v_{\beta_k}^k \rangle = \nabla F_i(x^k)^T v_{\beta_k}^k < 0,$$

lo que significa que

$$-\alpha_k \langle e^i, JF(x^k)v_{\beta_k}^k \rangle = \alpha_k |\langle e^i, JF(x^k)v_{\beta_k}^k \rangle|.$$

Por lo tanto obtenemos que

$$\varepsilon_k^1 \leq \frac{2\beta_{\max}}{\sigma} \sum_{i=1}^r (C_i^k - F_i(x^{k+1}))$$

y por (4.1.9)

$$\varepsilon_k^1 \leq \frac{2\beta_{\max}}{\sigma} \sum_{i=1}^r Q_{k+1}(C_i^k - C_i^{k+1}).$$

Como $Q_{k+1} = 1 + \sum_{i=0}^k \prod_{m=0}^i \eta_{k-m} \leq 1 + \sum_{i=0}^k \eta_{\text{máx}}^{i+1} \leq \sum_{i=0}^{\infty} \eta_{\text{máx}}^i \leq \frac{1}{1 - \eta_{\text{máx}}}$ obtenemos que

$$0 \leq \varepsilon_k^1 \leq \frac{2\beta_{\text{máx}}}{\sigma(1 - \eta_{\text{máx}})} \sum_{i=1}^r (C_i^k - C_i^{k+1}). \quad (4.3.10)$$

Por otro lado, por (4.1.9), $C_i^k = C_i^{k-1} + \frac{F_i(x^k) - C_i^{k-1}}{Q_k}$ para todo $k \geq 1$ y tenemos que $C_i^k - F_i(x^k) = (1 - \frac{1}{Q_k})(C_i^{k-1} - F_i(x^k))$. Entonces

$$\varepsilon_k^2 = 2\beta_{\text{máx}}\alpha_k \frac{Q_k - 1}{Q_k} \sum_{i=1}^r (C_i^{k-1} - F_i(x^k))$$

y usando (4.1.9)

$$\varepsilon_k^2 = 2\beta_{\text{máx}}\alpha_k(Q_k - 1) \sum_{i=1}^r (C_i^{k-1} - C_i^k).$$

Como $Q_k - 1 \leq \frac{\eta_{\text{máx}}}{1 - \eta_{\text{máx}}}$ y $\alpha_k \leq 1$ obtenemos

$$\varepsilon_k^2 \leq \frac{2\beta_{\text{máx}}\eta_{\text{máx}}}{(1 - \eta_{\text{máx}})} \sum_{i=1}^r (C_i^{k-1} - C_i^k). \quad (4.3.11)$$

Sumando de $k = 0$ a $k = N$ en las desigualdades (4.3.10) y (4.3.11) donde N es cualquier número entero positivo y definiendo $C^{-1} = C^0$, tenemos

$$\sum_{k=0}^N \varepsilon_k \leq \frac{2\beta_{\text{máx}}}{\sigma(1 - \eta_{\text{máx}})} \sum_{i=1}^r (C_i^0 - C_i^{N+1}) + \frac{2\beta_{\text{máx}}\eta_{\text{máx}}}{(1 - \eta_{\text{máx}})} \sum_{i=1}^r (C_i^0 - C_i^N).$$

Como N es un entero positivo arbitrario, $\beta_{\text{máx}} > 0$, $\sigma > 0$, $\eta_{\text{máx}} < 1$ y $F(x^*) \leq C^k$ para todo k , concluimos que $\sum_{k=0}^{\infty} \varepsilon_k < \infty$ como queríamos probar. \square

El teorema anterior formula las condiciones bajo las cuales la sucesión generada por el Algoritmo no monótono 4.1 es una sucesión Quasi-Féjer. Haremos a continuación algunos comentarios sobre la suposición $L \neq \emptyset$. Cuando la búsqueda lineal de Armijo se considera la suposición utilizada es la siguiente [35, 36]:

Suposición. Cada sucesión \mathbb{R}_+ -decreciente $\{y^k\} \subset F(C) := \{F(x) : x \in C\}$ es \mathbb{R}_+ -acotada inferiormente por un elemento de $F(C)$, es decir, para cada $\{y^k\}$ contenida en $F(C)$ con $y^{k+1} < y^k$ para todo k , existe $\bar{x} \in C$ tal que $F(\bar{x}) \leq y^k$ para todo k .

En el caso de optimización escalar sin restricciones (convexa), esta condición, conocida como \mathbb{R}_+^r -completo, es equivalente a la existencia de soluciones del problema de optimización y sirve para asegurar la existencia de puntos eficientes o Pareto optimales para problemas de optimización vectorial.

Por lo tanto, en el caso monótono [35, 36], como la sucesión $\{F(x^k)\}$ es \mathbb{R}_+^r -decreciente, por la suposición, existe un punto $\hat{x} \in \{x \in C : F(x) \leq F(x^k)\}$ y se prueba un resultado análogo al Lema 4.3.3 con la desigualdad

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + 2\beta_{\max}\alpha_k |\langle w^k, JF(x^k)v_{\beta_k}^k \rangle|,$$

en lugar de (4.3.4).

En el caso no monótono, consideramos el conjunto L definido por (4.3.8) considerando C^k en lugar de $F(x^k)$ ya que $\{C^k\}$ es una sucesión \mathbb{R}_+^r -decreciente. Es por eso que aparece el último término en (4.3.4) y ε_k^2 en la prueba del Teorema 4.3.7.

Observemos finalmente que la suposición $L \neq \emptyset$ se verifica cuando F es un mapeo lineal o no lineal pero con $F(C)$ conjunto convexo o incluso cuando el problema original tiene una solución.

Vamos a enunciar por completitud el Lema 3.5 de [36] y el Teorema 3.10 de [36], que serán utilizados para la demostración el próximo teorema.

Lema 4.3.12. *(Lema 3.5 de [36]) Sea x^* un punto Pareto optimal débil para F . Entonces x^* es un punto estacionario. Si, además, F es \mathbb{R}_+^r -convexa, esta condición es suficiente para un Pareto optimal débil.*

Entonces, bajo \mathbb{R}_+^r -convexidad, los conceptos de eficiencia débil y de punto estacionario son equivalentes.

Teorema 4.3.13. *(Teorema 3.10 de [36]) Si una sucesión $\{u^k\}$ es Quasi-Féjer convergente a un conjunto no vacío $U \subset \mathbb{R}^n$, entonces $\{u^k\}$ está acotado. Si, además, $\{u^k\}$ tiene un punto límite u que pertenece a U , entonces $\lim_{k \rightarrow \infty} u^k = u$.*

Teorema 4.3.14. *Supongamos que $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$ es \mathbb{R}_+^r -convexo y las hipótesis del Teorema 4.3.7 se satisfacen. Entonces, cualquier sucesión $\{x^k\}$ generada por el Algoritmo 4.1 converge a un punto Pareto optimal débil.*

Demostración. Por el Teorema 4.3.7, como $\{x^k\}$ es una sucesión Quasi-Féjer convergente a L , por el Teorema 4.3.13, $\{x^k\}$ es acotado. Entonces, como C es cerrado, $\{x^k\} \subset C$ tiene al menos un punto límite factible. Por lo tanto, existen $x^* \in C$ y $K \subset \mathbb{N}$ tales que $\lim_{k \in K} x^k = x^*$. Por el Teorema 4.2.13, x^* es un punto estacionario del problema (4.0.1). Usando que F es \mathbb{R}_+^r -convexo, del Lema 4.3.12, se tiene que $x^* \in C$ es un punto Pareto optimal débil de (4.0.1).

Demostremos ahora que $x^* \in L$. Como $C^{k+1} \leq C^k$ para todo k , entonces, para cada k_0 fijo tenemos que $C^k \leq C^{k_0}$ para todo $k > k_0$. Entonces, $F(x^k) \leq C^{k_0}$ para todo $k > k_0$. Por lo tanto, tomando límite de $k \in K$ y usando la continuidad de F obtenemos $F(x^*) \leq C^{k_0}$. Entonces, $F(x^*) \leq C^k$ para todo $k \in K$. Luego $x^* \in L$. Por el Teorema 4.3.13, como $\{x^k\}$ tiene un punto límite $x^* \in L$ concluimos que $\{x^k\}$ converge a $x^* \in C$ como queríamos probar. \square

En este capítulo presentamos un nuevo algoritmo para optimización multiobjetivo con restricciones convexas. En cada iteración la dirección de búsqueda se calcula usando un paso variable que puede definirse considerando las ideas espectrales del caso escalar. La novedad es el uso de una técnica de búsqueda lineal no monótona en lugar de la clásica estrategia de Armijo.

Probamos la convergencia global a puntos límites estacionarios en el caso general y, en el caso convexo, la convergencia de toda la sucesión a un punto Pareto optimal débil bajo suposiciones razonables.

Como un desafío para el trabajo futuro, nos gustaría estudiar una implementación del Algoritmo 4.1 incorporando el paso espectral β_k . En realidad, en el Algoritmo 4.1, el cálculo de $v_{\beta_k}(x^k)$ implica, en cada iteración, la minimización del problema cuadrático

$$\begin{aligned} & \text{Minimizar } \beta \varphi_{x^k}(v) + \frac{\|v\|^2}{2} \\ & \text{sujeto a } v \in C - \{x^k\} \end{aligned} \quad (4.3.15)$$

El problema (4.3.15) es similar al subproblema cuadrático utilizado en [16] que calcula la dirección de búsqueda para resolver el problema escalar con restricciones

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeto a } x \in C \end{aligned}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función continuamente diferenciable y $C \subset \mathbb{R}^n$ es un conjunto cerrado y convexo. En [16] los autores sugieren calcular la dirección de búsqueda considerando la solución del problema cuadrático

$$\begin{aligned} & \text{Minimizar } Q_k(d) \\ & \text{sujeto a } x^k + d \in C \end{aligned} \quad (4.3.16)$$

donde $Q_k(d) = \frac{1}{2}d^T B_k d + \nabla f(x^k)^T d$ y $B_k \in \mathbb{R}^{n \times n}$ y es definida positiva tal que $\|B_k\| \leq \tilde{L}$ y $\|B_k^{-1}\| \leq \tilde{L}$. En realidad, ellos consideran el caso particular: $B_k = \frac{1}{\lambda_k^{spg}} I$ donde λ_k^{spg} es la elección del coeficiente espectral definida por

$$\lambda_k^{spg} = \begin{cases} \text{mín} \left\{ \lambda_{\text{máx}}, \text{máx} \left\{ \lambda_{\text{mín}}, \frac{(s^k)^T s^k}{(s^k)^T y^k} \right\} \right\} & \text{si } (s^k)^T y^k > 0 \\ \lambda_{\text{máx}} & \text{si no,} \end{cases}$$

donde $s^k = x^k - x^{k-1}$ y $y^k = \nabla f(x^k) - \nabla f(x^{k-1})$.

Al comparar los problemas (4.3.15) y (4.3.16) y teniendo en cuenta que la condición de optimalidad $\theta(x) = 0$ se satisface si consideramos

$$v(x) := \operatorname{argmin}_{v \in C - \{x\}} \varphi_x(v) + \frac{\|v\|^2}{2\beta} \quad (4.3.17)$$

y $\theta(x) := \varphi_x(v(x)) + \frac{\|v(x)\|^2}{2\beta}$, para cualquier $\beta > 0$ proponemos usar (4.3.17) para obtener la dirección de búsqueda en el Algoritmo 4.1.

Notemos que, si $r = 1$,

$$(s^k)^T y^k = (s^k)^T (\nabla f(x^k) - \nabla f(x^{k-1})) = \varphi_{x^k}(s^k) - \varphi_{x^{k-1}}(s^k).$$

Por lo tanto podemos considerar, en el caso multiobjetivo, una sucesión de parámetros $\{\beta_k\}$ definidos como

$$\beta_k = \begin{cases} \min \left\{ \beta_{\max}, \max \left\{ \beta_{\min}, \frac{(s^k)^T s^k}{\varphi_{x^k}(s^k) - \varphi_{x^{k-1}}(s^k)} \right\} \right\} & \text{si } \varphi_{x^k}(s^k) - \varphi_{x^{k-1}}(s^k) > 0 \\ \beta_{\max} & \text{si no,} \end{cases}$$

donde $s^k = x^k - x^{k-1}$ y φ_x está dado por (4.1.2). Este será un tema para futuras investigaciones.

Capítulo 5

El método Lagrangiano Aumentado

El objetivo de este capítulo es desarrollar un método Lagrangiano Aumentado para resolver el problema multiobjetivo utilizando una función Lagrangiana Aumentada vectorial.

Para analizar la posible extensión al caso multiobjetivo analizamos en primer lugar el caso escalar. Este estudio se presenta en la Sección 5.1. Se considera un problema escalar con dos conjuntos de restricciones: en el nivel superior las restricciones generales que se penalizan y en el nivel inferior restricciones de caja que no son penalizadas. Se analiza la convergencia global a puntos *KKT*.

En la Sección 5.2 generalizamos lo obtenido en la Sección 5.1 para un conjunto más general en el nivel inferior. Para obtener el resultado clásico de convergencia se asume que el conjunto de restricciones del nivel inferior es cerrado, convexo y que las restricciones cumplen la condición de Abadie.

La novedad en las Secciones 5.1 y 5.2 es el uso de la condición de Quasinormalidad para el análisis de la convergencia global a un punto *KKT*.

Finalmente en la Sección 5.3 analizamos la extensión del método Lagrangiano Aumentado al problema multiobjetivo. El subproblema para el método Lagrangiano Aumentado para el *MOP* es vectorial y se estudia la convergencia global a puntos regulares débiles utilizando la condición de Quasinormalidad.

5.1. Problema escalar con restricciones de cotas en las variables en el nivel inferior

En esta sección consideramos el siguiente problema escalar

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeto a } h(x) = 0, \quad g(x) \leq 0, \\ & x \in X = \{x \in \mathbb{R}^n : l \leq x \leq u\} \end{aligned} \quad (5.1.1)$$

donde consideramos a las funciones $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ y $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ con derivadas primeras continuas en un conjunto abierto que contiene a Ω .

Definimos el conjunto $\Omega = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$ que contiene las restricciones que se penalizan en el método Lagrangiano Aumentado y las llamamos restricciones del nivel superior. El conjunto $X = \{x \in \mathbb{R}^n : l \leq x \leq u\}$ para $l, u \in \mathbb{R}^n$, $l_i < u_i$, contiene las restricciones del nivel inferior que no se penalizan. En cada iteración del método Lagrangiano Aumentado, para resolver el problema (5.1.1), se busca encontrar una solución aproximada del subproblema

$$\begin{aligned} & \text{Minimizar } L(x, \lambda^k, \mu^k, \rho_k) \\ & \text{sujeto a } x \in X \end{aligned}$$

donde $L(x, \lambda, \mu, \rho) = f(x) + \phi(h(x), g(x), \lambda, \mu, \rho)$ es la función Lagrangiana Aumentada siendo ϕ la función de penalidad cuadrática:

$$\phi(h(x), g(x), \lambda, \mu, \rho) = \frac{\rho}{2} \left[\sum_{i=1}^m \left(h_i(x) + \frac{\lambda_i}{\rho} \right)^2 + \sum_{j=1}^p \left(\max \left\{ 0, g_j(x) + \frac{\mu_j}{\rho} \right\} \right)^2 \right] \quad (5.1.2)$$

para $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$ estimados de los multiplicadores de Lagrange y $\rho > 0$ el parámetro de penalidad.

Algoritmo 5.1 Los parámetros utilizados por el algoritmo son los siguientes:

$$\begin{aligned} & \tau \in [0, 1), \gamma > 1, \rho_0 \in \mathbb{R}_+, \\ & -\infty < \bar{\lambda}_i^{\text{mín}} < \bar{\lambda}_i^{\text{máx}} < \infty \quad \forall i = 1, \dots, m, \\ & 0 \leq \bar{\mu}_j^{\text{mín}} < \bar{\mu}_j^{\text{máx}} < \infty \quad \forall j = 1, \dots, p, \\ & \bar{\lambda}_i^0 \in [\bar{\lambda}_i^{\text{mín}}, \bar{\lambda}_i^{\text{máx}}] \quad \forall i = 1, \dots, m, \\ & \bar{\mu}_j^0 \in [0, \bar{\mu}_j^{\text{máx}}] \quad \forall j = 1, \dots, p, \\ & \sigma_j^0 = \max \left\{ g_j(x^0), -\frac{\bar{\mu}_j^0}{\rho_0} \right\} \quad \forall j = 1, \dots, p. \end{aligned}$$

Hacer $k \leftarrow 0$.

Paso 1. Resolver el subproblema

Calcular (si es posible) $x^k \in \mathbb{R}^n$ una solución aproximada de

$$\begin{aligned} & \text{Minimizar } L(x, \bar{\lambda}^k, \bar{\mu}^k, \rho_k) \\ & \text{sujeeto a } x \in X \end{aligned} \quad (5.1.3)$$

Paso 2. Estimar los multiplicadores de Lagrange

Para $i = 1, \dots, m$ calcular

$$\begin{aligned} \lambda_i^{k+1} &= \bar{\lambda}_i^k + \rho_k h_i(x^k), \\ \bar{\lambda}_i^{k+1} &\in [\bar{\lambda}_i^{\text{mín}}, \bar{\lambda}_i^{\text{máx}}]. \end{aligned} \quad (5.1.4)$$

Para todo $j = 1, \dots, p$ calcular

$$\begin{aligned} \mu_j^{k+1} &= \text{máx} \{0, \bar{\mu}_j^k + \rho_k g_j(x^k)\}, \\ \bar{\mu}_j^{k+1} &\in [0, \bar{\mu}_j^{\text{máx}}], \end{aligned} \quad (5.1.5)$$

$$\sigma_j^k = \text{máx} \left\{ g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k} \right\}. \quad (5.1.6)$$

Paso 3. Adaptar el parámetro de penalidad. Si $k = 0$ o

$$\text{máx} \{ \|h(x^k)\|_\infty, \|\sigma^k\|_\infty \} \leq \tau \text{máx} \{ \|h(x^{k-1})\|_\infty, \|\sigma^{k-1}\|_\infty \}$$

definir $\rho_{k+1} = \rho_k$.

Sino, definir $\rho_{k+1} = \gamma \rho_k$.

Hacer $k \leftarrow k + 1$. Ir al Paso 1.

El siguiente teorema y su corolario son análogos a el Teorema 3.1.8 y el Corolario 3.1.9 en el Capítulo 3. Por completitud los enunciamos.

Teorema 5.1.7. *Supongamos que la sucesión generada por el Algoritmo 5.1 es tal que $\lim_{k \in K} x^k = x^*$. Entonces si $g_i(x^*) < 0$ se tiene que $\lim_{k \in K} \mu_i^{k+1} = 0$*

Corolario 5.1.8. *Si x^* es un punto factible y $\lim_{k \in K} x^k = x^*$ entonces para todo $i = 1, \dots, p$*

$$\lim_{k \in K} \text{mín} \{ \mu_i^{k+1}, -g_i(x^k) \} = 0.$$

Suposición 2. En el Paso 1 del Algoritmo 5.1, decimos que x^k es una solución aproximada de (5.1.3) si

$$\|P_X(x^k - \nabla L(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)) - x^k\|_\infty < \epsilon_k \quad (5.1.9)$$

para $\{\epsilon_k\}$ tal que $\lim_{k \rightarrow \infty} \epsilon_k = 0$, $\epsilon_k \geq 0$ siendo $P_X(y)$ la proyección de y sobre X .

Enunciamos a continuación una proposición sobre proyecciones que será utilizada en la prueba de la convergencia global.

Proposición 5.1.10. [61]

$$\|P_X(u + tv) - u\| \leq \|P_X(u + v) - u\|$$

para todo $u, v \in X$, $t \in [0, 1]$.

En el próximo teorema se analiza la convergencia a puntos factibles de la sucesión generada por el algoritmo.

Teorema 5.1.11. *Supongamos que la sucesión $\{x^k\}$ es generada por el Algoritmo 5.1 satisfaciendo la Suposición 2 y que x^* es un punto límite de $\{x^k\}$. Entonces x^* es punto estacionario de*

$$\begin{aligned} & \text{Minimizar } \frac{1}{2} (\|h(x)\|^2 + \|\max\{0, g(x)\}\|^2) \\ & \text{sujeto a } x \in X. \end{aligned} \quad (5.1.12)$$

Demostración. Sea $K_1 \subset \mathbb{N}$ tal que $\lim_{k \in K_1} x^k = x^*$. Como $x^k \in X$ y X es cerrado tenemos que $x^* \in X$. Supongamos que la sucesión $\{\rho_k\}$ de parámetros de penalidad generada por el Algoritmo 5.1, cumple $\rho_{k+1} = \rho_k$ para todo k a partir de cierto k_0 . Luego $\lim_{k \in K_1} h(x^k) = 0$ y $\lim_{k \in K_1} \sigma^k = 0$. Entonces $h(x^*) = 0$.

Ahora, si $g_j(x^*) > 0$ para algún $j = 1, \dots, p$ entonces $g_j(x^k) \geq c > 0$ para todo $k \in K_1$, $k \geq k_1$.

Como $\{\rho_k\}$ y $\{\bar{\mu}^k\}$ son acotados se contradice que $\lim_{k \in K_1} \max \left\{ g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k} \right\} = 0$.

Por lo tanto $g_j(x^*) \leq 0$ para todo $j = 1, \dots, p$. Luego x^* es un punto factible y por lo tanto es un punto estacionario de (5.1.12).

Luego, supongamos que $\lim_{k \rightarrow \infty} \rho_k = \infty$. Por (5.1.9)

$$\lim_{k \in K_1} \|P_X(x^k - \nabla L(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)) - x^k\|_\infty = 0, \quad (5.1.13)$$

siendo

$$\begin{aligned} \nabla L(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k) = \nabla f(x^k) + \sum_{i=1}^m \bar{\lambda}_i^k \nabla h_i(x^k) + \rho_k \sum_{i=1}^m h_i(x^k) \nabla h_i(x^k) + \\ \sum_{j=1}^p \max \{0, \bar{\mu}_j^k + \rho_k g_j(x^k)\} \nabla g_j(x^k). \end{aligned} \quad (5.1.14)$$

Dividiendo por ρ_k en (5.1.14) y usando (5.1.13) y la Proposición 5.1.10

$$\lim_{k \in K_1} \left\| P_X \left(x^k - \frac{\nabla f(x^k)}{\rho_k} - \frac{1}{\rho_k} \sum_{i=1}^m \bar{\lambda}_i^k \nabla h_i(x^k) - \sum_{i=1}^m h_i(x^k) \nabla h_i(x^k) - \sum_{j=1}^p \max \left\{ 0, g_j(x^k) + \frac{\bar{\mu}_j^k}{\rho_k} \right\} \nabla g_j(x^k) \right) - x^k \right\|_{\infty} = 0. \quad (5.1.15)$$

Como $\{\bar{\lambda}^k\}$, $\{\bar{\mu}^k\}$ son acotados y ∇f , ∇h_i , ∇g_j son continuos, se obtiene que

$$\left\| P_X \left(x^* - \sum_{i=1}^m h_i(x^*) \nabla h_i(x^*) - \sum_{j=1}^p \max \{0, g_j(x^*)\} \nabla g_j(x^*) \right) - x^* \right\|_{\infty} = 0.$$

Por lo tanto, utilizando el Lema 2.1.7, x^* es punto estacionario de

$$\min_{x \in X} \frac{1}{2} (\|h(x)\|^2 + \|\max \{0, g(x)\}\|^2).$$

□

Observar que si $\{\rho_k\}$ es acotado se demuestra que todo punto límite es factible. Si $\{\rho_k\}$ no es acotado se demuestra que todo punto límite es un punto estacionario de (5.1.12). Luego, en este caso, la convergencia a un punto factible depende de la solución del problema (5.1.12) en el que se minimiza el cuadrado de la infactibilidad.

El teorema de convergencia global del Algoritmo 5.1 establece condiciones bajo las cuales un punto límite de la sucesión $\{x^k\}$ es un punto *KKT* de (5.1.1).

Teorema 5.1.16. *Un punto $x^* \in X \cap \Omega$ es *KKT* para el problema (5.1.1) si y solo si existen $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$ tales que*

$$P_X \left(x^* - \nabla f(x^*) - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) - x^* = 0$$

$$\mu_j = 0 \text{ si } g_j(x^*) < 0 \text{ con } j = 1, \dots, p.$$

Demostración.

\implies) Supongamos que x^* es un punto *KKT* de (5.1.1) entonces existen $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$, $\bar{\lambda} \in \mathbb{R}^n$, $\bar{\mu} \in \mathbb{R}^n$, $\bar{\lambda} \geq 0$, $\bar{\mu} \geq 0$ tales que $\mu_j g_j(x^*) = 0$, $j = 1, \dots, p$, $\bar{\lambda}_j (l_j - x_j^*) = 0$, $\bar{\mu}_j (x_j^* - u_j) = 0$ $j = 1, \dots, n$,

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) + \sum_{j=1}^n \bar{\lambda}_j (-e_j) + \sum_{j=1}^n \bar{\mu}_j e_j = 0. \quad (5.1.17)$$

Sea $y^* = P_X \left(x^* - \left[\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right)$. Quiero ver que $y^* = x^*$.

Por definición de proyección, y^* es solución del problema convexo

$$\min_{y \in X} \frac{1}{2} \left\| y - \left(x^* - \left[\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right) \right\|^2.$$

Pero por (5.1.17) x^* es un punto *KKT* para el problema anterior y como el problema tiene solución única debe ser $y^* = x^*$.

\impliedby) Ahora supongamos que

$$x^* = P_X \left(x^* - \left[\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right)$$

con $\mu \geq 0$, $\mu_j g_j(x^*) = 0$ para $j = 1, \dots, p$. Luego x^* es solución del problema convexo

$$\min_{y \in X} \frac{1}{2} \left\| y - \left(x^* - \left[\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right) \right\|^2$$

entonces existen $\bar{\lambda}$, $\bar{\mu} \geq 0$, $\bar{\lambda}_j (l_j - x_j^*) = 0$, $\bar{\mu}_j (x_j^* - u_j) = 0$, $j = 1, \dots, n$, tales que

$$x^* - \left(x^* - \left[\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right) + \sum_{j=1}^n \bar{\lambda}_j (-e_j) + \sum_{j=1}^n \bar{\mu}_j e_j = 0.$$

Por lo tanto x^* es un punto *KKT* para el problema (5.1.1). \square

Teorema 5.1.18. *Supongamos que la sucesión $\{x^k\}$ es generada por el Algoritmo 5.1 satisfaciendo la Suposición 2 y que x^* es un punto límite de $\{x^k\}$ que es Quasinormal para $\Omega \cap X$ en el sentido de la Definición 2.3.16. Entonces x^* es un punto *KKT* para el problema (5.1.1).*

Demostración. Sea $K \subset \mathbb{N}$ tal que $\lim_{k \in K} x^k = x^*$. Por (5.1.9) tenemos que

$$\|P_X(x^k - \nabla L(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)) - x^k\|_\infty < \epsilon_k. \quad (5.1.19)$$

Luego por (5.1.4) y (5.1.5)

$$\nabla L(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k) = \nabla f(x^k) + \sum_{i=1}^m \lambda_i^{k+1} \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^{k+1} \nabla g_j(x^k). \quad (5.1.20)$$

Supongamos que la sucesión $\{(1, \lambda^{k+1}, \mu^{k+1})\}$ es acotada, luego existen $K_1 \subset K$, $\bar{\lambda} \in \mathbb{R}^m$, $\bar{\mu} \in \mathbb{R}^p$, $\bar{\mu} \geq 0$, y por el Teorema 5.1.7 $\bar{\mu}_j = 0$ si $g_j(x^*) < 0$ tales que $\lim_{k \in K_1} (1, \lambda^{k+1}, \mu^{k+1}) = (1, \bar{\lambda}, \bar{\mu})$. Luego, tomando límite para $k \in K_1$ en (5.1.19) y usando (5.1.20) obtenemos

$$P_X \left(x^* - \left(\nabla f(x^*) + \sum_{i=1}^m \bar{\lambda}_i \nabla h_i(x^*) + \sum_{j=1}^p \bar{\mu}_j \nabla g_j(x^*) \right) \right) = x^*.$$

Luego por el Teorema 5.1.16 tenemos que x^* es un punto *KKT* para el problema (5.1.1).

Si la sucesión $\{(1, \lambda^{k+1}, \mu^{k+1})\}$ no es acotada, definimos $\delta_k = \|(1, \lambda^{k+1}, \mu^{k+1})\|$. Se tiene que $\delta_k \rightarrow \infty$ y además dividiendo por δ_k en (5.1.20) obtenemos, por (5.1.19) y la Proposición 5.1.10,

$$\left\| P_X \left(x^k - \frac{1}{\delta_k} \nabla f(x^k) - \sum_{i=1}^m \frac{\lambda_i^{k+1}}{\delta_k} \nabla h_i(x^k) - \sum_{j=1}^p \frac{\mu_j^{k+1}}{\delta_k} \nabla g_j(x^k) \right) - x^k \right\|_\infty < \epsilon_k. \quad (5.1.21)$$

Como $\left\{ \left(\frac{1}{\delta_k}, \frac{\lambda^{k+1}}{\delta_k}, \frac{\mu^{k+1}}{\delta_k} \right) \right\}$ es acotado existen $\bar{K} \subset K$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ tales que

$$\lim_{k \in \bar{K}} \left(\frac{1}{\delta_k}, \frac{\lambda^{k+1}}{\delta_k}, \frac{\mu^{k+1}}{\delta_k} \right) = (0, \lambda, \mu),$$

con $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$, no todos nulos y $\mu \geq 0$. Por lo tanto, por la continuidad de $\nabla h(x)$ y $\nabla g(x)$, tomando límite para $k \in \bar{K}$ en (5.1.21) obtenemos

$$\left\| P_X \left(x^* - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) - x^* \right\|_\infty = 0$$

que equivale, por la Observación 2.3.23 a

$$-\sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \in N_X(x^*).$$

Como $\lim_{k \in \bar{K}} \frac{\lambda^{k+1}}{\delta_k} = \lambda$ y $\lambda^{k+1} = \bar{\lambda}^k + \rho_k h(x^k)$, entonces, por ser $\{\bar{\lambda}^k\}$ una sucesión acotada, $\lim_{k \in \bar{K}} \frac{\bar{\lambda}^k}{\delta_k} = 0$ y $\lim_{k \in \bar{K}} \frac{\rho_k h(x^k)}{\delta_k} = \lambda$.

Luego si $\lambda_i \neq 0$ existe $k_1 \in \bar{K}$ tal que para todo $k \in \bar{K}$, $k \geq k_1$, $h_i(x^k)\lambda_i > 0$.

De manera análoga, $\lim_{k \in \bar{K}} \frac{\mu^{k+1}}{\delta_k} = \mu$ y $\mu^{k+1} = \max\{0, \bar{\mu}^k + \rho_k g(x^k)\}$, entonces como $\{\bar{\mu}^k\}$ es acotada $\lim_{k \in \bar{K}} \frac{\bar{\mu}^k}{\delta_k} = 0$ y $\lim_{k \in \bar{K}} \frac{\rho_k \sigma^k}{\delta_k} = \mu$.

Luego si $\mu_j > 0$ existe $k_2 \in \bar{K}$ tal que para todo $k \in \bar{K}$, $k \geq k_2$, $\sigma_j^k \mu_j > 0$, entonces $g_j(x^k)\mu_j > 0$.

Y por lo tanto se contradice la definición de Quasinormalidad, luego debe ser la sucesión $\{(1, \lambda^{k+1}, \mu^{k+1})\}$ acotada. \square

5.2. Problema escalar con restricciones que satisfacen la condición de Abadie en el nivel inferior

En esta sección consideremos el siguiente problema escalar con restricciones generales

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{sujeto a } h(x) = 0, \quad g(x) \leq 0, \\ & x \in \bar{\Omega} = \{x \in \mathbb{R}^n : \underline{h}(x) = 0, \underline{g}(x) \leq 0\} \end{aligned} \quad (5.2.1)$$

donde $\Omega = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$ contiene las restricciones que se penalizan, llamadas restricciones del nivel superior y $\bar{\Omega} = \{x \in \mathbb{R}^n : \underline{h}(x) = 0, \underline{g}(x) \leq 0\}$ contiene las restricciones del nivel inferior que no se penalizan. Consideramos $\underline{h} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\underline{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p$. Como ya hemos mencionado, en cada iteración del Lagrangiano Aumentado para resolver el problema general se busca encontrar una solución aproximada del subproblema

$$\begin{aligned} & \text{Minimizar } L(x, \lambda^k, \mu^k, \rho_k) \\ & \text{sujeto a } x \in \bar{\Omega} \end{aligned} \quad (5.2.2)$$

siendo $L(x, \lambda^k, \mu^k, \rho_k) = f(x) + \phi(h(x), g(x), \lambda, \mu, \rho)$ para ϕ dada por (5.1.2).

Para obtener los resultados clásicos de convergencia global vamos a suponer que el conjunto $\bar{\Omega} = \{x \in \mathbb{R}^n : \underline{h}(x) = 0, \underline{g}(x) \leq 0\}$ es un conjunto cerrado, convexo y que las restricciones $\underline{h}, \underline{g}$ cumplen la condición de Abadie.

Algoritmo 5.2 Los parámetros utilizados por el algoritmo son los siguientes:

$$\begin{aligned} \tau &\in [0, 1), \gamma > 1, \rho_0 \in \mathbb{R}_+ \\ -\infty &< \bar{\lambda}_i^{\min} < \bar{\lambda}_i^{\max} < \infty \quad \forall i = 1, \dots, m, \\ \bar{\lambda}_i^0 &\in [\bar{\lambda}_i^{\min}, \bar{\lambda}_i^{\max}] \quad \forall i = 1, \dots, m, \\ 0 &\leq \bar{\mu}_j^{\min} < \bar{\mu}_j^{\max} < \infty \quad \forall j = 1, \dots, p, \\ \bar{\mu}_j^0 &\in [0, \bar{\mu}_j^{\max}] \quad \forall j = 1, \dots, p, \\ \sigma_j^0 &= \max \left\{ g_j(x^0), -\frac{\bar{\mu}_j^0}{\rho_0} \right\} \quad \forall j = 1, \dots, p. \end{aligned}$$

Hacer $k \leftarrow 0$.

Paso 1. Resolver el subproblema

Calcular (si es posible) $x^k \in \mathbb{R}^n$ una solución aproximada de

$$\begin{aligned} \text{Minimizar } &L(x, \bar{\lambda}^k, \bar{\mu}^k, \rho_k) \\ \text{sujeto a } &x \in \bar{\Omega} \end{aligned} \tag{5.2.3}$$

Paso 2. Estimar los multiplicadores de Lagrange

Para $i = 1, \dots, m$ calcular

$$\begin{aligned} \lambda_i^{k+1} &= \bar{\lambda}_i^k + \rho_k h_i(x^k), \\ \bar{\lambda}_i^{k+1} &\in [\bar{\lambda}_i^{\min}, \bar{\lambda}_i^{\max}]. \end{aligned} \tag{5.2.4}$$

Para todo $j = 1, \dots, p$ calcular

$$\mu_j^{k+1} = \max \left\{ 0, \bar{\mu}_j^k + \rho_k g_j(x^k) \right\}, \bar{\mu}_j^{k+1} \in [0, \bar{\mu}_j^{\max}] \tag{5.2.5}$$

$$\sigma_j^k = \max \left\{ g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k} \right\} \tag{5.2.6}$$

Paso 3. Adaptar los parámetros de penalidad Si $k = 0$ o

$$\max \{ \|h(x^k)\|_\infty, \|\sigma^k\|_\infty \} \leq \tau \max \{ \|h(x^{k-1})\|_\infty, \|\sigma^{k-1}\|_\infty \}$$

definir $\rho_{k+1} = \rho_k$.

Sino, definir $\rho_{k+1} = \gamma\rho_k$.

Hacer $k \leftarrow k + 1$. Ir al Paso 1.

Observación 5.2.7. Las restricciones del conjunto $\bar{\Omega}$ cumplen la condición de Abadie entonces, para cada $x \in \bar{\Omega}$

$$L_{\bar{\Omega}}(x) = T_{\bar{\Omega}}(x).$$

Luego

$$L_{\bar{\Omega}}^*(x) = T_{\bar{\Omega}}^*(x),$$

y, por la definición de cono normal y el Teorema 2.3.8 se tiene que

$$G(x) = N_{\bar{\Omega}}(x).$$

Por lo tanto para cada $x \in \bar{\Omega}$

$$N_{\bar{\Omega}}(x) = \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j=1}^p \mu_j \nabla g_j(x) : \mu_j \geq 0, \mu_j = 0 \text{ si } g_j(x) < 0 \right\}.$$

Las condiciones (i) y (ii) de la definición 2.3.16 se satisfacen en $x^* \in \bar{\Omega}$ si existen $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$ tales que

$$- \left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) \in N_{\bar{\Omega}}(x^*).$$

Es decir si existen $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$ tales que

$$- \left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) = \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*).$$

Suposición 3. En el Paso 1 del Algoritmo 5.2, decimos que x^k es una solución aproximada de (5.2.3) si

$$\|P_{\bar{\Omega}}(x^k - \nabla L(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)) - x^k\|_{\infty} < \epsilon_k \quad (5.2.8)$$

para $\{\epsilon_k\}$ tal que $\lim_{k \rightarrow \infty} \epsilon_k = 0$, $\epsilon_k \geq 0$ siendo $P_{\bar{\Omega}}(y)$ la proyección de y sobre $\bar{\Omega}$.

Observar que se puede proyectar sobre $\bar{\Omega}$ porque suponemos que es un conjunto convexo y cerrado.

Teorema 5.2.9. Un punto $x^* \in \bar{\Omega} \cap \Omega$ es KKT de (5.2.1) si y solo si existen $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ $\mu \geq 0$ tales que

$$P_{\bar{\Omega}} \left(x^* - \nabla f(x^*) - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) - x^* = 0$$

$$\mu_j = 0 \text{ si } g_j(x^*) < 0 \text{ para todo } j = 1, \dots, p.$$

Demostración.

\implies) Supongamos que x^* es un punto *KKT* de (5.2.1) entonces existen $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$, $\underline{\lambda} \in \mathbb{R}^m$, $\underline{\mu} \in \mathbb{R}^p$, $\underline{\mu} \geq 0$, tales que $\mu_j g_j(x^*) = 0$, $j = 1, \dots, p$, $\underline{\mu}_j \underline{g}_j(x^*) = 0$ $j = 1, \dots, p$ y

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) + \sum_{i=1}^m \underline{\lambda}_i \nabla \underline{h}_i(x^*) + \sum_{j=1}^p \underline{\mu}_j \nabla \underline{g}_j(x^*) = 0. \quad (5.2.10)$$

Sea $y^* = P_{\Omega} \left(x^* - \left[\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right)$. Quiero ver que $y^* = x^*$.

Por definición de proyección, y^* es solución del problema convexo

$$\min_{y \in \Omega} \frac{1}{2} \left\| y - \left(x^* - \left[\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right) \right\|^2.$$

Pero por (5.2.10) x^* es *KKT* para el problema anterior y como el problema tiene solución única debe ser $y^* = x^*$.

\Leftarrow) Ahora supongamos que

$$x^* = P_{\Omega} \left(x^* - \left[\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right)$$

con $\mu \geq 0$, $\mu_j g_j(x^*) = 0$. Luego x^* es solución del problema convexo

$$\min_{y \in \Omega} \frac{1}{2} \left\| y - \left(x^* - \left[\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right) \right\|^2$$

entonces existen $\underline{\lambda} \in \mathbb{R}^m$, $\underline{\mu} \in \mathbb{R}^p$, $\underline{\mu} \geq 0$, $\underline{\mu}_j \underline{g}_j(x^*) = 0$ $j = 1, \dots, p$, tales que

$$x^* - \left(x^* - \left[\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right) + \sum_{i=1}^m \underline{\lambda}_i \nabla \underline{h}_i(x^*) + \sum_{j=1}^p \underline{\mu}_j \nabla \underline{g}_j(x^*) = 0.$$

Por lo tanto x^* es un punto *KKT* de (5.2.1) . □

La demostración del siguiente teorema es análoga a la demostración del Teorema 5.1.11.

Teorema 5.2.11. *Supongamos que la sucesión $\{x^k\}$ es generada por el Algoritmo 5.2 satisfaciendo la Suposición 3 y que x^* es un punto límite de $\{x^k\}$. Entonces x^* es punto estacionario de*

$$\begin{aligned} & \text{Minimizar } \frac{1}{2} (\|h(x)\|^2 + \|\text{máx}\{0, g(x)\}\|^2) \\ & \text{sujeto a } x \in \bar{\Omega}. \end{aligned} \quad (5.2.12)$$

En el siguiente teorema se establece la convergencia global del Algoritmo 5.2.

Teorema 5.2.13. *Supongamos que la sucesión $\{x^k\}$ es generada por el Algoritmo 5.2 satisfaciendo la Suposición 3 y que x^* es un punto límite de $\{x^k\}$ que satisface la definición de Quasinormalidad (2.3.16) y además, las restricciones del conjunto $\bar{\Omega}$ cumplen la condición de Abadie. Entonces x^* es un punto KKT de (5.2.1).*

Demostración. Sea $K \subset \mathbb{N}$ tal que $\lim_{k \in K} x^k = x^*$. Por (5.2.8) tenemos que

$$\|P_{\bar{\Omega}}(x^k - \nabla L(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)) - x^k\|_{\infty} < \epsilon_k. \quad (5.2.14)$$

Luego, por (5.2.4) y (5.2.5) tenemos que

$$\nabla L(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k) = \nabla f(x^k) + \sum_{i=1}^m \lambda_i^{k+1} \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^{k+1} \nabla g_j(x^k). \quad (5.2.15)$$

Supongamos que la sucesión $\{(1, \lambda^{k+1}, \mu^{k+1})\}$ es acotada, luego existen $K_1 \subset K$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$ y por el Teorema 5.1.7 $\mu_j = 0$ si $g_j(x^*) < 0$ tales que $\lim_{k \in K_1} (1, \lambda^{k+1}, \mu^{k+1}) = (1, \lambda, \mu)$. Luego, tomando límite para $k \in K_1$ en (5.2.14) y usando (5.2.15) obtenemos

$$P_{\bar{\Omega}} \left(x^* - \left(\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) \right) = x^*.$$

Por lo tanto, por el Teorema 5.2.9, x^* es un punto KKT de (5.2.1).

Si la sucesión $\{(1, \lambda^{k+1}, \mu^{k+1})\}$ no es acotada, definimos $\delta_k = \|(1, \lambda^{k+1}, \mu^{k+1})\|$. Se tiene que $\delta_k \rightarrow \infty$ y además dividiendo por δ_k en (5.2.15) y usando la Proposición 5.1.10, por (5.2.14) se tiene

$$\left\| P_{\bar{\Omega}} \left(x^k - \frac{1}{\delta_k} \nabla f(x^k) - \sum_{i=1}^m \frac{\lambda_i^{k+1}}{\delta_k} \nabla h_i(x^k) - \sum_{j=1}^p \frac{\mu_j^{k+1}}{\delta_k} \nabla g_j(x^k) \right) - x^k \right\|_{\infty} < \epsilon_k. \quad (5.2.16)$$

Como la sucesión $\left\{ \left(\frac{1}{\delta_k}, \frac{\lambda^{k+1}}{\delta_k}, \frac{\mu^{k+1}}{\delta_k} \right) \right\}$ es acotada existe $\bar{K} \subset K$, tal que

$$\lim_{k \in \bar{K}} \left(\frac{1}{\delta_k}, \frac{\lambda^{k+1}}{\delta_k}, \frac{\mu^{k+1}}{\delta_k} \right) = (0, \lambda, \mu),$$

para $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$, no todos nulos y $\mu \geq 0$. Por lo tanto, por la continuidad de $\nabla h(x)$ y $\nabla g(x)$, tomando límite en (5.2.16) obtenemos

$$\left\| P_{\bar{\Omega}} \left(x^* - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) - x^* \right\|_{\infty} = 0.$$

Que equivale, utilizando la Observación 5.2.7 a $-\sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \in N_{\bar{\Omega}}(x^*)$, pues

$$x^* = P_{\bar{\Omega}} \left(x^* - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right)$$

es equivalente a que x^* es solución del problema convexo

$$\min_{y \in \bar{\Omega}} \frac{1}{2} \left\| y - \left(x^* - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) \right\|^2.$$

Luego

$$-x^* + \left(x^* - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) = \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*)$$

Es decir existen $\underline{\lambda} \in \mathbb{R}^m$, $\underline{\mu} \in \mathbb{R}^p$, $\underline{\mu} \geq 0$ tales que

$$-\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) = \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*).$$

Luego, las condiciones (i) y (ii) de Quasinormalidad para $\bar{\Omega} \cap \Omega$ se satisfacen ya que $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$, son no todos nulos tales que

$$-\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) \in N_{\bar{\Omega}}(x^*).$$

Como $\lim_{k \in \bar{K}} \frac{\lambda^{k+1}}{\delta_k} = \lambda$ y $\lambda^{k+1} = \bar{\lambda}^k + \rho_k h(x^k)$, entonces, como $\{\bar{\lambda}^k\}$ es acotada,
 $\lim_{k \in \bar{K}} \frac{\bar{\lambda}^k}{\delta_k} = 0$ y $\lim_{k \in \bar{K}} \frac{\rho_k h(x^k)}{\delta_k} = \lambda$.

Luego si $\lambda_i \neq 0$ existe $k_1 \in \bar{K}$ tal que para todo $k \in \bar{K}$, $k \geq k_1$, $h_i(x^k)\lambda_i > 0$.

Además, como $\lim_{k \in \bar{K}} \frac{\mu^{k+1}}{\delta_k} = \mu$ y $\mu^{k+1} = \max\{0, \bar{\mu}^k + \rho_k g(x^k)\}$, entonces como $\{\bar{\mu}^k\}$ es acotada $\lim_{k \in \bar{K}} \frac{\bar{\mu}^k}{\delta_k} = 0$ y $\lim_{k \in \bar{K}} \frac{\rho_k \sigma^k}{\delta_k} = \mu$.

Luego si $\mu_j > 0$ existe $k_2 \in \bar{K}$ tal que para todo $k \in \bar{K}$, $k \geq k_2$, $\sigma_j^k \mu_j > 0$, entonces $g_j(x^k)\mu_j > 0$.

Y por lo tanto se contradice la condición de Quasinormalidad, luego la sucesión $\{(1, \lambda^{k+1}, \mu^{k+1})\}$ es acotada. \square

5.3. Problema multiobjetivo con restricciones que satisfacen la condición de Abadie en el nivel inferior

En esta sección analizamos la extensión del método Lagrangiano Aumentado de la sección anterior al problema de optimización multiobjetivo

$$\begin{aligned} & \text{Minimizar } F(x) \\ & \text{sujeto a } h(x) = 0, \quad g(x) \leq 0, \\ & x \in X = \{x \in \mathbb{R}^n : \underline{h}(x) = 0, \underline{g}(x) \leq 0\}. \end{aligned} \tag{5.3.1}$$

donde $F : \mathbb{R}^n \rightarrow \mathbb{R}^r$, $F(x) = (F_1(x), \dots, F_r(x))$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, son funciones dadas. El conjunto $\Omega = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$ contiene las restricciones que serán penalizadas y las llamamos restricciones del nivel superior. El conjunto $\bar{\Omega} = \{x \in \mathbb{R}^n : \underline{h}(x) = 0, \underline{g}(x) \leq 0\}$ siendo $\underline{h} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\underline{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p$, contiene las restricciones del nivel inferior que no serán penalizan. Suponemos que $\bar{\Omega}$ es un conjunto cerrado, convexo y las restricciones que lo definen satisfacen la condición de Abadie.

En cada iteración del método Lagrangiano Aumentado para resolver el problema (5.3.1) se busca encontrar una solución aproximada del problema multiobjetivo

$$\begin{aligned} & \text{Minimizar } \mathcal{L}(x, \lambda, \mu, \rho) \\ & \text{sujeto a } x \in \bar{\Omega} \end{aligned}$$

donde $\mathcal{L}(x, \lambda, \mu, \rho) = F(x) + \phi(h(x), g(x), \lambda, \mu, \rho)(1, \dots, 1)^T$ es la función Lagrangiana

Aumentada vectorial siendo

$$\phi(h(x), g(x), \lambda, \mu, \rho) = \frac{\rho}{2} \left[\sum_{i=1}^m \left(h_i(x) + \frac{\lambda_i}{\rho} \right)^2 + \sum_{j=1}^p (\max\{0, \mu_j + \rho g_j(x)\})^2 \right]$$

para $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ estimados de los multiplicadores de Lagrange, $\rho > 0$ el parámetro de penalidad y el vector $(1, \dots, 1) \in \mathbb{R}^r$. La función \mathcal{L} es una función vectorial dada por $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_r)$ siendo

$$\mathcal{L}_i(x, \lambda, \mu, \rho) = F_i(x) + \frac{\rho}{2} \left[\sum_{i=1}^m \left(h_i(x) + \frac{\lambda_i}{\rho} \right)^2 + \sum_{j=1}^p (\max\{0, \mu_j + \rho g_j(x)\})^2 \right],$$

para cada $i = 1, \dots, r$.

En el Paso 1 del algoritmo se calcula (si es posible) $x^k \in \mathbb{R}^n$ una solución aproximada del problema multiobjetivo

$$\begin{aligned} & \text{Minimizar } \mathcal{L}(x, \lambda^k, \mu^k, \rho_k) \\ & \text{sujeto a } x \in \bar{\Omega} \end{aligned} \tag{5.3.2}$$

siendo λ^k, μ^k estimados de los multiplicadores de Lagrange y ρ_k parámetro de penalidad. En la siguiente definición estableceremos el concepto que le damos a una solución aproximada de (5.3.2).

Suposición 4. Decimos que x^k es una solución aproximada de (5.3.2) si existe $w(x^k) \in \mathbb{R}^r$, $\sum_{i=1}^r w_i(x^k) = 1$, tal que

$$\|v(x^k)\|_{\infty} = \|P_{\bar{\Omega}}(x^k - \beta_k J\mathcal{L}(x^k, \lambda^k, \mu^k, \rho_k)^T w(x^k)) - x^k\|_{\infty} < \epsilon_k$$

para $\{\epsilon_k\} \subset \mathbb{R}$ tal que $\lim_{k \rightarrow \infty} \epsilon_k = 0$, $\epsilon_k \geq 0$ y $\beta_k > 0$ tal que $\beta_k \in [\beta_{\min}, \beta_{\max}]$.

Observación 5.3.3. El vector $v(x^k)$ en este caso se puede calcular con el algoritmo del gradiente proyectado para el MOP presentado en el capítulo anterior.

Algoritmo 5.3 Los parámetros utilizados por el algoritmo son los siguientes:

$$\begin{aligned} & \tau \in [0, 1), \gamma > 1, \rho_0 \in \mathbb{R}_+, \\ & -\infty < \bar{\lambda}_i^{\min} < \bar{\lambda}_i^{\max} < \infty \quad \forall i = 1, \dots, m, \\ & \bar{\lambda}_i^0 \in [\bar{\lambda}_i^{\min}, \bar{\lambda}_i^{\max}] \quad \forall i = 1, \dots, m, \\ & 0 \leq \bar{\mu}_j^{\min} < \bar{\mu}_j^{\max} < \infty \quad \forall j = 1, \dots, p, \\ & \bar{\mu}_j^0 \in [0, \bar{\mu}_j^{\max}] \quad \forall j = 1, \dots, p, \\ & \sigma_j^0 = \max \left\{ g_j(x^0), -\frac{\bar{\mu}_j^0}{\rho_0} \right\} \quad \forall j = 1, \dots, p. \end{aligned}$$

Hacer $k \leftarrow 0$.

Paso 1. Resolver el subproblema

Calcular (si es posible) $x^k \in \mathbb{R}^n$ una solución aproximada del subproblema multiobjetivo

$$\begin{aligned} & \text{Minimizar } \mathcal{L}(x, \bar{\lambda}^k, \bar{\mu}^k, \rho_k) \\ & \text{sujeto a } x \in \bar{\Omega} \end{aligned} \quad (5.3.4)$$

Paso 2. Estimar los multiplicadores de Lagrange

Para $i = 1, \dots, m$ calcular

$$\begin{aligned} \lambda_i^{k+1} &= \bar{\lambda}_i^k + \rho_k h_i(x^k), \\ \bar{\lambda}_i^{k+1} &\in [\bar{\lambda}_i^{\text{mín}}, \bar{\lambda}_i^{\text{máx}}]. \end{aligned} \quad (5.3.5)$$

Para todo $j = 1, \dots, p$ calcular

$$\mu_j^{k+1} = \text{máx} \{0, \bar{\mu}_j^k + \rho_k g_j(x^k)\}, \bar{\mu}_j^{k+1} \in [0, \bar{\mu}_j^{\text{máx}}] \quad (5.3.6)$$

$$\sigma_j^k = \text{máx} \left\{ g_j(x^k), -\frac{\bar{\mu}_j^k}{\rho_k} \right\} \quad (5.3.7)$$

Paso 3. Adaptar los parámetros de penalidad Si $k = 0$ o

$$\text{máx} \{ \|h(x^k)\|_\infty, \|\sigma^k\|_\infty \} \leq \tau \text{máx} \{ \|h(x^{k-1})\|_\infty, \|\sigma^{k-1}\|_\infty \}$$

definir $\rho_{k+1} = \rho_k$.

Sino, definir $\rho_{k+1} = \gamma \rho_k$.

Hacer $k \leftarrow k + 1$. Ir al Paso 1.

Teorema 5.3.8. *Supongamos que la sucesión $\{x^k\}$ es generada por el Algoritmo 5.3 satisfaciendo la Suposición 4 y que x^* es un punto límite de $\{x^k\}$. Entonces x^* es punto estacionario de*

$$\begin{aligned} & \text{Minimizar } \frac{1}{2} (\|h(x)\|^2 + \|\text{máx} \{0, g(x)\}\|^2) \\ & \text{sujeto a } x \in \bar{\Omega} \end{aligned} \quad (5.3.9)$$

Demostración. Sea $K_1 \subset \mathbb{N}$ tal que $\lim_{k \in K_1} x^k = x^*$, como $x^k \in \bar{\Omega}$ y $\bar{\Omega}$ es cerrado tenemos que $x^* \in \bar{\Omega}$. La sucesión $\{\beta_k\}$ es acotada, luego, sin pérdida de generalidad existe $\beta > 0$ tal que $\lim_{k \in K_1} \beta_k = \beta$.

Supongamos que la sucesión $\{\rho_k\}$ de parámetros de penalidad generada por el algoritmo cumple $\rho_{k+1} = \rho_k$ para todo k a partir de cierto k_0 . Entonces $\lim_{k \in K_1} h(x^k) = 0$ y $\lim_{k \in K_1} \sigma_k = 0$ entonces $h(x^*) = 0$ y verifica que $g(x^*) \leq 0$. Por lo tanto x^* es factible y entonces x^* es punto estacionario de (5.3.9).

Luego, supongamos que $\lim_{k \in K_1} \rho_k = \infty$. Por la Suposición 4,

$$\lim_{k \in K_1} \|P_{\bar{\Omega}}(x^k - \beta_k J\mathcal{L}(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)^T w(x^k)) - x^k\|_{\infty} = 0 \quad (5.3.10)$$

siendo

$$\beta_k J\mathcal{L}(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)^T w(x^k) = \beta_k JF^T(x^k)w(x^k) + \beta_k \sum_{i=1}^m (\bar{\lambda}_i^k + \rho_k h_i(x^k)) \nabla h_i(x^k) + \beta_k \sum_{j=1}^p \max\{0, \bar{\mu}_j^k\} \nabla g_j(x^k) \quad (5.3.11)$$

donde hemos usado que $\sum_{i=1}^r w_i(x^k) = 1$.

Dividiendo por ρ_k en (5.3.11) como $\{\bar{\lambda}_i^k\}$, $\{\bar{\mu}_j^k\}$ son acotados para $k \in K_1$ y ∇F_l , ∇h_i , ∇g_j son continuos, tomando límite para $k \in K_1$ se obtiene que

$$\left\| P_{\bar{\Omega}} \left(x^* - \beta \left[\sum_{i=1}^m h_i(x^*) \nabla h_i(x^*) - \sum_{j=1}^p \max\{0, g_j(x^*)\} \nabla g_j(x^*) \right] \right) - x^* \right\|_{\infty} = 0.$$

Por lo tanto, por el Teorema 5.2.9 y por ser $\beta > 0$, x^* es punto estacionario de $\min_{x \in \bar{\Omega}} \|h(x)\|^2 + \|\max\{0, g(x)\}\|^2$ siendo $\bar{\Omega} = \{x \in \mathbb{R}^n : \underline{h}(x) = 0, \underline{g}(x) \leq 0\}$ cerrado y convexo. \square

En el siguiente teorema establecemos una equivalencia con la definición de regularidad débil para el problema (5.3.1).

Teorema 5.3.12. *Un punto x^* es regular débil para (5.3.1) si y solo si existen $\theta \in \mathbb{R}^r$ $\theta \neq 0$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ $\mu \geq 0$, $\mu_j g_j(x^*) = 0$ tales que*

$$P_{\bar{\Omega}} \left(x^* - \sum_{l=1}^r \theta_l \nabla F_l(x^*) - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) - x^* = 0.$$

Demostración.

\implies) Supongamos que x^* es un punto regular débil de (5.3.1) entonces existen $\theta \in \mathbb{R}^r$ $\theta \neq 0$ $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$, $\underline{\lambda} \in \mathbb{R}^m$, $\underline{\mu} \in \mathbb{R}^p$, $\underline{\mu} \geq 0$, $\mu_j g_j(x^*) = 0$, $\bar{\mu}_j g_j(x^*) = 0$ tales que

$$\sum_{l=1}^r \theta_l \nabla F_l(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) + \sum_{i=1}^m \underline{\lambda}_i \nabla \underline{h}_i(x^*) + \sum_{j=1}^p \underline{\mu}_j \nabla \underline{g}_j(x^*) = 0.$$

Sea $y^* = P_{\bar{\Omega}} \left(x^* - \left[\sum_{l=1}^r \theta_l \nabla F_l(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right)$ quiero ver que $y^* = x^*$. Por definición de proyección y^* es solución del problema convexo

$$\min_{y \in \bar{\Omega}} \frac{1}{2} \left\| y - \left(x^* - \left[\sum_{l=1}^r \theta_l \nabla F_l(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right) \right\|^2.$$

Pero x^* es un punto *KKT* para el problema anterior y como tiene solución única debe ser $y^* = x^*$.

\Leftarrow) Ahora supongamos que

$$x^* = P_{\bar{\Omega}} \left(x^* - \left[\sum_{l=1}^r \theta_l \nabla F_l(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right)$$

para $\theta \neq 0$, $\mu \geq 0$, $\mu_j g_j(x^*) = 0$. Luego x^* es solución del problema convexo

$$\min_{y \in \bar{\Omega}} \frac{1}{2} \left\| y - \left(x^* - \left[\sum_{l=1}^r \theta_l \nabla F_l(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right) \right\|^2$$

entonces existen $\underline{\lambda} \in \mathbb{R}^m$, $\underline{\mu} \in \mathbb{R}^p$, $\underline{\mu} \geq 0$, $\underline{\mu}_j g_j(x^*) = 0$ tales que

$$x^* - \left(x^* - \left[\sum_{l=1}^r \theta_l \nabla F_l(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right] \right) + \sum_{i=1}^m \underline{\lambda}_i \nabla h_i(x^*) + \sum_{j=1}^p \underline{\mu}_j \nabla g_j(x^*) = 0.$$

Por lo tanto x^* es un punto regular débil de (5.3.1). \square

Teorema 5.3.13. *Supongamos que x^* es un punto límite de una sucesión $\{x^k\}$ generada por el Algoritmo 5.3. Supongamos que x^* es Quasinormal para $\Omega \cap \bar{\Omega}$ en el sentido de la definición (2.3.16). Entonces x^* es un punto regular débil de (5.3.1).*

Demostración. Sea $K \subset \mathbb{N}$ tal que $\lim_{k \in K} x^k = x^*$. Como x^k es una solución aproxima-

mada de (5.3.2) existe $w(x) \in \mathbb{R}^r$, $\sum_{i=1}^r w_i(x^k) = 1$, tal que

$$\|v(x^k)\|_{\infty} = \|P_{\bar{\Omega}}(x^k - \beta_k J\mathcal{L}(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)^T w(x^k)) - x^k\|_{\infty} < \epsilon_k.$$

Luego, se tiene que

$$\lim_{k \in K} \|P_{\bar{\Omega}}(x^k - \beta_k J\mathcal{L}(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)^T w(x^k)) - x^k\|_{\infty} = 0 \quad (5.3.14)$$

donde, por (5.3.5) y (5.3.6) se tiene

$$\begin{aligned} \beta_k J\mathcal{L}(x^k, \bar{\lambda}^k, \bar{\mu}^k, \rho_k)^T w(x^k) &= \beta_k JF^T(x^k)w(x^k) + \beta_k \sum_{i=1}^m \lambda_i^{k+1} \nabla h_i(x^k) + \\ &\beta_k \sum_{j=1}^p \mu_j^{k+1} \nabla g_j(x^k). \end{aligned} \quad (5.3.15)$$

donde hemos usado que $\sum_{i=1}^r w_i(x^k) = 1$.

La sucesión $\{\beta_k\}$ es acotada entonces tiene una subsucesión convergente, es decir, existe $K_1 \subset K$ y $\beta > 0$ tal que $\lim_{k \in K_1} \beta_k = \beta$.

Supongamos que $\{(\beta_k w(x^k), \beta_k \lambda^{k+1}, \beta_k \mu^{k+1})\}$ es acotada. Luego tomando límite para $k \in K_1$ existen $\theta \in \mathbb{R}^r$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ tales que

$$\lim_{k \in K_1} (\beta_k w(x^k), \beta_k \lambda^{k+1}, \beta_k \mu^{k+1}) = (\theta, \lambda, \mu)$$

y $\mu_j \geq 0$, cumple, por el Teorema 5.1.7 $\mu_j = 0$ si $g_j(x^*) < 0$ y por (5.3.15) y (5.3.14)

$$P_{\bar{\Omega}} \left(x^* - \left(\sum_{l=1}^r \theta_l \nabla F_l(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) \right) = x^*.$$

Como $\|w(x^k)\| = 1$ tenemos que $\|\beta_k w(x^k)\| = \beta_k$ y como $\lim_{k \in K_1} \beta_k w(x^k) = \theta$ se tiene que $\|\theta\| = \beta \neq 0$. Luego, por el Teorema 5.3.12 x^* es un punto regular débil para el problema.

Si no, definimos $\delta_k = \|(\beta_k w(x^k), \beta_k \lambda^{k+1}, \beta_k \mu^{k+1})\|$. Se tiene que $\delta_k \rightarrow \infty$ y además dividiendo por δ_k en (5.3.15) y usando (5.3.14) se tiene

$$\left\| P_{\bar{\Omega}} \left(x^k - \frac{\beta_k w^T(x^k) JF(x^k)}{\delta_k} - \frac{\beta_k}{\delta_k} \sum_{i=1}^m \lambda_i^{k+1} \nabla h_i(x^k) - \frac{\beta_k}{\delta_k} \sum_{j=1}^p \mu_j^{k+1} \nabla g_j(x^k) \right) - x^k \right\|_{\infty} < \epsilon_k. \quad (5.3.16)$$

La sucesión $(\frac{\beta_k w(x^k)}{\delta_k}, \frac{\beta_k \lambda^{k+1}}{\delta_k}, \frac{\beta_k \mu^{k+1}}{\delta_k})$ es acotada pues su norma es 1. Luego, existen $\bar{K} \subset K_1$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ tales que

$$\lim_{k \in \bar{K}} \left(\frac{\beta_k w(x^k)}{\delta_k}, \frac{\beta_k \lambda^{k+1}}{\delta_k}, \frac{\beta_k \mu^{k+1}}{\delta_k} \right) = (0, \lambda, \mu),$$

siendo $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$, no todos igual a 0 y $\mu \geq 0$. Observar que $\lim_{k \in \bar{K}} \frac{\beta_k w(x^k)}{\delta_k} = 0$ ya que $\{\beta_k w(x^k)\}$ es acotada. Por lo tanto, por la continuidad de $\nabla h(x)$ y $\nabla g(x)$, tomando límite para $k \in \bar{K}$ en (5.3.16) se obtiene

$$\left\| P_{\bar{\Omega}} \left(x^* - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) - x^* \right\|_{\infty} = 0$$

que es equivalente a $-\sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \in N_{\bar{\Omega}}(x^*)$, pues

$$x^* = P_{\bar{\Omega}} \left(x^* - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right).$$

Luego x^* es solución del problema convexo

$$\min_{y \in \bar{\Omega}} \left\| y - \left(x^* - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) \right\|^2,$$

es decir existen $\underline{\lambda} \in \mathbb{R}^m$, $\underline{\mu} \in \mathbb{R}^p$, $\underline{\mu} \geq 0$, $\underline{\mu}_j g_j(x^*) = 0$

$$-x^* + \left(x^* - \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) = \sum_{i=1}^m \underline{\lambda}_i \nabla h_i(x^*) + \sum_{j=1}^p \underline{\mu}_j \nabla g_j(x^*)$$

que equivale a

$$-\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) = \sum_{i=1}^m \underline{\lambda}_i \nabla h_i(x^*) + \sum_{j=1}^p \underline{\mu}_j \nabla g_j(x^*).$$

Luego, la condición (1) de Quasinormalidad para $\bar{\Omega}$ se satisface ya que

$$-\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) \right) \in N_{\bar{\Omega}}(x^*)$$

con $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$, no todos igual a 0.

Como $\lim_{k \in \bar{K}} \frac{\lambda^{k+1}}{\delta_k} = \lambda$ y $\lambda^{k+1} = \bar{\lambda}^k + \rho_k h(x^k)$, entonces como $\{\bar{\lambda}^k\}$ es acotada $\lim_{k \in \bar{K}} \frac{\bar{\lambda}^k}{\delta_k} = 0$ y $\lim_{k \in \bar{K}} \frac{\rho_k h(x^k)}{\delta_k} = \lambda$. Luego si $\lambda_i \neq 0$ existe $k_1 \in \bar{K}$ tal que para todo $k \geq k_1$ $h_i(x^k) \lambda_i > 0$.

Como $\lim_{k \in \bar{K}} \frac{\mu^{k+1}}{\delta_k} = \mu$ y $\mu_{k+1} = \max \{0, \bar{\mu}^k + \rho_k g(x^k)\}$, entonces como $\{\bar{\mu}^k\}$ es acotada $\lim_{k \in \bar{K}} \frac{\bar{\mu}^k}{\delta_k} = 0$ y $\lim_{k \in \bar{K}} \frac{\rho_k \sigma^k}{\delta_k} = \mu$.

Luego si $\mu_j > 0$ existe $k_2 \in \bar{K}$ tal que para todo $k \geq k_2$ $\sigma_j^k \mu_j > 0$, entonces $g_j(x^k) \mu_j > 0$.

Y por lo tanto se contradice la Quasinormalidad, luego $\{\delta_k\}$ es acotada. □

En este capítulo se define una función del tipo Lagrangiana Aumentada vectorial y se define el algoritmo general del tipo Lagrangiano Aumentado que puede utilizar el método PGM del Capítulo 4 en la resolución del subproblema. Se demuestra convergencia global utilizando la definición de Quasinormalidad para el conjunto factible junto con la condición de Abadie para el conjunto de restricciones del nivel inferior o que no fueron penalizadas.

Capítulo 6

Conclusiones

En esta tesis abordamos el estudio de métodos para la resolución de problemas de optimización con restricciones.

En el Capítulo 3 estudiamos una extensión del conocido método Lagrangiano Aumentado escalar al caso vectorial. En ese capítulo demostramos que el algoritmo clásico Lagrangiano Aumentado escalar converge a un punto KKT utilizando una condición que no había sido considerada hasta este momento: Quasinormalidad. Para la extensión del método al contexto multiobjetivo la sucesión generada satisface la condición $AKKT$ para el MOP dada recientemente en [40] que es una extensión de la conocida condición $AKKT$ del caso escalar definida mediante una escalarización adecuada. La convergencia global de la sucesión a puntos regulares débiles se obtiene usando la Quasinormalidad o la condición CCP . Demostramos que la condición $AKKT$ para el MOP se puede obtener usando una clase diferenciable de funciones débilmente crecientes que tienen gradiente no nulo.

En el Capítulo 4 presentamos un nuevo algoritmo para optimización multiobjetivo con restricciones convexas. En cada iteración, la dirección de búsqueda se calcula considerando un paso variable basado en ideas espectrales del caso escalar. La novedad es el uso de una técnica de búsqueda lineal no monótona en lugar de la clásica estrategia de Armijo. Probamos que cualquier punto límite de la sucesión generada por el algoritmo propuesto es un punto estacionario del problema original. En el caso convexo probamos la convergencia de toda la sucesión a un punto Pareto optimal débil bajo suposiciones razonables.

Finalmente, en el Capítulo 5 se define una función del tipo Lagrangiana Aumentada vectorial y el algoritmo Lagrangiano Aumentado en el cual el subproblema es vectorial. El subproblema del algoritmo propuesto puede ser resuelto utilizando el algoritmo PGM propuesto en el capítulo 4. Se demuestra convergencia global a puntos regulares débiles utilizando la definición de Quasinormalidad para el conjunto

factible junto con la condición de Abadie para el conjunto de restricciones del nivel inferior o que no fueron penalizadas.

Bibliografía

- [1] Abadie, J.: On the Kuhn-Tucker Theorem. In J. Abadie, editor. *Non-linear Programming*, (1967).
- [2] Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal Optimization* 18, 1286-1309, (2007).
- [3] Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification. *Mathematical Programming* 112, 5-32, (2008).
- [4] Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. *Optimization* 60, 627-641, (2011).
- [5] Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. *SIAM Journal Optimization* 22, 1109-1135, (2012).
- [6] Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: A Cone-Continuity Constraint qualification and Algorithmic Consequences. *SIAM Journal Optimization* 26, 96-110, (2016).
- [7] Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal Optimization* 20, 3533-3554, (2010).
- [8] Andreani, R., Martínez, J.M., Schuverdt, M.L.: On the relation between Constant Positive Linear Dependence condition and Quasinormality Constraint Qualification. *Journal of Optimization Theory and Applications* 125, 473-485, (2005).
- [9] Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. *IMA Journal of Numerical Analysis* 8, 141-148, (1988).

-
- [10] Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming: Theory and Algorithms, second edition. *John Wiley Sons, New York*, (1993).
- [11] Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. *Athena Scientific*, (1996).
- [12] Bertsekas, D.P.: Nonlinear Programming, second edition. *Athena Scientific*, (1999).
- [13] Bertsekas, D.P., Ozdaglar, A.E.: Pseudonormality and a Lagrange Multiplier Theory for Constrained Optimization. *Journal of Optimization Theory and Applications* 114, 287-343, (2002).
- [14] Bigi, G.: Optimality and Lagrangian Regularity in Vector Optimization. *PhD thesis, Dipartimento di Matematica, Universita di Pisa, Italy*, (2003).
- [15] Bigi, G., Pappalardo, M.: Regularity Conditions in Vector Optimization. *Journal of Optimization Theory and Applications* 102, 83-96, (1999).
- [16] Birgin, E.G., Martínez, J.M., Raydan M.: Inexact spectral projected gradient methods on convex sets. *IMA Journal of Numerical Analysis* 23(4), 539-559, (2003).
- [17] Birgin, E.G., Martínez J.M., Raydan M.: Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal Optimization* 10(4), 1196-1211, (1999).
- [18] Birgin, E.G., Floudas, C.A., Martínez, J.M.: Global minimization using an augmented Lagrangian method with variable lower-level constraints. *Mathematical Programming* 125, 139-162, (2010).
- [19] Birgin, E.G., Martínez, J.M., Prudente, L.F.: Optimality properties of an Augmented Lagrangian method on infeasible problems. *Technical report MCDO 110214, University of Campinas, Campinas, SP, Brazil*, (2014).
- [20] Birgin, E.G., Martínez, J.M.: On the application of an Augmented Lagrangian algorithm to some portfolio problems. *EURO Journal on Computational Optimization* 4, 79-92, (2016)
- [21] Birgin, E.G, Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. *SIAM Publications. Series: Fundamentals of Algorithms, Philadelphia*, (2014).

- [22] Bonnel, H., Iusem, A.N., Svaiter, B.F.: Proximal methods in vector optimization. *SIAM Journal on Optimization* 15, 953-970, (2005).
- [23] Burachik, R., Graña Drummond, L.M., Iusem, A.N., Svaiter B.F.: Full convergence of the steepest descent method with inexact line searches. *Optimization* 32(2), 137-146, (1995).
- [24] Carrizo, G.A., Lotito, P.A., Maciel, M.C.: Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem. *Mathematical Programming* 159, 339-369, (2015).
- [25] Carrizosa, E., Frenk, J.B.G.: Dominating sets for convex functions with some applications. *Journal of Optimization Theory and Applications* 96, 281-295, (1998).
- [26] Da Cunha, N.O., Polak, E.: Constraint minimization under vector-value criteria in finite dimensional spaces. *Journal of Mathematical Analysis and Applications* 19, 103-124, (1967).
- [27] Dai, Y.H.: On the nonmonotone line search. *Journal of Mathematical Analysis and Applications* 112(2), 315-330, (2002).
- [28] Diniz-Ehrhardt, M.A., Martínez, J.M., Pedroso, L.G.: Derivative-free methods for nonlinear programming with general lower-level constraints. *Computational and Applied Mathematics* 30, 19-52, (2011).
- [29] Echebest, N., Sánchez, M.D., Schuverdt, M.L.: Convergence Results of an Augmented Lagrangian Method Using the Exponential Penalty Function. *Journal of Optimization Theory and Applications*, 92-108, (2015).
- [30] Eschenauer, H., Koski, J., Osycza, A.: Multicriteria Design Optimization. *Procedures and Applications*, Springer-Verlag, Berlin, (1990).
- [31] Evans, G.W.: An overview of techniques for solving multiobjective mathematical programs. *Management Science* 30, 1268-1282, (1984).
- [32] Fletcher, R., Practical Methods of Optimization, second edition. *J. Wiley Sons, Chichester*, (1991).
- [33] Fliege J., Graña Drummond L.M., Svaiter B.F.: Newton's method for multiobjective optimization. *SIAM Journal on Optimization* 20, 602-626 (2009).
- [34] Fukuda, E.H., Graña Drummond, L.M.: On the convergence of the projected gradient method for vector optimization. *Optimization* 60, 1009-1021, (2011).

- [35] Fukuda, E.H., Graña Drummond, L.M.: Inexact projected gradient method for vector optimization. *Computational Optimization and Applications* 54, 473-493, (2013).
- [36] Fukuda, E.H., Graña Drummond, L.M.: A survey on multiobjective descent methods. *Pesquisa Operacional* 34, 585-620, (2014).
- [37] Fukuda, E.H., Graña Drummond, L.M., Raupp, F. M. P.: An external penalty-type method for multicriteria. *TOP* 24(2), 493-513, (2016).
- [38] Gass, S., Saaty, T.: The computational algorithm for the parametric objective function. *Naval Research Logistics Quarterly* 2, 39-45, (1955).
- [39] Geoffrion A.M.: Proper efficiency and the theory of vector maximization. *Journal of Mathematical Analysis and Applications* 22, 618-630, (1968).
- [40] Giorgi, G., Jiménez, B., Novo, V.: Approximate Karush-Kuhn-Tucker condition in multiobjective optimization. *Journal of Optimization Theory and Applications* 171(1), 70-89, (2016).
- [41] Graña Drummond, L.M., Iusem, A.N.: A projected gradient method for vector optimization problems. *Computational Optimization and Applications* 28, 5-29, (2004).
- [42] Graña Drummond, L.M., Svaiter, B.F.: A steepest descent method for vector optimization. *Journal of Computational and Applied Mathematics* 175, 395-414, (2005).
- [43] Grippo, L., Lampariello, F., Lucidi, S.: A Nonmonotone Line Search Technique for Newton's Method. *SIAM Journal on Numerical Analysis* 23, 707-716, (1986).
- [44] Haeser, G., Schuverdt, M.L.: On approximate KKT condition and its extension to continuous variational inequalities. *Journal of Optimization Theory and Applications* 23, 707-716, (2011).
- [45] Hager, W.W., Zhang, H.: Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent. *ACM Transactions on Mathematical Software* 32(1), 113-137, (2006).
- [46] Hestenes, M.R.: Optimization theory: the finite dimensional case. *John Wiley*, (1975).
- [47] Leschine, T.M., Wallenius, H., Verdini, W.A.: Interactive multiobjective analysis and assimilative capacity-based ocean disposal decisions. *European Journal of Operational Research* 56, 278-289, (1992).

- [48] Luc, D.T.: Theory of vector optimization. *Lecture Notes in Economics and Mathematical Systems*, Springer, (1989).
- [49] Mangasarian, O.L., Fromovitz, S.: The Fritz-John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints. *Journal of Mathematical Analysis and Applications* 17, 37-47, (1967).
- [50] Miettinen, K.M.: Nonlinear Multiobjective Optimization, *Kluwer Academic Publishers, Boston*, (1999).
- [51] Mita, K., Fukuda, E.H., Yamashita, N.: On using nonmonotone line search techniques in steepest descent methods for multiobjective optimization. *Proceedings of the 61st Annual Conference of the Institute of Systems, Control and Information Engineers (ISCIE), Kyoto*, (2017).
- [52] Ozdaglar, A.E., Bertsekas, D.P.. The relation between pseudonormality and quasiregularity in constrained optimization, *Optimization Methods and Software* 19, 493-506, (2004).
- [53] Panier, E.R., Tits, A.L.: Avoiding the maratos effect by means of a nonmonotone line search. i: General constrained problems. *SIAM Journal on Numerical Analysis* 28(4), 1183-1195, (1991).
- [54] Pareto, V.: Cours d'Economie Politique. *F. Rouge, Lausanne, Switzerland*, (1986).
- [55] Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. *SIAM Journal on Optimization* 10, 963-981 (2000).
- [56] Qu, S., Ji, Y., Jiang, J., Zhang, Q.: Nonmonotone gradient methods for vector optimization with a portfolio optimization application. *European Journal of Operational Research* 263(2), 356-366 (2017).
- [57] Raydan, M.: On the Barzilai and Borwein choice of steplength for the gradient method. *IMA Journal of Numerical Analysis* 13, 321-326, (1993).
- [58] Raydan, M.: The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM Journal on Optimization* 7(1), 26-33, (1997).
- [59] Rockafellar, R. T., Wets, R.: Variational Analysis. *Springer, Berlin*, (1998).
- [60] Sánchez, M. D., Schuverdt, M. L.: A second order convergence Augmented Lagrangian Method using non-quadratic penalty functions, submitted, (2016).

-
- [61] Schuverdt, M. L.: Convergencia del métodos de Lagrangiano Aumentado usando la condición de dependencia linear positiva constante. *Tese de Doutorado, Departamento de Matemática Aplicada, Universidade Estadual de Campinas*, (2006).
- [62] Zhang, H., Hager, W.W.: A nonmonotone line search technique and its application to unconstrained optimization. *SIAM Journal Optimization* 14(4), 1043-1056, (2004).