

Aproximación Numérica Equiespaciada de la Variedad
y el Frente de Pareto para Problemas de Optimización
Multiobjetivo.

Autor: Claudio Javier Tablada
Director: Germán Ariel Torres

Facultad de Matemática, Astronomía y Física
Universidad Nacional de Córdoba
Córdoba, Argentina

5 de Agosto de 2011

Índice general

1. Introducción	1
1.1. Un poco de historia	2
1.2. Optimización simple o de un objetivo	4
1.2.1. Programación lineal	5
1.2.2. Programación no lineal	8
2. Optimización multiobjetivo	14
2.1. Conos y órdenes relacionados	17
2.2. Soluciones eficientes y puntos no dominados	19
2.2.1. Cotas del conjunto no dominado	30
2.2.2. Caracterización geométrica de las soluciones eficientes	31
2.3. Eficiencia y no dominancia propia	34
2.4. Caracterización de soluciones eficientes	41
2.4.1. Caracterización por escalarización	41
2.4.2. Caracterización como solución de problemas con restricciones (ε -Constraint)	51
2.5. Conectividad de conjuntos no dominados	51
2.6. Condiciones de optimalidad multiobjetivo	54
3. Método de continuación global	59
3.1. Cálculo de la variedad y el frente de Pareto	60
3.2. Generalización a problemas con más de dos objetivos	65
3.3. Problemas de testeo	66
3.4. Conclusiones	71

Agradecimientos

Mi más profundo agradecimiento a mi familia por el apoyo incondicional que me ha brindado, a mis profesores por su dedicación y paciencia, y a mis compañeros porque también he aprendido mucho de ellos.

Resumen

En este trabajo se presenta un algoritmo que tiene aplicación en problemas de optimización multiobjetivo convexo irrestricto. Este método de continuación global, desarrollado en [16], hace uso del método de Newton y de restricciones adicionales de equiespaciado para proporcionar una aproximación numérica uniformemente espaciada de la variedad de Pareto (conjunto de soluciones eficientes) o del frente de Pareto (conjunto de puntos no dominados). Las ventajas que presenta el método son su bajo costo de implementación, el muestreo uniforme obtenido de los puntos óptimos y la posible paralelización del procedimiento computacional. Se dan ejemplos aplicados a funciones de testeo para ver la performance del método.

Para motivar esta presentación se introduce el concepto de optimización simple y posteriormente se dan ejemplos donde los objetivos de interés están en conflicto, lo cual hace imposible, sin información adicional, definir una única solución óptima. Puesto que se considera optimalidad en el sentido de Pareto, se define eficiencia y nodominancia de Pareto junto con los principales resultados teóricos de optimización multiobjetivo.

Capítulo 1

Introducción

La teoría de optimización¹ intenta dar respuesta a un tipo general de problemas donde se desea elegir la mejor opción entre un conjunto de elementos. La optimización simple o de un objetivo consiste en encontrar los valores máximos o mínimos de una función objetivo, y la optimización multiobjetivo (donde se tienen en cuenta varias funciones objetivos) consiste en encontrar el conjunto de puntos que, de acuerdo a un orden pre-establecido, son preferidos al resto y tal que ninguno de ellos dentro de este conjunto es preferido uno del otro.

Además de su interés teórico implícito, en la actualidad la teoría de optimización tiene gran relevancia debido a sus aplicaciones en áreas tan diversas como ingeniería, economía, industria y ciencia, donde muchas veces se plantean problemas tales como minimizar o maximizar ciertas cantidades de acuerdo a objetivos especificados. Analistas, ingenieros, managers y científicos confrontan todo el tiempo con problemas de esta naturaleza. Tales problemas pueden ser de diversos tipos, desde considerar diseños y planeamientos de operaciones industriales óptimas hasta calcular la trayectoria óptima que debe recorrer un módulo espacial para economizar combustible. En el pasado, un amplio rango de soluciones a muchos de estos problemas se consideraron aceptables, pero en la actualidad y debido en parte a la gran competencia existente y a la complejidad cada vez mayor de tales problemas, una solución aceptable no es suficiente y es necesario rever éstos para llevar la optimización de los mismos a estándares más elevados.

A lo largo de este informe plantearemos en primer lugar el problema de optimización simple (*Bazaraa et al.* [1], *Hestenes* [9], *Mangasarian* [12], *Martínez et al.* [13]) en sus dos versiones: *irrestringido* y *con restricciones*. La introducción al problema de optimización multiobjetivo (*Coello et al.* [2], *Collette et al.* [3], *Ehrgott* [7], *Eichfelder* [8], *Jahn* [10], *Miettinen* [14], *Nakayama et al.* [15], *Sawaragi et al.* [19]) será motivada a través de ejemplos, los cuales, al disponer de objetivos en conflicto, muestran que para tales planteamientos la búsqueda de una única solución es generalmente imposible, obteniéndose en este caso un *conjunto de soluciones*, las cuales (y a falta de otra información) son igualmente preferidas. De esta forma, elegir una única solución de este conjunto implica realizar un balance agregado o compromiso entre las soluciones (comúnmente conocido en la literatura como *trade-off*), el cual no consideraremos en este trabajo. Al igual que en el caso de un objetivo, el problema multiobjetivo puede ser clasificado en *irrestringido* y *con restricciones*. Varios temas y resultados teóricos respecto de este conjunto de soluciones son también considerados.

Finalmente y como tema central, introduciremos un método de continuación global (*Pereyra* [16]) el cual permite obtener una aproximación numérica equiespaciada de la variedad de Pareto o el frente de Pareto para problemas de optimización multiobjetivos convexos irrestringidos. Este método se desarrolla en primer lugar para el caso bi-objetivo, proponiéndose *a posteriori* su

¹Al conjunto de teoremas, algoritmos, métodos y técnicas para resolver problemas de optimización se lo conoce como *programación matemática*.

extensión al caso multiobjetivo general. Tal algoritmo hace uso del método de Newton y considera restricciones de espaciado uniforme, características éstas que lo convierten en un método novedoso respecto de otros métodos por su bajo costo computacional y obtención uniforme de los puntos óptimos.

1.1. Un poco de historia

Si bien el desarrollo del cálculo fue fundamental para desarrollar la teoría de optimización actual, los primeros problemas de optimización datan de la antigüedad². El geómetra griego Euclides (300 a. C.) consideró el problema de hallar la distancia mínima entre un punto y una recta, y probó que entre todos los rectángulos con un cierto perímetro fijo, el cuadrado es la figura que maximiza el área. Zenodorus (200 a. C.) estudió el *problema de Dido*, el cual consiste en hallar, de entre todas las curvas planas cerradas que tienen una longitud fija, la que encierra una mayor área en su interior (la solución a tal problema es un semicírculo). Heron (100 a. C.) propuso en su obra *Catóptrica* que la luz viaja por el camino geoméricamente más corto (hoy se sabe que esto no es cierto, según el principio de Fermat).

En los siglos XVII y XVIII algunos problemas aislados de optimización fueron estudiados. Kepler (1615) halla las dimensiones óptimas para un barril de vino y formula por primera vez el *problema de la secretaria* (seleccionar la mejor postulante entre un total de m candidatas, asumiendo ciertas suposiciones para realizar la elección), el cual es una aplicación clásica de programación dinámica. Fermat (1646) muestra que en un punto extremo de una función su gradiente se anula. En 1657 también establece el *principio de Fermat para óptica*, el cual muestra que el trayecto seguido por la luz al propagarse de un punto a otro es tal que el tiempo empleado en recorrerlo es un mínimo. En este período, Newton (1660) y Leibniz (1670) crean el *análisis matemático*, el cual es la base de la teoría de optimización. Los hermanos Johann y Jacob Bernoulli (1696) estudian el problema de la *curva braquistócrona* (curva entre dos puntos que es recorrida en el menor lapso de tiempo por un cuerpo que comienza en el punto inicial con velocidad cero y que debe desplazarse a lo largo de la curva hasta llegar al segundo punto, bajo acción de una fuerza de gravedad constante y suponiendo que no existe fricción). Knig (1712) muestra que la forma de un panal es óptima. Euler (1740) incursiona por primera vez en la teoría general del *cálculo de variaciones*, el cual es la base teórica de optimización en dimensión infinita. Maupertuis (1746) formula el *principio de mínima acción* para explicar fenómenos físicos. Lagrange (1760) formula el *problema de superficies mínimas* (problema de Plateau). Monge (1784) investiga un problema de optimización combinatoria conocido como el *problema del transporte* (caso particular de un problema de programación lineal en el cual se debe minimizar el costo del abastecimiento a una serie de puntos de demanda a partir de un grupo de puntos de oferta, teniendo en cuenta los distintos precios de envío de cada punto de oferta a cada punto de demanda).

En el siglo XIX se obtienen los primeros algoritmos sobre optimización y la teoría del cálculo de variaciones es desarrollada principalmente por Weierstrass, Steiner, Hamilton y Jacobi. Legendre (1806) presenta el *método de mínimos cuadrados*. La idea de una función (quasi)cóncava aparece en economía alrededor de 1815 en trabajos realizados por Malthus, Torrens, West y Ricardo. Fourier (1826) formula el problema de *programación lineal* para tratar problemas en mecánica y teoría de probabilidad. Faustmann (1846) propone su modelo de optimización para producción forestal. Cauchy (1847) presenta el *método del gradiente*. Gibbs (1857) propone que el equilibrio químico se alcanza cuando la *energía de Gibbs* del sistema está en su valor mínimo. En 1870, debido a los trabajos de Jevons, Menger y Walras, surge la revolución marginalista en

²fuentes <http://hse-econ.fi/kitti/opthist.html>

economía, la cual aplica procedimientos de maximización a las diferentes variables económicas razonando en el margen, es decir, sobre la última unidad del bien consumido, producido, intercambiado o retenido.

En el siglo XX se logran avances en el cálculo de variaciones principalmente por Bolza, Bliss y Caratheodory. Farkas (1902) presenta el *lema de Farkas*, el cual es usado en la prueba del teorema de Karush-Kuhn-Tucker. Jensen (1905) introduce el concepto de *funciones convexas* y Minkowski (1911) obtiene los primeros resultados sobre *conjuntos convexos*. Hancock (1917) publica el primer libro de texto sobre optimización, *Theory of Mínima and Máxima* (Teoría de Mínimo y Máximo). El biomatemático Thompson (1917) escribe su libro *On Growth and Form* (Sobre Crecimiento y Forma), en el cual aplica optimización para analizar las formas de organismos vivos. Morse (1925) presenta su teoría que generaliza el cálculo de variaciones. Ramsey (1928) aplica el cálculo de variaciones en su estudio sobre el crecimiento económico óptimo. Menger (1932) presenta una formulación general del *problema del viajante*, el cual consiste en encontrar un recorrido de longitud mínima para un viajante que debe visitar varias ciudades y volver al punto de partida, conocida la distancia existente entre dos ciudades. Kantorovich (1939) presenta un modelo de programación lineal y un algoritmo para resolverlo. Después de la segunda guerra mundial, el desarrollo de la teoría de optimización se expande simultáneamente con la *investigación operativa* (uso efectivo de la tecnología por las organizaciones). Von Neumann y Morgenstern (1944) resuelven problemas de decisión secuenciales al usar la idea de *programación dinámica*. Massé (1944) presenta otra aplicación de la programación dinámica al problema de manejo de reservas. Dantzig (1947) presenta el *método simplex* para resolver problemas de programación lineal y Von Neumann establece la *teoría de dualidad* para tales problemas. En 1949 se realiza el primer simposio internacional sobre programación matemática. Kuhn y Tucker (1951) reinventan las condiciones de optimalidad para problemas no lineales (Karush ya había presentado en 1939 condiciones similares). Markowitz (1951) presenta su *teoría de portafolio* (teoría de inversión que trata de maximizar el retorno y minimizar el riesgo, mediante la cuidadosa elección de los componentes). Las investigaciones sobre problemas de redes llevadas a cabo por Ford y Fulkerson (1954) dan origen a la investigación de *optimización combinatoria*. En este período también son desarrollados los *métodos de quasi-Newton* y de *gradiente conjugado* para problemas irrestrictos y la *teoría de control optimal* comienza su desarrollo como una disciplina separada del cálculo de variaciones, incentivada en parte por el comienzo de la era espacial. En 1954 se funda la *IEEE Control Systems Society* (organización internacional dedicada al avance de investigaciones, desarrollo y práctica en automatización y sistemas de control). Pontryagin (1956) presenta su *principio del máximo* y Bellman (1957) presenta el *principio de optimalidad*. Zoutendijk (1960) presenta los *métodos de direcciones factibles* con la idea de generalizar el método simplex a programas no lineales. Wilson (1963) presenta el *método de programación cuadrática secuencial*. En 1973 se funda la *Mathematical Programming Society* (actualmente la *Mathematical Optimization Society*), dedicada a la promoción y mantenimiento de estándares profesionales en el tema de optimización matemática. Karmarkar (1984) introduce el *algoritmo de Karmarkar* para resolver problemas de programación lineal, produciéndose a posteriori una proliferación de los *métodos de puntos interiores*.

En la actualidad, con el advenimiento de las computadoras modernas, ha sido posible implementar algoritmos para optimización global y resolver problemas de gran escala, lo que hace de la teoría de optimización una herramienta importante para plantear y resolver problemas en diversos campos.

1.2. Optimización simple o de un objetivo

Un problema de optimización simple puede plantearse como

$$\min_{x \in \mathcal{X}} f(x), \quad (1.1)$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función a valores reales (función objetivo) y $\mathcal{X} \subseteq \mathbb{R}^n$. Si existe $x^* \in \mathcal{X}$ tal que la desigualdad $f(x^*) \leq f(x)$ se cumple para todo x en \mathcal{X} , entonces x^* es una solución del problema (1.1). En este caso se dice que x^* es un punto de mínimo (global) de f en \mathcal{X} (o que es un punto óptimo) y que $f(x^*)$ es el valor mínimo (o valor óptimo) correspondiente. Si existe un entorno \mathcal{N} de x^* tal que la desigualdad $f(x^*) \leq f(x)$ se cumple para todo x en $\mathcal{N} \cap \mathcal{X}$, entonces se dice que x^* es un punto de mínimo *local* de f en \mathcal{X} .

Obsérvese que este planteamiento también considera el problema

$$\max_{x \in \mathcal{X}} f(x) = \min_{x \in \mathcal{X}} \{-f(x)\},$$

y por lo tanto el problema de hallar los puntos máximos de f en \mathcal{X} es equivalente a encontrar los puntos mínimos de $-f$ en \mathcal{X} .

Por otra parte, se dice que x^* es un punto de mínimo *irrestringido* de f en \mathcal{X} si $x^* \in \mathcal{X}^\circ$ (interior de \mathcal{X}), mientras que x^* es un punto de mínimo *restringido* de f en \mathcal{X} si $x^* \in \partial\mathcal{X}$ (frontera de \mathcal{X}). Esta última definición permite dividir el problema (1.1) en dos tipos:

1) Problema *irrestringido*

$$\min_{x \in \mathcal{X}} f(x), \quad (1.2)$$

donde \mathcal{X} es un abierto en \mathbb{R}^n (posiblemente $\mathcal{X} = \mathbb{R}^n$).

2) Problema *con restricciones*

$$\min_{x \in \mathcal{X}} f(x), \quad (1.3)$$

donde $\mathcal{X} \subsetneq \mathbb{R}^n$ no es un abierto.

En la mayoría de las aplicaciones el conjunto \mathcal{X} en (1.3) es definido por restricciones de igualdad y desigualdad de la forma

$$\begin{aligned} g_\alpha(x) &\leq 0, \quad \alpha = 1, \dots, l, \\ g_\beta(x) &= 0, \quad \beta = l + 1, \dots, m, \end{aligned}$$

para ciertas funciones $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ($i = 1, \dots, m$), o por restricciones de caja

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid l_i \leq x_i \leq u_i, \quad l_i, u_i \in \mathbb{R}, \quad i = 1, \dots, n\}.$$

El conjunto \mathcal{X} es conocido como el *conjunto factible* del problema de optimización. Con excepción del caso lineal, el problema (1.3) es mucho más complejo que el problema (1.2). Una revisión más detallada de (1.2) y (1.3) puede encontrarse en la literatura mencionada.

1.2.1. Programación lineal

En el caso de que la función objetivo y las restricciones en el problema (1.3) son lineales, obtenemos un problema de *programación lineal*. Este tipo de problema es el que mejor se ha desarrollado dentro de la teoría de optimización general. Si la función objetivo f es lineal y estamos considerando m funciones de restricción lineales, entonces podemos plantear el problema (1.3) con $f(x) = c^T x$ ($c \in \mathbb{R}^n$) y el conjunto factible \mathcal{X} definido por restricciones de igualdad o desigualdad lineales.

Un problema de programación lineal puede entonces ser planteado como

$$\text{Minimizar } c^T x \text{ sujeto a } Ax \geq b, x \geq 0, \quad (1.4)$$

para el caso de restricciones de desigualdad, o como

$$\text{Minimizar } c^T x \text{ sujeto a } Ax = b, x \geq 0, \quad (1.5)$$

para el caso de restricciones de igualdad, donde A es una matriz $m \times n$ (suponemos $m < n$) y $b \in \mathbb{R}^m$. Obsérvese que se pide además que el vector x sea no negativo, un requerimiento importante en problemas de este tipo.

El problema (1.5) correspondiente a restricciones de igualdad se conoce como un problema lineal estándar. Todo problema de la forma (1.4) puede ser transformado a un problema estándar al introducir la *variable de holgura* $w = Ax - b$. De esta manera obtenemos el nuevo sistema:

$$(A \quad -I) \begin{pmatrix} x \\ w \end{pmatrix} = b$$

donde I denota la matriz identidad. Las restricciones en el nuevo sistema son $x \geq 0$, $w \geq 0$, las cuales corresponden perfectamente con el problema estándar (1.5).

Ilustraremos estas ideas con un ejemplo sencillo.

Problema de la dieta (Strang [20], cap. 8). Supongamos que estamos realizando una dieta con dos fuentes de proteínas: A y B. Cada porción del alimento A proporciona una unidad de proteína y la misma porción del alimento B proporciona dos unidades de proteína, y se requieren al menos cuatro unidades de proteína en la dieta. Por lo tanto, una dieta que contenga x porciones del alimento A e y porciones del alimento B está restringida por $x + 2y \geq 4$ y también por $x \geq 0$ e $y \geq 0$. Si cada porción del alimento A cuesta \$2 y la misma porción de B cuesta \$3, entonces el costo total de la dieta es $2x + 3y$. El problema es entonces proporcionar las cantidades x e y de tales alimentos que satisfagan las restricciones impuestas y que minimicen el costo total de la dieta.

En términos de (1.4) tenemos que

$$f(x, y) = (2 \quad 3) \begin{pmatrix} x \\ y \end{pmatrix} = 2x + 3y$$

$$\mathcal{X} = \{(x, y) \in \mathbb{R}^2 \mid x + 2y \geq 4, x \geq 0, y \geq 0\}$$

donde $c^T = (2 \quad 3)$ y $A = (1 \quad 2)$. El problema de la dieta se muestra en la figura 1.1. El conjunto \mathcal{X} se muestra como una región sombreada y se conoce como el *conjunto factible* del problema lineal. Nótese que $f(x, y) = d$ resulta ser una recta en \mathbb{R}^2 para todo $d \in \mathbb{R}$. Si $d_1 < d_2$,

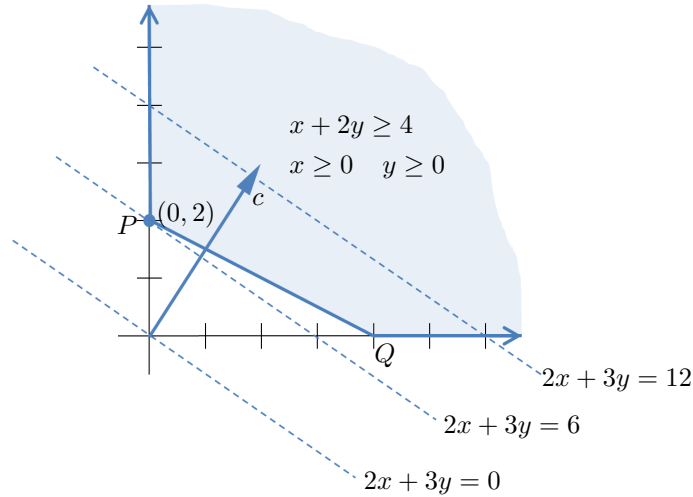


Figura 1.1: Problema de la dieta.

entonces la recta $f(x, y) = d_2$ se encuentra por encima de la recta $f(x, y) = d_1$ (todas son paralelas). Encontrar la solución (x^*, y^*) a este problema es equivalente a encontrar el menor valor $d^* \in \mathbb{R}$ tal que la recta $f(x^*, y^*) = d^*$ intersekte al conjunto \mathcal{X} . Como puede observarse en la figura 1.1, la solución óptima al problema de la dieta es el punto $(0, 2)$ correspondiente al costo $2x + 3y = 6$, el cual coincide con el vértice P del conjunto factible. Esta solución es factible porque se encuentra en \mathcal{X} , es óptima porque minimiza la función objetivo f y el costo 6 se conoce como el valor mínimo del programa lineal. Por lo tanto, la dieta óptima consiste en ingerir 2 porciones del alimento B y nada del alimento A. También podemos observar de la figura que esta solución es única y se encuentra en la frontera del conjunto factible. Esto está garantizado por la geometría, puesto que las rectas dadas por la función objetivo se mueven paralelamente hacia arriba. Si las rectas hubieran sido paralelas a la arista PQ , entonces existirían infinitas soluciones al problema lineal. \square

Como podemos observar en el problema de la dieta, la solución óptima a tal problema se encuentra en un vértice del conjunto factible. De acuerdo al teorema fundamental de programación lineal, ésta es la regla general para programas lineales que tienen solución. Antes de enunciar este teorema, necesitamos introducir una definición formal de vértice del conjunto factible. Para tal fin consideremos el problema estándar (1.5) y supondremos que la matriz A tiene rango m . Entonces podemos elegir de las n columnas de A una cantidad de m columnas linealmente independientes. Por simplicidad supongamos que elegimos las primeras m columnas de A y denotamos por B a la matriz $m \times m$ formada por estas columnas. Entonces B es no singular y por lo tanto podemos resolver de manera única la ecuación

$$Bx_B = b \tag{1.6}$$

donde $x_B \in \mathbb{R}^m$. Al definir $x = (x_B, 0)$ como el vector en \mathbb{R}^n con las primeras m componentes iguales a las de x_B y las restantes $n - m$ componentes igualadas a cero, obtenemos una solución a la ecuación $Ax = b$. De acuerdo a esto damos la siguiente definición.

Definición 1.2.1. Dado el programa lineal estándar (1.5), sea B la matriz no singular $m \times m$ formada por m columnas de A linealmente independientes y sea $x_B \in \mathbb{R}^m$ la solución a la ecuación (1.6). Si las $n - m$ componentes de $x \in \mathbb{R}^n$ no asociadas a las columnas de B son

igualadas a cero y las restantes m componentes coinciden con las de x_B , entonces se dice que x es un punto factible básico de (1.5).

Estos puntos factibles básicos corresponden a los vértices del conjunto factible \mathcal{X} . El siguiente teorema sugiere una manera directa de hallar las soluciones de un programa lineal estándar.

Teorema 1.2.1. (*Teorema fundamental de programación lineal*) Consideremos el programa lineal estándar (1.5), donde supondremos que la matriz A tiene rango m . Si existe una solución óptima del problema (1.5), entonces existe una solución óptima que es factible básica.

La demostración del teorema es dada en Luenberger et al. [11].

Del teorema anterior vemos que en principio podemos resolver tales problemas lineales encontrando todos los vértices del conjunto factible y calculando sus costos. El vértice con menor costo es una solución x^* al problema de programación lineal. Puesto que encontramos un vértice del conjunto factible al seleccionar m columnas de A linealmente independientes de un total de n columnas, entonces esto nos da una cantidad de $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ vértices.

Pero realizar esto en la práctica sería imposible, pues este número podría ser muy grande. Un método novedoso para resolver tales problemas fué propuesto por George Dantzig en 1947 y se conoce como el *método simplex* (Dantzig et al. [5], cap. 3). Este método es considerado una de las ideas más brillantes en matemática computacional. El mismo consta de dos fases:

- *Fase I.* Encontrar un vértice del conjunto factible \mathcal{X} .
- *Fase II.* Ir de vértice en vértice a lo largo de aristas que garanticen un decrecimiento en el costo (valor de la función objetivo f).

La parte importante del método radica en la fase II. En cada vértice podemos escoger cualquiera de las aristas que concurren a éste; algunas de estas aristas nos alejan del punto x^* óptimo pero desconocido y otros nos acercan al mismo. El método elige continuar el recorrido por una arista que garantice que el costo decrece; esa arista conducirá a otro vértice con un costo menor, y así sucesivamente. En algún momento se alcanzará un vértice tal que todas las aristas que concurren al mismo provocan un incremento en el costo. Este vértice es el punto x^* óptimo buscado.

Haremos un último comentario sobre programas lineales. Un concepto importante en programación lineal es el de *dualidad*, el cual resulta fundamental en aplicaciones. Cada problema en programación lineal tiene asociado un problema dual. Si el problema original es una minimización, entonces su dual es una maximización y la solución de uno coincide con la solución del otro. En el caso de (1.4), su dual se define como

$$\text{Maximizar } yb \text{ sujeto a } y \geq 0, yA \leq c \tag{1.7}$$

con y un vector fila en \mathbb{R}^m . Como podemos ver de (1.7), el problema dual comienza con los mismos A , b y c del problema original (1.4), pero invierte todo; lo único que permanece es la restricción de no negatividad $y \geq 0$. Volviendo al problema de la dieta, así como el consumidor debe calcular qué cantidades tiene que comprar de los alimentos A y B para obtener proteína a un costo mínimo, el problema dual lo enfrenta el farmacéutico que vende proteína sintética; él quiere maximizar el precio p de su producto, pero ese precio está sujeto a restricciones lineales debido a que tiene que competir con los alimentos A y B. Si quiere vender su producto, éste

no debe costar más que la proteína proporcionada por el alimento A, la cual cuesta \$2 la unidad (es decir $p \leq 2$) o que la proteína proporcionada por el alimento B, la cual cuesta \$3 las dos unidades (es decir $2p \leq 3$). Además es claro que el precio debe ser no negativo ($p \geq 0$). Como los requerimientos de la dieta son de cuatro unidades de proteína, entonces el ingreso del farmacéutico será de $4p$, y el problema dual de la dieta de acuerdo a (1.7) (con $y = p$) puede ser planteado como

$$\text{Maximizar } 4p \text{ sujeto a } p \geq 0, p \begin{pmatrix} 1 \\ 2 \end{pmatrix} \leq \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Obviamente, las restricciones pueden reducirse a $0 \leq 2p \leq 3$ (ó $0 \leq p \leq \frac{3}{2}$) y el precio máximo de la proteína sintética es de \$1.50. Por lo tanto, el ingreso máximo del farmacéutico es de $4p = \$6$ por cada venta. Este valor coincide con el costo mínimo del problema de la dieta y el consumidor termina pagando el mismo precio por la proteína natural y la sintética.

Una descripción más detallada sobre programación lineal puede encontrarse en *Luenberger et al.* [11] y *Dantzig et al.* [5].

1.2.2. Programación no lineal

El gran desarrollo que han tenido los algoritmos y el software para resolver problemas de programación lineal hacen de ésta una herramienta importante para resolver problemas en diversas áreas. No obstante, muchos problemas reales no pueden ser representados o aproximados adecuadamente como un programa lineal debido a la naturaleza no lineal de la función objetivo o de las funciones de restricción.

Un problema de programación no lineal puede ser planteado como en (1.2) para el caso irrestricto o como en (1.3) para el caso con restricciones. En esta sección daremos un resumen de programación no lineal y mencionaremos algunos resultados teóricos que resultarán de utilidad más adelante.

A modo de ilustración consideremos el siguiente problema sencillo (*Bazaraa et al.* [1], cap. 1).

Ejemplo. Hallar la distancia entre el punto $(3, 2)$ en \mathbb{R}^2 y el conjunto delimitado por las ecuaciones $y = x^2 - 3$, $y = 1$ y $x = 0$ (figura 1.2).

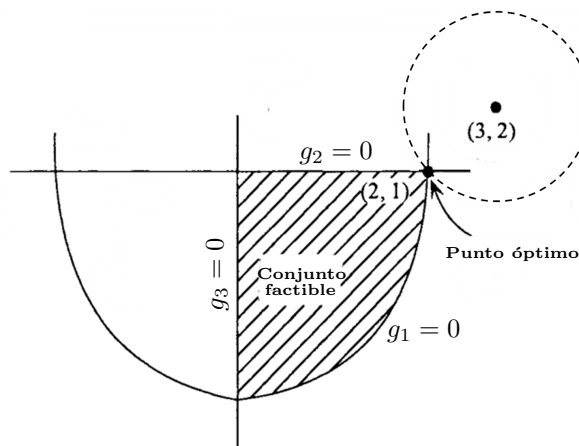


Figura 1.2: Ejemplo de programación no lineal con restricciones.

Este problema puede ser planteado como el siguiente programa no lineal con restricciones:

Minimizar

$$f(x, y) = (x - 3)^2 + (y - 2)^2$$

sujeto a

$$g_1(x, y) = x^2 - y - 3 \leq 0,$$

$$g_2(x, y) = y - 1 \leq 0,$$

$$g_3(x, y) = -x \leq 0.$$

La figura 1.2 ilustra el problema. Nótese que para toda constante $r > 0$, la ecuación $f(x, y) = r^2$ representa un círculo de radio r con centro en el punto $(3, 2)$. Puesto que queremos minimizar f , debemos encontrar el menor valor de r tal que el círculo $f(x, y) = r^2$ intersekte el conjunto factible $\mathcal{X} = \{(x, y) \in \mathbb{R}^2 \mid g_i(x, y) \leq 0, i = 1, 2, 3\}$. Como podemos observar de la geometría de la figura, un círculo con el menor radio posible debe intersectar al conjunto factible en el punto $(2, 1)$. Por lo tanto, la solución óptima al programa no lineal es $(x^*, y^*) = (2, 1)$, el cual proporciona un valor mínimo (distancia del punto $(3, 2)$ al conjunto \mathcal{X}) igual a $(2 - 3)^2 + (1 - 2)^2 = 2$. \square

Naturalmente, la tendencia geométrica usada en el ejemplo anterior sólo resulta útil para problemas sencillos, pero resulta poco práctico para problemas con más de dos variables o con funciones objetivos o de restricción más sofisticadas.

Los principales resultados teóricos respecto de programación no lineal se basan en determinar condiciones de primer y segundo orden necesarias y suficientes para la existencia de puntos óptimos. En los siguientes teoremas supondremos que la función objetivo involucrada es de clase C^2 en un entorno del punto $x^* \in \mathcal{X}$.

Teorema 1.2.2. *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función a valores reales, \mathcal{X} un abierto en \mathbb{R}^n y supongamos que x^* es un punto de mínimo (local) de f en \mathcal{X} . Entonces*

$$\nabla f(x^*) = 0 \quad (\text{condición necesaria de primer orden})$$

y para todo vector $h \neq 0$ en \mathbb{R}^n se cumple que

$$h^T \nabla^2 f(x^*) h \geq 0 \quad (\text{condición necesaria de segundo orden})$$

Recíprocamente, si x^* es un punto en \mathcal{X} tal que

$$\nabla f(x^*) = 0 \quad (\text{condición suficiente de primer orden})$$

y para todo vector $h \neq 0$ en \mathbb{R}^n se cumple que

$$h^T \nabla^2 f(x^*) h > 0 \quad (\text{condición suficiente de segundo orden})$$

entonces existen números positivos δ y m tales que

$$f(x) \geq f(x^*) + m \|x - x^*\|^2 \quad \text{para todo } x \text{ tal que } \|x - x^*\| < \delta$$

y por lo tanto x^* es un punto de mínimo local de f en \mathcal{X} .

La demostración del teorema es dada en *Hestenes* [9] (Teoremas 3.1 y 3.2) y en *Martínez et al.* [13] (Corolario 2.1.9 y Teorema 2.1.10). Aquí usamos la notación

$$\nabla f(x^*) = \left(\frac{\partial f}{\partial x_1}(x^*), \dots, \frac{\partial f}{\partial x_n}(x^*) \right)^T$$

para denotar el gradiente de f en x^* ,

$$\nabla^2 f(x^*) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x^*) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x^*) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^*) \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x^*) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x^*) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x^*) \end{pmatrix}$$

para denotar la matriz Hessiana de f en x^* , y

$$\|x\| = \left(\sum_{i=1}^n x_i^2 \right)^{1/2}$$

para denotar la norma euclídea en \mathbb{R}^n . Nótese que la condición necesaria de segundo orden en el teorema nos dice que $\nabla^2 f(x^*)$ resulta definida no negativa y la condición suficiente de segundo orden establece que $\nabla^2 f(x^*)$ debe ser definida positiva.

La teoría de optimización se simplifica considerablemente para el caso convexo.

Definición 1.2.2.

- a) Un conjunto $\mathcal{X} \subseteq \mathbb{R}^n$ se dice convexo si para dos puntos cualesquiera $x, y \in \mathcal{X}$ y para todo escalar $0 \leq \alpha \leq 1$ se cumple que $(1 - \alpha)x + \alpha y \in \mathcal{X}$.
- b) Si $\mathcal{X} \subseteq \mathbb{R}^n$ es convexo, entonces una función $f : \mathcal{X} \rightarrow \mathbb{R}$ se dice convexa si se cumple que

$$f((1 - \alpha)x + \alpha y) \leq (1 - \alpha)f(x) + \alpha f(y) \quad \forall x, y \in \mathcal{X}, 0 \leq \alpha \leq 1$$

Si sólo se cumple la desigualdad estricta, entonces decimos que f es estrictamente convexa.

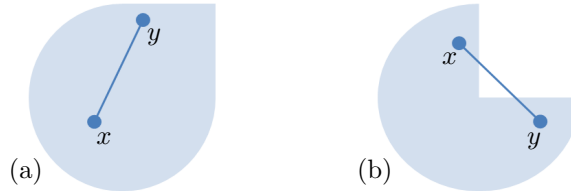


Figura 1.3: (a) Conjunto convexo. (b) Conjunto no convexo.

La definición 1.2.2 establece que un conjunto es convexo si para dos puntos cualesquiera x e y en el conjunto se cumple que éste contiene al segmento que une ambos puntos (el cual puede parametrizarse como $(1 - \alpha)x + \alpha y$, $0 \leq \alpha \leq 1$), como se muestra en la figura 1.3(a). Un ejemplo de función estrictamente convexa se muestra en la figura 1.4.

A continuación se introducen dos teoremas que resultarán ser de utilidad más adelante.

Teorema 1.2.3. Sea $\mathcal{X} \subseteq \mathbb{R}^n$ un conjunto convexo y abierto y sea $f : \mathcal{X} \rightarrow \mathbb{R}$ una función convexa. Si x^* es un punto de mínimo local de f en \mathcal{X} , entonces x^* es también un punto de mínimo global. El conjunto \mathcal{D} de puntos mínimos de f en \mathcal{X} es convexo. Si f es estrictamente convexa, entonces \mathcal{D} contiene a lo sumo un punto de mínimo f en \mathcal{X} .

Las hipótesis del teorema 1.2.2 pueden ser ampliamente debilitadas para el caso convexo.

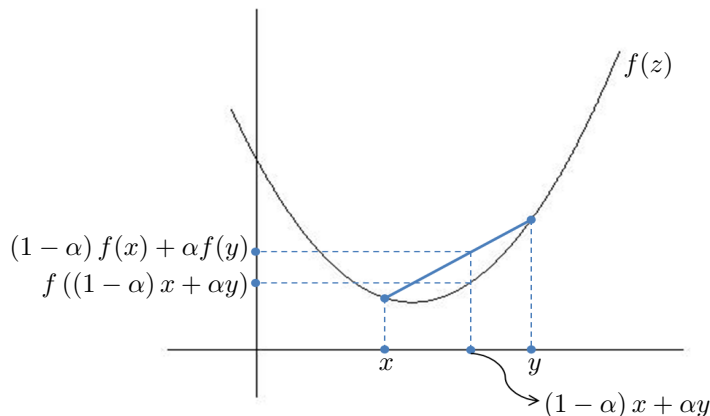


Figura 1.4: Ejemplo de función estrictamente convexa $f : \mathbb{R} \rightarrow \mathbb{R}$.

Teorema 1.2.4. Sea $\mathcal{X} \subseteq \mathbb{R}^n$ un conjunto convexo y sea $f : \mathcal{X} \rightarrow \mathbb{R}$ una función convexa. Entonces x^* es un punto de mínimo irrestricto de f en \mathcal{X} si y sólo si $\nabla f(x^*) = 0$.

Las demostraciones de los teoremas 1.2.3 y 1.2.4 son dadas en *Hestenes* [9] (teoremas 5.1 y 5.3). Una demostración del teorema 1.2.3 es también dada en *Martínez et al.* [13] (teorema 3.1.17).

Los teoremas mencionados anteriormente sólo consideran el caso en el que \mathcal{X} es un abierto en \mathbb{R}^n (es decir, el caso de puntos de mínimo irrestrictos). A continuación daremos algunos resultados teóricos para un programa no lineal con restricciones del tipo (1.3), donde el conjunto factible \mathcal{X} está dado por restricciones de igualdad y desigualdad.

Sea entonces $\mathcal{X} \subsetneq \mathbb{R}^n$ definido por las restricciones

$$g_\alpha(x) \leq 0, \quad \alpha = 1, \dots, l, \quad (1.8a)$$

$$g_\beta(x) = 0, \quad \beta = l + 1, \dots, m, \quad (1.8b)$$

para ciertas funciones $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ($i = 1, \dots, m$) las cuales son, en general, no lineales. Si $l = 0$, entonces sólo tenemos restricciones de igualdad y si $l = m$, entonces sólo tenemos restricciones de desigualdad. Consideraremos un análisis local de estas restricciones en un punto x^* en \mathcal{X} . Supondremos que estas funciones de restricción son de clase C^1 en un entorno de x^* cuando consideremos derivadas de primer orden y de clase C^2 en un entorno de x^* cuando involucremos derivadas de segundo orden. También asumiremos que x^* no es un punto aislado de \mathcal{X} , es decir, para todo entorno \mathcal{N} de x^* se cumple que $(\mathcal{N} \cap \mathcal{X}) \setminus \{x^*\} \neq \emptyset$.

Haremos una distinción entre restricciones activas e inactivas en (1.8). Una restricción del tipo (1.8a) se dice que es *activa* en x^* si $g_\alpha(x^*) = 0$ y se dice *inactiva* si $g_\alpha(x^*) < 0$. De manera similar, diremos que un índice α es activo si la restricción correspondiente a ese índice es activa en el punto x^* y diremos que es inactivo en caso contrario. Naturalmente, todas las restricciones del tipo (1.8b) son activas en un punto $x^* \in \mathcal{X}$ y por lo tanto también lo son los índices $\beta = l + 1, \dots, m$.

Definimos las *restricciones tangenciales* en el punto x^* por

$$\langle \nabla g_\alpha(x^*), h \rangle \leq 0, \quad \alpha = \alpha_1, \dots, \alpha_r \quad (\alpha_j \leq l), \quad (1.9a)$$

$$\langle \nabla g_\beta(x^*), h \rangle = 0, \quad \beta = l + 1, \dots, m, \quad (1.9b)$$

donde $\alpha_1, \dots, \alpha_r$ son los índices activos menores o iguales a l . Aquí usamos

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i$$

para denotar el producto escalar en \mathbb{R}^n .

A continuación introducimos dos conceptos importantes en programación no lineal.

Definición 1.2.3.

- a) Un vector $h \in \mathbb{R}^n$ se dice que es tangente al conjunto \mathcal{X} en el punto x^* si existe una curva suave $x(t)$ ($0 \leq t \leq \delta$) contenida en \mathcal{X} para algún $\delta > 0$ tal que $x(0) = x^*$ y $x'(0) = h$.
- b) Decimos que el problema de programación no lineal (1.3) satisface la condición de regularidad de Kuhn-Tucker en un punto $x^* \in \mathcal{X}$ si todo vector $h \in \mathbb{R}^n$ que satisface las restricciones tangenciales (1.9) en el punto x^* es un vector tangente a \mathcal{X} en x^* .

La definición 1.2.3(a) es representada en la figura 1.5.

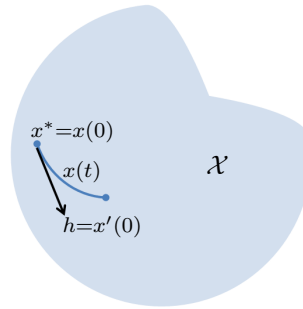


Figura 1.5: Vector tangente h en el punto x^* .

El siguiente resultado es uno de los más importantes en programación no lineal.

Teorema 1.2.5. *Supongamos que el problema de programación no lineal (1.3) satisface la condición de regularidad de Kuhn-Tucker en $x^* \in \mathcal{X}$. Entonces existen números reales $\lambda_1, \dots, \lambda_m$ tales que*

$$\lambda_\alpha \geq 0 \quad (\alpha = 1, \dots, l), \quad \text{con } \lambda_\alpha = 0 \text{ si } \alpha \text{ es un índice inactivo.} \quad (1.10a)$$

$$\nabla F(x^*) = 0, \quad \text{donde } F = f + \sum_{j=1}^m \lambda_j g_j. \quad (1.10b)$$

La función F es conocida como el *Lagrangiano* y los números $\lambda_1, \dots, \lambda_m$ se conocen como los *multiplicadores de Lagrange*. Las condiciones (1.10) son conocidas como las *condiciones de optimalidad de Karush-Kuhn-Tucker* (condiciones KKT). Una prueba del teorema es dada en *Hestenes* [9], teorema 7.1 (con una hipótesis de quasiregularidad sobre el punto x^* más general que la hipótesis de regularidad de Kuhn-Tucker). Las condiciones KKT del teorema 1.2.5 para un problema de programación no lineal (1.3) pueden ser escritas de manera equivalente como:

$$\nabla f(x) + \sum_{j=1}^m \lambda_j \nabla g_j(x) = 0 \quad (1.11a)$$

$$g_\alpha(x) \leq 0, \quad \alpha = 1, \dots, l \quad (1.11b)$$

$$g_\beta(x) = 0, \quad \beta = l + 1, \dots, m \quad (1.11c)$$

$$\lambda_\alpha g_\alpha(x) = 0 \quad (1.11d)$$

$$\lambda_\alpha \geq 0, \quad \alpha = 1, \dots, l \quad (1.11e)$$

Las ecuaciones (1.11b) y (1.11c) son las restricciones (1.8a) y (1.8b) respectivamente. Las ecuaciones (1.11a), (1.11d) y (1.11e) son equivalentes a las ecuaciones (1.10) (ver *Martínez et al.* [13], sección 2.4).

La importancia del teorema 1.2.5 radica en que nos permite obtener condiciones suficientes de primer y segundo orden para la existencia de puntos de mínimo restringidos.

Teorema 1.2.6. *Sea $x^* \in \mathcal{X}$ un punto en donde se satisfacen las condiciones KKT (1.11) y supongamos que no existe un vector $h \neq 0$ tal que satisfaga las restricciones tangenciales (1.9) en x^* y que además cumpla*

$$\langle \nabla f(x^*), h \rangle = 0$$

entonces x^ es un punto de mínimo local de f en \mathcal{X} .*

Teorema 1.2.7. *Sea $x^* \in \mathcal{X}$ y supongamos que existe un vector $h \neq 0$ el cual satisface las restricciones tangenciales (1.9) en x^* y que además cumpla*

$$\langle \nabla f(x^*), h \rangle = 0$$

Supongamos además que para cada vector $h \neq 0$ de este tipo existe una función Lagrangiana F definida por (1.10b) tal que

$$h^T \nabla^2 F(x^*) h > 0$$

Entonces existe un entorno \mathcal{N} de x^ y una constante $m > 0$ tal que la desigualdad*

$$f(x) \geq f(x^*) + m \|x - x^*\|^2$$

se cumple para todo $x \in \mathcal{N} \cap \mathcal{X}$, es decir, x^ es un punto de mínimo local de f en \mathcal{X} .*

Las demostraciones de los teoremas 1.2.6 y 1.2.7 son dadas en *Hestenes* [9] (teoremas 7.2 y 7.3 respectivamente). Una revisión más detallada sobre programación no lineal puede ser hallada en la literatura mencionada.

Capítulo 2

Optimización multiobjetivo

En la sección anterior consideramos el problema de minimizar (o maximizar) una función objetivo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ a valores escalares. Sin embargo, en una gran cantidad de aplicaciones es necesario resolver problemas de optimización que por su naturaleza involucran dos o más funciones objetivos, las cuales por lo general están en conflicto, de modo que es imposible minimizar todas simultáneamente. Los siguientes ejemplos ilustrarán más claramente lo mencionado.

Ejemplo 1.1 (Ehrgott [7], ejemplo 1.1). Supongamos que queremos comprar un automóvil y tenemos como opciones un VW Golf, un Opel Astra, un Ford Focus y un Toyota Corolla. Nuestra decisión de compra se basa en considerar el precio, el consumo de combustible y la potencia. Estamos interesados en un vehículo con el menor precio y consumo y con máxima potencia. A continuación se muestra una tabla con los valores objetivos (los cuales son inventados) de cada vehículo.

Vehículo	Precio (1.000 €) f_1	Consumo ($\frac{l}{100km}$) f_2	Potencia (kW) f_3
VW Golf	16.2	7.2	66.0
Opel Astra	14.9	7.0	62.0
Ford Focus	14.0	7.5	55.0
Toyota Corolla	15.2	8.2	71.0

Este es un problema de optimización multiobjetivo que involucra tres funciones objetivos: f_1 (precio), f_2 (consumo) y f_3 (potencia). Queremos seleccionar un automóvil (solución óptima) que minimice f_1 , f_2 y $-f_3$ simultáneamente (maximizar f_3 es equivalente a minimizar $-f_3$).

De la tabla podemos ver que tal solución óptima no existe. En efecto, el Ford Focus es la opción mas económica, pero el Opel Astra tiene el menor consumo y el Toyota Corolla la mayor potencia, por lo tanto ninguno de estos vehículos minimiza f_1 , f_2 y $-f_3$ simultáneamente. De esta forma, sin ninguna otra información adicional es imposible seleccionar una única solución. Sin embargo, observe que si consideramos en el problema una única función objetivo f_i , $i = 1, 2, 3$ (problema de optimización simple) entonces una única elección óptima es obtenida directamente. \square

Ejemplo 1.2. Para una mejor ilustración, consideremos ahora el problema de optimización multiobjetivo anterior que involucra únicamente las funciones objetivos f_1 y f_2 (precio y consumo). La figura 2.1 ilustra f_2 en función de f_1 (espacio objetivo). En éste podemos ver que tanto el VW Golf como el Toyota Corolla pueden ser descartados pues, por ejemplo, el Opel Astra es más económico y de menor consumo que ambos (y por lo tanto es preferido a éstos).

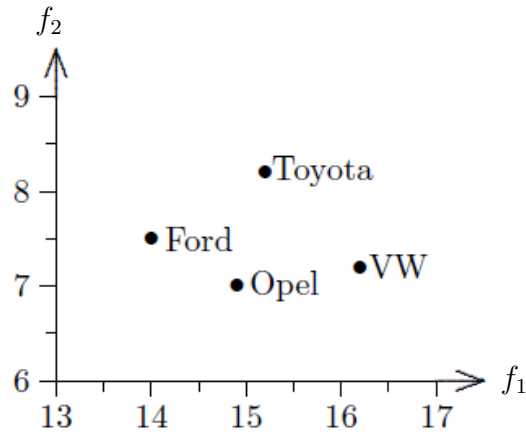


Figura 2.1: Espacio objetivo de ejemplo 1.2.

Sin embargo, no podemos descartar el Ford Focus, pues si bien el Opel tiene menor consumo, el Ford es más económico. Nuevamente, no podemos hacer una única elección que minimice los objetivos f_1 y f_2 simultáneamente y por lo tanto decimos que ambos, Ford y Opel, son soluciones igualmente válidas de nuestro problema de optimización multiobjetivo y forman, de acuerdo a lo que definiremos más adelante, el *conjunto de soluciones eficientes*. \square

Ejemplo 2 (Collette et al. [3], sección 1.6.3). Supongamos que queremos calcular las dimensiones de una viga de longitud L y sección transversal cuadrada de lado a , la cual soporta una carga puntual P en su centro (figura 2.2). Nuestro objetivo es obtener una viga con volumen mínimo, con el fin de ahorrar material, y que ofrezca la menor deformación posible ante la carga aplicada, debido a que una deformación más allá de lo deseable puede producir el colapso de elementos de terminación tales como cielorrasos o ventanales. Este es un problema típico en ingeniería.

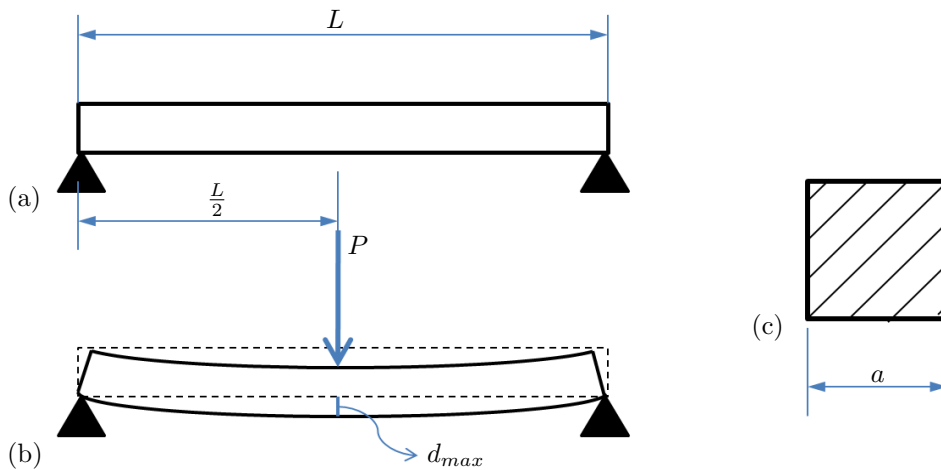


Figura 2.2: (a) Viga sin carga. (b) Viga con carga. (c) Sección transversal.

Supondremos en este ejemplo que se trata de una viga formada por un material homogéneo (por ejemplo hierro) tal que se cumplen la ley de Hooke y las hipótesis de Navier-Bernoulli. Bajo estas suposiciones, la deformación máxima que sufre la viga por el efecto de una carga

puntual en su centro es

$$d_{max} = \frac{PL^3}{48EI}$$

donde I es el momento de inercia correspondiente a la geometría de la sección transversal (en este caso $I = \frac{a^4}{12}$) y E es el módulo de elasticidad del material. Para los fines del cálculo tomaremos los valores $L = 100$ cm, $P = 100$ kg y $E = 2 \cdot 10^5$ kg/cm². También supondremos que, por motivos de diseño, el lado a de la sección transversal no debe ser mayor a 10 cm. Puesto que en este caso minimizar el volumen de la viga es equivalente a minimizar el área de su sección transversal, entonces el problema puede ser planteado como:

Minimizar

$$S(a) = a^2, \quad d(a) = \frac{10^2 \cdot (10^2)^3}{48 \cdot 2 \cdot 10^5 \cdot \frac{a^4}{12}}$$

sujeto a

$$0 \leq a \leq 10$$

donde S y d denotan el área y la deformación respectivamente.

En la figura 2.3 se muestran los valores (S, d) , los cuales fueron obtenidos seleccionando valores aleatorios para a (con distribución uniforme) en el intervalo $[0, 10]$ y luego calculando los valores correspondientes $S(a)$ y $d(a)$. Como puede observarse en la figura, las funciones S y d están en conflicto; en efecto, para dos pares cualesquiera (S_1, d_1) y (S_2, d_2) tenemos que $d_2 < d_1$ cada vez que $S_1 < S_2$ y viceversa. Esto es completamente intuitivo, pues es de esperarse que una viga con un área transversal menor sufrirá una mayor deformación bajo los efectos de la misma carga que una viga con un área transversal mayor. Luego, sin ninguna otra información adicional es imposible seleccionar una única solución óptima al problema y por lo tanto es necesario realizar un balance agregado o compromiso (*trade-off*) entre todos los valores factibles de a con el fin de elegir una única solución. El conjunto $\{(S(a), d(a)) \mid 0 \leq a \leq 10\}$ es lo que definiremos más adelante como el *conjunto de puntos no dominados*.

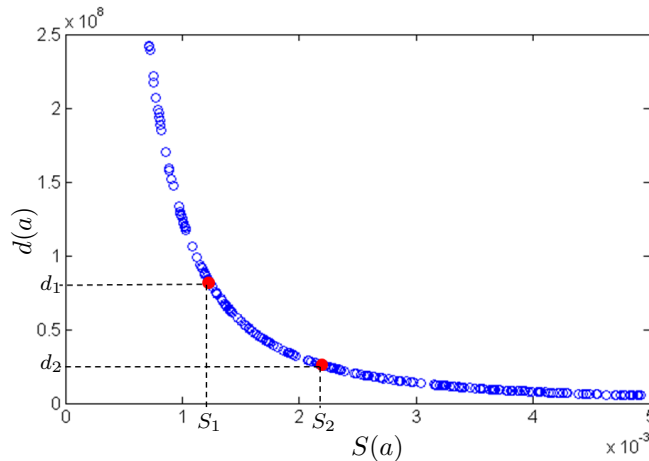


Figura 2.3: Espacio objetivo de ejemplo 2.

Obsérvese que para este ejemplo el planteo de un problema de optimización simple considerando sólo una función objetivo, S ó d , nos conduce a una solución sin sentido, pues la solución al problema de minimizar solamente el área S nos lleva a obtener un “área nula”, mientras que

el problema de minimizar solamente la deformación d nos conduce a un “sobredimensionamiento” de la viga. \square

Un problema de optimización multiobjetivo puede plantearse de manera general como

$$\min_{x \in \mathcal{X}} F(x), \quad (2.1)$$

donde $F = (f_1, \dots, f_p) : \mathbb{R}^n \rightarrow \mathbb{R}^p$ es una función a valores vectoriales y cada función f_j , $j = 1, \dots, p$, es una función objetivo. En estos problemas \mathbb{R}^n es llamado el *espacio variable* o *espacio de decisión* y \mathbb{R}^p el *espacio objetivo* o *espacio criterio*. El conjunto $\mathcal{X} \subset \mathbb{R}^n$ es llamado el *conjunto factible* en el espacio variable (en los ejemplos 1.1 y 1.2 tenemos que $\mathcal{X} = \{\text{VW Golf, Opel Astra, Ford Focus, Toyota Corolla}\}$ y en el ejemplo 2 resulta $\mathcal{X} = [0, 10]$) y el conjunto $\mathcal{Y} = F(\mathcal{X}) \subset \mathbb{R}^p$ es llamado el *conjunto factible* en el espacio objetivo.

Como en el caso de optimización simple, un problema de optimización multiobjetivo puede clasificarse como *irrestringido* si \mathcal{X} es un abierto en \mathbb{R}^n o *con restricciones* si $\mathcal{X} \subsetneq \mathbb{R}^n$ no es un abierto, donde en este último caso el conjunto \mathcal{X} puede estar dado para fines prácticos por restricciones de igualdad y desigualdad del tipo

$$\begin{aligned} g_\alpha(x) &\leq 0, \quad \alpha = 1, \dots, l, \\ g_\beta(x) &= 0, \quad \beta = l + 1, \dots, m, \end{aligned}$$

para ciertas funciones $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$, o por restricciones de caja

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid l_i \leq x_i \leq u_i, \quad l_i, u_i \in \mathbb{R}, \quad i = 1, \dots, n\}.$$

Si las funciones objetivos f_j y las funciones de restricción g_i son lineales, entonces estamos ante un problema de optimización multiobjetivo lineal (ver ejemplo 6 de sección 2.2).

Como podemos ver de los ejemplos dados anteriormente, un problema de optimización multiobjetivo también puede clasificarse como continuo si los objetivos f_j son funciones continuas (ejemplo 2 y ejemplos 3, 4 y 6 de sección 2.2) o como discreto si los objetivos f_j son funciones discretas (ejemplos 1.1, 1.2 y ejemplo 5 de sección 2.2).

2.1. Conos y órdenes relacionados

En el caso de optimización simple, el significado de minimizar una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sobre un conjunto \mathcal{X} es claro: el punto $x^* \in \mathcal{X}$ es una solución a tal problema si $f(x^*) \leq f(x)$ para todo $x \in \mathcal{X}$. Como puede verse en este caso, el significado de *mínimo* está directamente relacionado con el orden canónico \leq en \mathbb{R} . Esto no ocurre en un problema de optimización multiobjetivo. Puesto que no existe un orden canónico en \mathbb{R}^p para $p \geq 2$, entonces la definición de *mínimo* en (2.1) estará relacionada con un determinado orden en \mathbb{R}^p previamente establecido. No daremos aquí una descripción detallada de órdenes, limitándonos sólo a los conceptos básicos necesarios. Referirse a *Sawaragi et al.* [19] (sección 2.3.1), *Nakayama et al.* [15] (secciones 1.1 y 1.2) y *Ehrgott* [7] (sección 1.4) para una descripción más detallada.

En primer lugar definiremos el concepto de *cono* en \mathbb{R}^p y posteriormente definiremos los conceptos de *pre-orden* y de *orden* en \mathbb{R}^p relacionado a un cono.

Definición 2.1.1. *Un conjunto $\mathcal{C} \subset \mathbb{R}^p$ es un cono si para todo $d \in \mathcal{C}$ y todo número real $\alpha > 0$ se cumple que $\alpha d \in \mathcal{C}$. Un cono \mathcal{C} en \mathbb{R}^p se dice que es:*

- *Punteado si para todo $d \in \mathcal{C}$, $d \neq 0$, se cumple que $-d \notin \mathcal{C}$, es decir, $\mathcal{C} \cap \{-\mathcal{C}\} \subset \{0\}$.*

- *Convexo* si éste es convexo como un subconjunto de \mathbb{R}^p .
- *Cerrado* si éste es cerrado como un subconjunto de \mathbb{R}^p .
- *Propio* si $\mathcal{C} \neq \emptyset$ y $\mathcal{C} \neq \mathbb{R}^p$.

De la definición de cono, podemos ver que \mathcal{C} es un *cono convexo* si $d_1 + d_2 \in \mathcal{C}$ para todo $d_1, d_2 \in \mathcal{C}$ (condición (\star)). En efecto, dados $d_1, d_2 \in \mathcal{C}$ tenemos que $(1 - \alpha)d_1 \in \mathcal{C}$ y $\alpha d_2 \in \mathcal{C}$ para todo $0 < \alpha < 1$ por ser \mathcal{C} un cono. Por lo tanto, si la condición (\star) se cumple, entonces $(1 - \alpha)d_1 + \alpha d_2 \in \mathcal{C} \quad \forall 0 \leq \alpha \leq 1$, es decir, \mathcal{C} es convexo.

Definición 2.1.2. Una relación binaria “ \prec ” en \mathbb{R}^p se dice que es un pre-orden si ésta es reflexiva y transitiva, es decir

- 1) $x \prec x \quad \forall x \in \mathbb{R}^p$.
- 2) Si $x \prec y$ e $y \prec z$, entonces $x \prec z \quad \forall x, y, z \in \mathbb{R}^p$.

Definición 2.1.3. Dados $x, y \in \mathbb{R}^p$, decimos que $x \prec y$ (x es preferido a y) respecto de un cono punteado y convexo $\mathcal{C} \subset \mathbb{R}^p$ si $y - x \in \mathcal{C}$, es decir, si existe $d \in \mathcal{C}$ tal que $y = x + d$.

Puede probarse (Ehrgott [7], teorema 1.17) que si \mathcal{C} es un cono punteado y convexo tal que $0 \in \mathcal{C}$, entonces la relación binaria “ \prec ” de la definición 2.1.3 resulta ser un pre-orden *antisimétrico*, es decir, también satisface

- 3) Si $x \prec y$ e $y \prec x$, entonces $x = y \quad \forall x, y \in \mathbb{R}^p$.

En este caso se dice que “ \prec ” es un *orden parcial* en \mathbb{R}^p relacionado al cono \mathcal{C} . Aquí estamos interesados principalmente en un tipo de orden en \mathbb{R}^p , el cual definiremos a continuación.

Sea $\mathcal{C} = \mathbb{R}_+^p = \{(x_1, \dots, x_p) \in \mathbb{R}^p \mid x_i \geq 0 \quad \forall i = 1, \dots, p\}$ (el “octante” no negativo de \mathbb{R}^p), entonces éste es claramente un cono punteado y convexo que contiene al 0. Consideremos el orden parcial “ \prec ” relacionado a $\mathcal{C} = \mathbb{R}_+^p$. Definimos una nueva relación binaria “ \leq ” en \mathbb{R}^p ($p \geq 2$) por:

$$x \leq y \iff x \prec y, \quad x \neq y. \quad (2.2a)$$

la cual podemos escribir de las maneras alternativas:

$$x \leq y \iff y - x \in \mathbb{R}_+^p, \quad x \neq y. \quad (2.2b)$$

$$x \leq y \iff y - x \in \mathbb{R}_+^p \setminus \{0\}. \quad (2.2c)$$

$$x \leq y \iff x_i \leq y_i \quad \forall i = 1, \dots, p, \quad \text{con } x_j < y_j \text{ para algún } j = 1, \dots, p. \quad (2.2d)$$

Es claro que la relación binaria definida en (2.2) resulta ser *asimétrica* (si $x \leq y$, entonces $y \not\leq x \quad \forall x, y \in \mathbb{R}^p$) y transitiva, es decir, un *orden parcial estricto*. Este orden “ \leq ” respecto del cono $\mathcal{C} = \mathbb{R}_+^p \setminus \{0\}$ se conoce como el *orden de Pareto* y el concepto de optimización multiobjetivo respecto de este orden se conoce como *optimalidad de Pareto*. La figura 2.4 ilustra el orden de Pareto en \mathbb{R}^2 .

Sawaragi et al. [19] desarrolla la teoría de optimización multiobjetivo para un cono arbitrario $\mathcal{C} \subset \mathbb{R}^p$ y Eichfelder [8] considera también órdenes de distinta naturaleza. Jahn [10] desarrolla la teoría de optimización multiobjetivo para espacios vectoriales generales. Ehrgott [7] (capítulos 2 a 4) y Miettinen [14] consideran optimalidad de Pareto ($\mathcal{C} = \mathbb{R}_+^p \setminus \{0\}$) que es el concepto de optimalidad con mayor cantidad de aplicaciones.

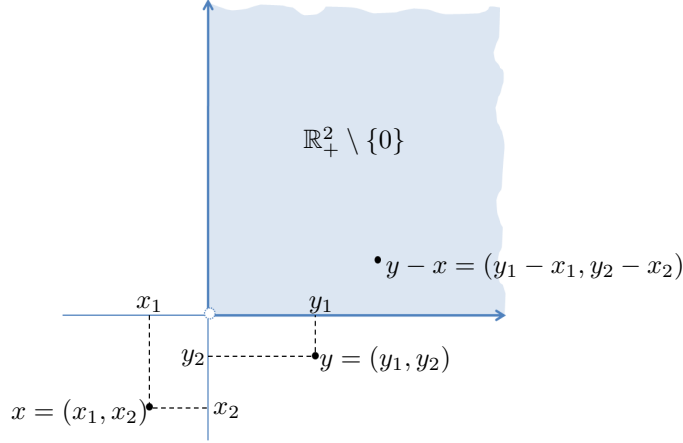


Figura 2.4: Ilustración del orden de Pareto en \mathbb{R}^2 ($x \leq y$).

2.2. Soluciones eficientes y puntos no dominados

En este trabajo estamos interesados en problemas de optimización multiobjetivo donde el significado de “mínimo” está relacionado a la optimalidad de Pareto. Tales problemas pueden plantearse como

$$\min_{x \in \mathcal{X}} F(x) = \min_{x \in \mathcal{X}} (f_1(x), \dots, f_p(x)), \quad p \geq 2 \quad (2.3)$$

donde $\mathcal{X} \subseteq \mathbb{R}^n$ y consideramos en el espacio objetivo \mathbb{R}^p el orden de Pareto definido en la sección anterior. Como ya mencionamos, si las funciones objetivos f_j , $j = 1, \dots, p$, están en conflicto, entonces es imposible, sin información adicional, seleccionar una única solución óptima. En estos casos estamos dispuestos a aceptar un conjunto de soluciones, las cuales son consideradas igualmente válidas para el problema en cuestión. Llamaremos a éste el *conjunto de soluciones eficientes* y a su imagen por F , el *conjunto de puntos no dominados*.

Teniendo en cuenta el orden de Pareto (2.2), podemos dar la siguiente definición.

Definición 2.2.1. Decimos que $x^* \in \mathcal{X}$ es una solución eficiente del problema de optimización multiobjetivo (2.3) si no existe un $x \in \mathcal{X}$ tal que $F(x) \leq F(x^*)$, donde “ \leq ” denota el orden de Pareto en \mathbb{R}^p definido en (2.2). El conjunto de soluciones eficientes es denotado por \mathcal{X}_E . Si $y^* = F(x^*)$, con $x^* \in \mathcal{X}_E$, entonces y^* es llamado un punto no dominado. El conjunto $\mathcal{Y}_N = F(\mathcal{X}_E)$ es llamado el conjunto de puntos no dominados.

Obsérvese que $\mathcal{X}_E \subset \mathcal{X}$ y que $\mathcal{Y}_N \subset \mathcal{Y} = F(\mathcal{X})$. El conjunto de soluciones eficientes también es llamado la *variedad de Pareto* y el conjunto de puntos no dominados, el *frente de Pareto*. A lo largo de esta sección usaremos las palabras *eficiencia* y *no dominancia* por ser las más utilizadas en la literatura. El término “variedad de Pareto” será introducido en la última sección, donde el conjunto de soluciones eficientes será obtenido como una variedad paramétrica de dimensión $p - 1$ en el espacio variable (cuando $n \geq p$).

Las siguientes definiciones equivalentes de eficiencia resultarán de utilidad:

- 1) x^* es una solución eficiente si no existe $x \in \mathcal{X}$ tal que $F(x^*) - F(x) \in \mathbb{R}_+^p \setminus \{0\}$.
- 2) x^* es una solución eficiente si no existe $x \in \mathcal{X}$ tal que $f_i(x) \leq f_i(x^*) \quad \forall i = 1, \dots, p$ y $f_j(x) < f_j(x^*)$ para algún $j = 1, \dots, p$.

3) x^* es una solución eficiente si $\mathcal{Y} \cap (F(x^*) - \mathbb{R}_+^p) = \{F(x^*)\}$.

Las definiciones 1 y 2 son consecuencias inmediatas de (2.2c) y (2.2d) respectivamente. La definición 3 se desprende de la definición 1. En efecto, si existe $x \in \mathcal{X}$ tal que $F(x) \in \mathcal{Y} \cap (F(x^*) - \mathbb{R}_+^p)$, con $F(x) \neq F(x^*)$, entonces en particular existe un $d \in \mathbb{R}_+^p$, $d \neq 0$, tal que $F(x) = F(x^*) - d$, es decir, $F(x^*) - F(x) = d \in \mathbb{R}_+^p \setminus \{0\}$, por lo que x^* no sería eficiente de acuerdo a la definición 1.

De la definición de solución eficiente vemos que

$$\mathcal{Y}_N = \{y^* \in \mathcal{Y} \mid \text{no existe } y \in \mathcal{Y} \text{ tal que } y \leq y^*\}. \quad (2.4)$$

El siguiente ejemplo ilustrará mejor estas ideas.

Ejemplo 3 (Ehrgott [7], ejemplo 1.3). Consideremos el problema de optimización bi-objetivo

$$\min_{x \in \mathcal{X}} (f_1(x), f_2(x))$$

donde $f_1(x) = \sqrt{x+1}$, $f_2(x) = (x-2)^2 + 1$ y $\mathcal{X} = \{x \in \mathbb{R} \mid x > 0\}$. En este ejemplo tenemos a \mathbb{R} como el espacio variable y a \mathbb{R}^2 como el espacio objetivo.

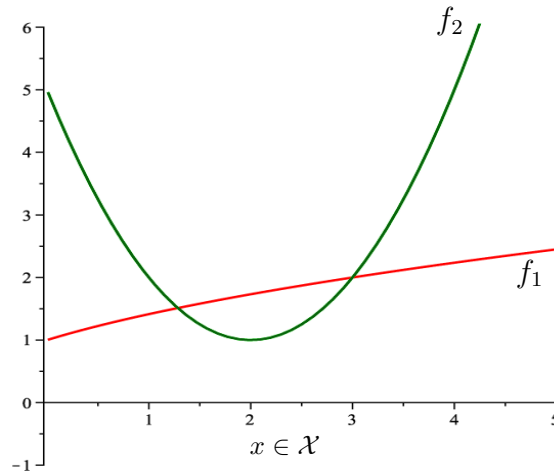


Figura 2.5: Gráficas de f_1 y f_2 de ejemplo 3.

En la figura 2.5 se muestran las gráficas de las funciones objetivos f_1 y f_2 para $x \in \mathcal{X}$. La función f_2 tiene un mínimo en $x = 2$ (como puede deducirse del hecho de que f_2 es convexa y $f_2'(2) = 0$, ver teorema 1.2.4). La función objetivo f_1 es estrictamente creciente en \mathcal{X} , mientras que f_2 es estrictamente decreciente en el intervalo $(0, 2]$ y estrictamente creciente en el intervalo $(2, \infty)$. De estas consideraciones podemos ver que cualquier $x \in (2, \infty)$ no puede ser una solución a nuestro problema, pues para un $y \in (2, \infty)$ con $y < x$ tenemos que $f_1(y) < f_1(x)$ y $f_2(y) < f_2(x)$, siendo por lo tanto y preferido a x . Considerando el orden de Pareto en \mathbb{R}^2 esto significa que $(f_1(y), f_2(y)) \leq (f_1(x), f_2(x))$.

No obstante, si $x \in (0, 2]$, entonces para cualquier $y \in \mathcal{X}$, $y \neq x$, se cumple una y sólo una de las siguientes opciones (para $x = 2$ la opción (b) no es considerada):

- a) $f_1(y) < f_1(x)$ y $f_2(y) > f_2(x)$ si $y \in (0, x)$,
- b) $f_1(y) > f_1(x)$ y $f_2(y) \leq f_2(x)$ si $y \in (x, 4 - x]$,

c) $f_1(y) > f_1(x)$ y $f_2(y) > f_2(x)$ si $y \in (4 - x, \infty)$.

Luego, ningún x en éste intervalo puede ser ignorado pues para cualquier y distinto vemos que x es preferido a y (opción (c)) ó y disminuye el valor de f_1 , pero aumenta el valor de f_2 (opción (a)) ó y disminuye (o iguala) el valor de f_2 , pero aumenta el valor de f_1 (opción (b)). En términos del orden de Pareto esto significa que si $x \in (0, 2]$, entonces no existe $y \in \mathcal{X}$ tal que $(f_1(y), f_2(y)) \leq (f_1(x), f_2(x))$, es decir, x es una solución eficiente. Concluimos que $\mathcal{X}_E = (0, 2]$.

Nótese que la obtención del conjunto eficiente \mathcal{X}_E requirió de una clasificación exhaustiva de varios casos que pueden aumentar su dificultad para problemas más complejos. Una manera mucho más simple de obtener tales soluciones es analizando el conjunto factible $\mathcal{Y} = F(\mathcal{X})$ en el espacio objetivo. Para tal fin definimos $y_1 = f_1(x)$, $y_2 = f_2(x)$ y expresamos y_2 en función de y_1 (teniendo en cuenta que $x > 0$):

$$y_1 = \sqrt{x+1} \Rightarrow x = y_1^2 - 1 \Rightarrow y_2 = ((y_1^2 - 1) - 2)^2 + 1 = y_1^4 - 6y_1^2 + 10$$

Puesto que $x > 0$, entonces $y_1 > 1$. En la figura 2.6 se muestra el conjunto factible $\mathcal{Y} = F(\mathcal{X})$ en el espacio objetivo \mathbb{R}^2 . La función $y_2 = y_1^4 - 6y_1^2 + 10$ tiene un punto de mínimo en $y_1 = \sqrt{3}$, como puede deducirse del hecho de que y_2 es convexa en el intervalo $(1, \infty)$ e $y_2'(\sqrt{3}) = 0$.

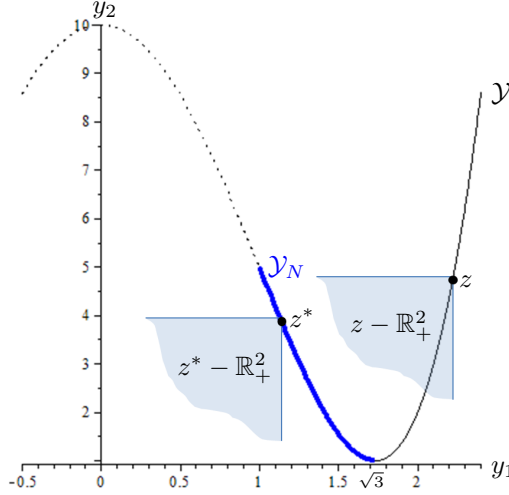


Figura 2.6: Conjunto factible \mathcal{Y} y conjunto no dominado \mathcal{Y}_N de ejemplo 3.

De la figura podemos ver que si $z = (y_1, y_2) \in \mathcal{Y}$ es tal que $y_1 \in (\sqrt{3}, \infty)$, entonces $\mathcal{Y} \cap (z - \mathbb{R}_+^2) \neq \{z\}$, mientras que si $z^* = (y_1^*, y_2^*) \in \mathcal{Y}$ es tal que $y_1^* \in (1, \sqrt{3}]$, entonces $\mathcal{Y} \cap (z^* - \mathbb{R}_+^2) = \{z^*\}$. Por la definición 3 de solución eficiente tenemos que $z^* = (y_1^*, y_2^*) \in \mathcal{Y}$ es un punto no dominado si y sólo si $y_1^* \in (1, \sqrt{3}]$, es decir $\mathcal{Y}_N = \{(y_1^*, y_2^*) \in \mathcal{Y} \mid y_1^* \in (1, \sqrt{3}]\}$. Se comprueba fácilmente que $\mathcal{X}_E = (f_1, f_2)^{-1}(\mathcal{Y}_N) = \{x \in \mathcal{X} \mid (f_1(x), f_2(x)) \in \mathcal{Y}_N\} = (0, 2]$. \square

En el ejemplo 3 se puede ver la utilidad gráfica del cono \mathbb{R}_+^p para determinar el conjunto \mathcal{Y}_N . En efecto, los puntos no dominados son aquellos $y^* \in \mathcal{Y}$ tales que la intersección del cono trasladado $y^* - \mathbb{R}_+^p$ con el conjunto factible \mathcal{Y} contiene a y^* como único punto (figura 2.7), es decir,

$$\mathcal{Y}_N = \{y^* \in \mathcal{Y} \mid \mathcal{Y} \cap (y^* - \mathbb{R}_+^p) = \{y^*\}\}. \quad (2.5)$$

La igualdad (2.5) es una consecuencia inmediata de la definición 3 de solución eficiente.

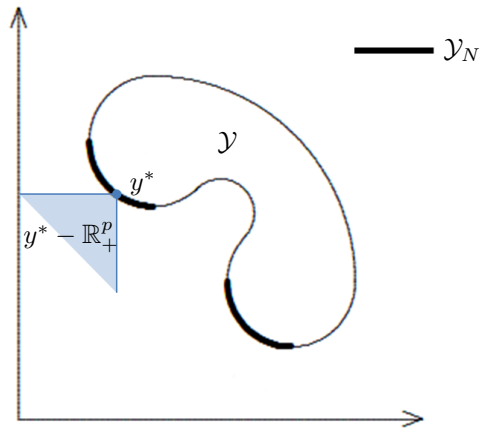


Figura 2.7: Conjunto no dominado \mathcal{Y}_N .

A continuación mostraremos los conjuntos no dominados de algunos problemas de optimización multiobjetivo. En la figura 2.8 se muestra el conjunto no dominado del ejemplo 1.2.

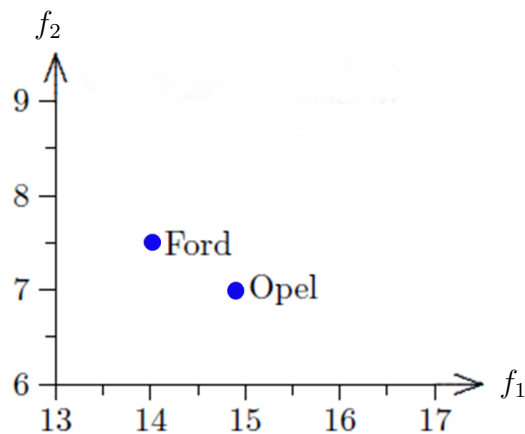


Figura 2.8: Conjunto no dominado de ejemplo 1.2.

En la figura 2.3 se muestra el conjunto no dominado del ejemplo 2, aunque el mismo puede ser obtenido directamente escribiendo la deformación d en función del área S :

$$S = a^2 \implies d = \frac{10^2 \cdot (10^2)^3}{48 \cdot 2 \cdot 10^5 \cdot \frac{S^2}{12}}$$

y graficando en el plano (S, d) . El conjunto no dominado obtenido de esta manera se muestra en la figura 2.9.

Ejemplo 4 (Collette et al. [3], sección 1.6.3.2). Consideremos nuevamente el ejemplo 2 de la viga, pero esta vez con una sección transversal hueca como se muestra en la figura 2.10(a). El objetivo es, como antes, minimizar el volumen de la viga y la deformación que sufre debido a una carga puntual centrada P . Tomando los mismos valores para la carga P , la longitud L y

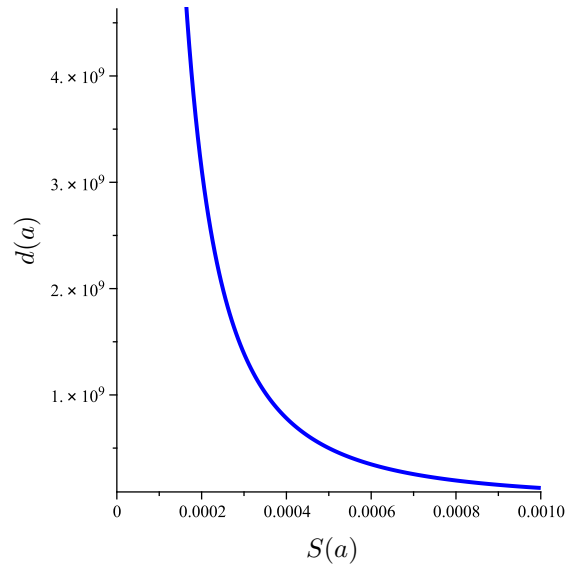


Figura 2.9: Conjunto no dominado de ejemplo 2.

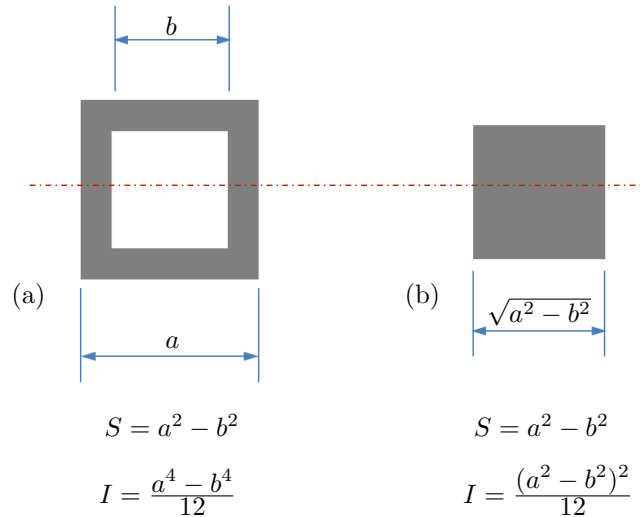


Figura 2.10: Secciones transversales con misma área pero distinta geometría.

el módulo de elasticidad E como en el ejemplo 2, podemos plantear el nuevo problema como sigue:

Minimizar

$$S(a, b) = a^2 - b^2, \quad d(a, b) = \frac{10^2 \cdot (10^2)^3}{48 \cdot 2 \cdot 10^5 \cdot \frac{(a^4 - b^4)}{12}}$$

sujeto a

$$0 \leq a \leq 10, \quad 0 \leq b \leq 6, \quad a - b \geq 1.$$

Las restricciones son impuestas para obtener valores de a y b no mayores a 10 cm y 6 cm respectivamente y un espesor de la pared de la sección transversal no menor a 0.5 cm.

En la figura 2.11 se muestra el conjunto no dominado \mathcal{Y}_N . Las curvas mostradas en rojo corresponden a distintos valores de b ; a medida que el valor de b aumenta, la curva obtenida se ubica por debajo de las otras curvas. Para valores a y b fijos, podemos deducir de la figura 2.11 que dos vigas con secciones transversales (a) y (b) de igual área (y por lo tanto de igual

volumen) pero con distinta geometría (figura 2.10), se deforman de manera diferente ante una misma carga P . En efecto, la viga con sección transversal (a) sufre una menor deformación ante la carga P que la viga con sección transversal (b). \square

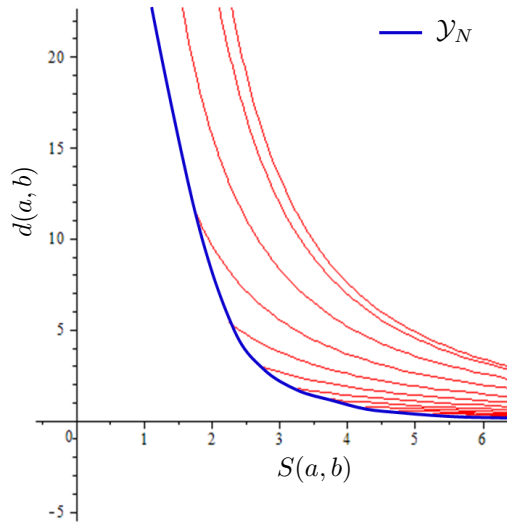


Figura 2.11: Conjunto no dominado de ejemplo 4. Cada curva en rojo corresponde a un valor fijo de b .

Ejemplo 5. Problema de la mochila (Correa et al. [4]). El problema clásico de la mochila (Knapsack problem) en su versión simple consiste en dar una mochila con cierta capacidad y una serie de productos o ítems con un peso y un beneficio asociado (ambos positivos) a cada uno de ellos. La tarea es entonces seleccionar el subconjunto de ítems que maximicen el beneficio total y que puedan ser colocados dentro de la mochila sin exceder su límite de capacidad. Este problema de optimización simple puede ser extendido al caso multiobjetivo al considerar dos o más mochilas. En este caso no existe una solución óptima única sino un conjunto de posibles combinaciones que permiten diferentes beneficios al introducir unos u otros elementos entre las diferentes mochilas respetando sus respectivas capacidades. Dado un conjunto de n ítems y un conjunto de p mochilas, sea $b_{i,j}$ el beneficio del j -ésimo ítem en la i -ésima mochila, $w_{i,j}$ el peso del j -ésimo ítem en la i -ésima mochila y c_i la capacidad de la i -ésima mochila. Entonces el problema puede ser planteado como:

Maximizar

$$f_i(x) = \sum_{j=1}^n b_{i,j} x_j, \quad i = 1, \dots, p,$$

sujeto a

$$x_j \in \{0, 1\}, \quad j = 1, \dots, n,$$

$$\sum_{j=1}^n w_{i,j} x_j \leq c_i, \quad i = 1, \dots, p,$$

donde $x_j = 1$ si el j -ésimo ítem es seleccionado y $x_j = 0$ en caso contrario.

En la figura 2.12 se muestra el conjunto no dominado para un problema con una cantidad de 100 ítems y 2 mochilas¹, el cual fue resuelto aplicando un algoritmo genético (ver [4]). \square

¹los datos pueden ser descargados de <http://www.tik.ee.ethz.ch/sop/download/supplementary/testProblemSuite>

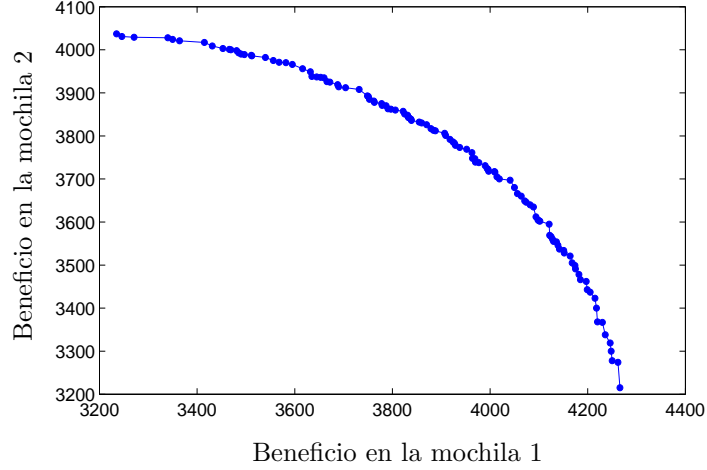


Figura 2.12: Conjunto no dominado de ejemplo 5.

Ejemplo 6. Consideremos el siguiente problema de optimización multiobjetivo lineal:
Minimizar

$$(z_1, z_2) = (2x_1 - x_2, -x_1 + 5x_2),$$

sujeto a

$$\begin{aligned} x_1 + x_2 &\leq 8, \\ -x_1 + x_2 &\leq 3, \\ 0 &\leq x_1 \leq 6, \\ 0 &\leq x_2 \leq 4. \end{aligned}$$

El conjunto factible \mathcal{X} en el espacio variable se muestra en la figura 2.13(a) y el conjunto factible \mathcal{Y} en el espacio objetivo junto con el conjunto no dominado \mathcal{Y}_N se muestra en la figura 2.13(b). El conjunto \mathcal{Y} es obtenido al tomar las imágenes de los vértices del conjunto \mathcal{X} , como se muestra en la tabla siguiente.

Vértices	x_1	x_2	z_1	z_2
A	0	0	0	0
B	0	3	-3	15
C	1	4	-2	19
D	4	4	4	16
E	6	2	10	4
F	6	0	12	-6

Otros ejemplos de optimización multiobjetivo, así como bibliografía, papers, tesis, software y funciones de testeo con sus respectivos conjuntos no dominados (frentes de Pareto) pueden ser consultados en la *EMOO Web Page* <http://delta.cs.cinvestav.mx/~ccoello/EMOO> \square

En lo que respecta al orden de Pareto, resulta de utilidad en optimización multiobjetivo definir un orden en \mathbb{R}^p más restrictivo, el cual lleva a un concepto más débil de solución eficiente. Consideremos el cono $\mathcal{C} = (\mathbb{R}_+^p)^\circ = \{(x_1, \dots, x_p) \in \mathbb{R}^p \mid x_i > 0 \ \forall i = 1, \dots, p\}$ y definamos el orden “<” en \mathbb{R}^p ($p \geq 2$) por:

$$x < y \iff y - x \in (\mathbb{R}_+^p)^\circ \quad (2.6a)$$

el cual podemos escribir de la manera alternativa:

$$x < y \iff x_i < y_i \ \forall i = 1, \dots, p. \quad (2.6b)$$

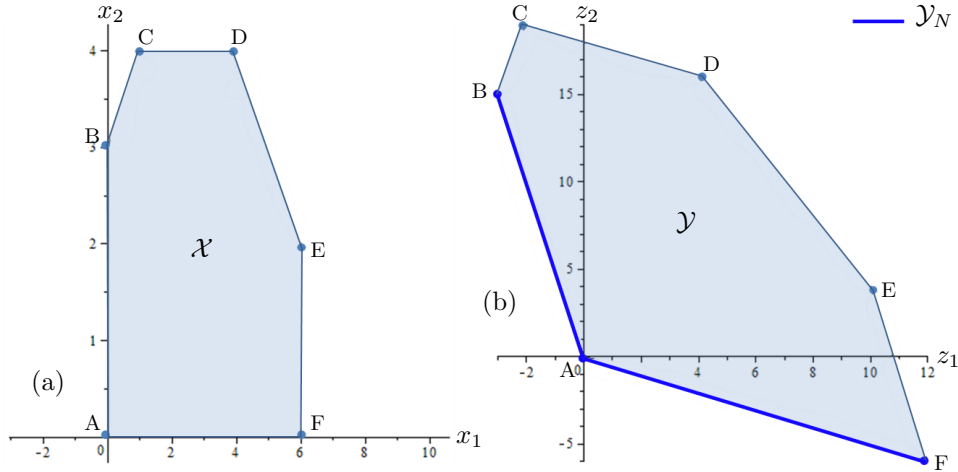


Figura 2.13: (a) Conjunto factible en el espacio variable. (b) Conjunto factible y conjunto no dominado en el espacio objetivo.

Teniendo en cuenta este orden damos la siguiente definición.

Definición 2.2.2. Decimos que $x^* \in \mathcal{X}$ es una solución débilmente eficiente del problema de optimización multiobjetivo (2.3) si no existe $x \in \mathcal{X}$ tal que $F(x) < F(x^*)$, donde “ $<$ ” denota el orden definido en (2.6). El conjunto de soluciones débilmente eficientes es denotado por \mathcal{X}_{wE} . Si $y^* = F(x^*)$, con $x^* \in \mathcal{X}_{wE}$, entonces y^* es llamado un punto débilmente no dominado. El conjunto $\mathcal{Y}_{wN} = F(\mathcal{X}_{wE})$ es llamado el conjunto de puntos débilmente no dominados.

Otras definiciones equivalentes de eficiencia débil son:

- 1) x^* es una solución débilmente eficiente si no existe $x \in \mathcal{X}$ tal que $F(x^*) - F(x) \in (\mathbb{R}_+^p)^\circ$.
- 2) x^* es una solución débilmente eficiente si no existe $x \in \mathcal{X}$ tal que $f_i(x) < f_i(x^*) \forall i = 1, \dots, p$.
- 3) x^* es una solución débilmente eficiente si $\mathcal{Y} \cap (F(x^*) - (\mathbb{R}_+^p)^\circ) = \emptyset$.

Las definiciones 1 y 2 son consecuencias inmediatas de (2.6a) y (2.6b) respectivamente. La definición 3 sigue de la definición 1. En efecto, si existe $x \in \mathcal{X}$ tal que $F(x) \in \mathcal{Y} \cap (F(x^*) - (\mathbb{R}_+^p)^\circ)$, entonces, en particular, existe $d \in (\mathbb{R}_+^p)^\circ$ tal que $F(x) = F(x^*) - d$, es decir, $F(x^*) - F(x) = d \in (\mathbb{R}_+^p)^\circ$, por lo que x^* no es débilmente eficiente de acuerdo a la definición 1.

De las definiciones anteriores se desprende que

$$\mathcal{Y}_{wN} = \{y^* \in \mathcal{Y} \mid \text{no existe } y \in \mathcal{Y} \text{ tal que } y < y^*\} = \{y^* \in \mathcal{Y} \mid \mathcal{Y} \cap (y^* - (\mathbb{R}_+^p)^\circ) = \emptyset\}. \quad (2.7)$$

Dado que $x < y$ implica que $x \leq y$ en \mathbb{R}^p , entonces $\mathcal{X}_E \subset \mathcal{X}_{wE}$ y por lo tanto $\mathcal{Y}_N \subset \mathcal{Y}_{wN}$.

Ejemplo 7. Consideremos el conjunto $\mathcal{Y} = [0, 1] \times [0, 1]$ (figura 2.14). De (2.7) vemos que $\mathcal{Y}_{wN} = \{(y_1, y_2) \in \mathcal{Y} \mid y_1 = 0 \text{ ó } y_2 = 0\}$. En este caso $\mathcal{Y}_N = \{(0, 0)\}$. \square

Los conjuntos (débilmente) no dominados poseen ciertas propiedades, algunas de las cuales son presentadas a continuación.

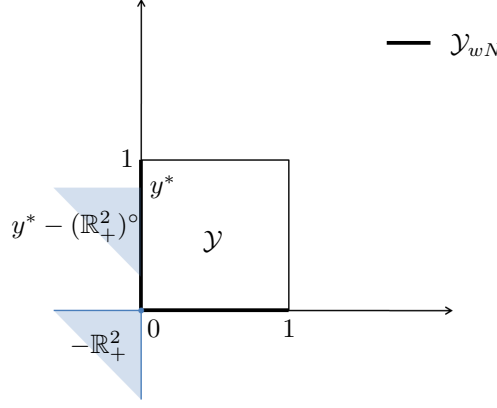


Figura 2.14: Conjunto débilmente no dominado \mathcal{Y}_{wN} .

Proposición 2.2.1. $\mathcal{Y}_{wN} \subset (\mathcal{Y} + (\mathbb{R}_+^p)^\circ)_{wN}$.

Demostración. Supongamos que $y \in \mathcal{Y}_{wN}$, pero $y \notin (\mathcal{Y} + (\mathbb{R}_+^p)^\circ)_{wN}$. Entonces existen $y' \in (\mathcal{Y} + (\mathbb{R}_+^p)^\circ)$ y $d' \in (\mathbb{R}_+^p)^\circ$ tales que $y = y' + d'$. Por otra parte, existen $y'' \in \mathcal{Y}$ y $d'' \in (\mathbb{R}_+^p)^\circ$ tales que $y' = y'' + d''$. Luego $y = y'' + (d' + d'')$, con $d' + d'' \in (\mathbb{R}_+^p)^\circ$ (por ser $(\mathbb{R}_+^p)^\circ$ convexo, ver sección 2.1). Esto implica que $y \notin \mathcal{Y}_{wN}$, lo cual es una contradicción. Por lo tanto $\mathcal{Y}_{wN} \subset (\mathcal{Y} + (\mathbb{R}_+^p)^\circ)_{wN}$. ■

La siguiente proposición nos dice que el conjunto \mathcal{Y}_N se ubica “abajo y a la izquierda” del conjunto \mathcal{Y} .

Proposición 2.2.2. $\mathcal{Y}_N = (\mathcal{Y} + \mathbb{R}_+^p)_N$.

Demostración. El resultado es trivial si $\mathcal{Y} = \emptyset$, por lo que asumiremos lo contrario. Supongamos que $y \in (\mathcal{Y} + \mathbb{R}_+^p)_N$, pero $y \notin \mathcal{Y}_N$. Si $y \notin \mathcal{Y}$, entonces existen $y' \in \mathcal{Y}$ y $d \in \mathbb{R}_+^p$, $d \neq 0$, tales que $y = y' + d$. Puesto que $\mathcal{Y} \subset (\mathcal{Y} + \mathbb{R}_+^p)$ (pues $0 \in \mathbb{R}_+^p$), esto dice que $y \notin (\mathcal{Y} + \mathbb{R}_+^p)_N$, lo cual es una contradicción. Si $y \in \mathcal{Y}$, entonces existe $y' \in \mathcal{Y} \subset (\mathcal{Y} + \mathbb{R}_+^p)$ tal que $y' \leq y$. Sea $d = y - y' \in \mathbb{R}_+^p \setminus \{0\}$, entonces resulta $y = y' + d$ (con $d \neq 0$), lo cual implica que $y \notin (\mathcal{Y} + \mathbb{R}_+^p)_N$, una contradicción. Por lo tanto $(\mathcal{Y} + \mathbb{R}_+^p)_N \subset \mathcal{Y}_N$.

Recíprocamente, supongamos que $y \in \mathcal{Y}_N$, pero $y \notin (\mathcal{Y} + \mathbb{R}_+^p)_N$. Entonces existen $y' \in (\mathcal{Y} + \mathbb{R}_+^p)$ y $d' \in \mathbb{R}_+^p$, $d' \neq 0$, tales que $y = y' + d'$. Por otra parte, existen $y'' \in \mathcal{Y}$ y $d'' \in \mathbb{R}_+^p$ tales que $y' = y'' + d''$. Luego $y = y'' + (d' + d'')$, con $d' + d'' \in \mathbb{R}_+^p$ (por ser \mathbb{R}_+^p convexo, ver sección 2.1) y $d' + d'' \neq 0$. Esto implica que $y \notin \mathcal{Y}_N$, lo cual es una contradicción. Por lo tanto $\mathcal{Y}_N \subset (\mathcal{Y} + \mathbb{R}_+^p)_N$. ■

El resultado de la proposición 2.2.2 es ilustrado en la figura 2.15.

Proposición 2.2.3. $(\mathcal{Y} + \mathcal{Y}')_N \subset \mathcal{Y}_N + \mathcal{Y}'_N$.

Demostración. Sea $\hat{y} \in (\mathcal{Y} + \mathcal{Y}')_N$, entonces existen $y \in \mathcal{Y}$, $y' \in \mathcal{Y}'$ tales que $\hat{y} = y + y'$. Supongamos que $y \notin \mathcal{Y}_N$, entonces existen $y'' \in \mathcal{Y}$ y $d \in \mathbb{R}_+^p$, $d \neq 0$, tales que $y = y'' + d$. Luego $\hat{y} = y'' + y' + d$, con $y'' + y' \in \mathcal{Y} + \mathcal{Y}'$, lo cual implica que $\hat{y} \notin (\mathcal{Y} + \mathcal{Y}')_N$, una contradicción. Por lo tanto $y \in \mathcal{Y}_N$. De igual manera tenemos que $y' \in \mathcal{Y}'_N$ y por lo tanto $\hat{y} = y + y' \in \mathcal{Y}_N + \mathcal{Y}'_N$. ■

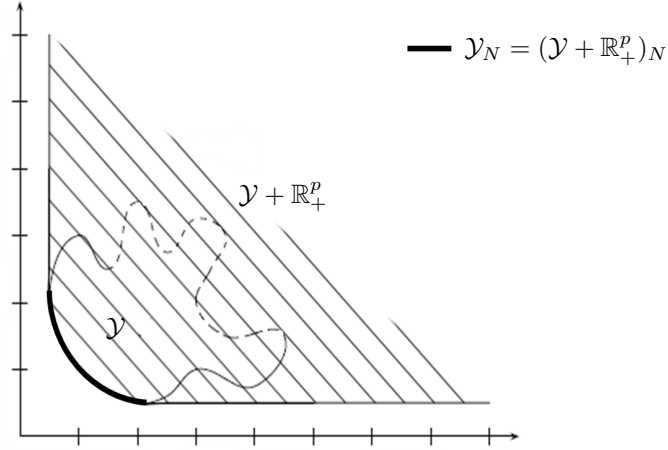


Figura 2.15: Ilustración de Proposición 2.2.2.

Proposición 2.2.4. $(\alpha\mathcal{Y})_N = \alpha\mathcal{Y}_N$ para todo número real $\alpha > 0$.

Demostración. Sea $y \in \alpha\mathcal{Y}_N$, entonces existe $y' \in \mathcal{Y}_N$ tal que $y = \alpha y'$. Puesto que $y' \in \mathcal{Y}$, entonces $y \in \alpha\mathcal{Y}$. Supongamos que $y \notin (\alpha\mathcal{Y})_N$, entonces existen $y'' \in \mathcal{Y}$ y $d \in \mathbb{R}_+^p$, $d \neq 0$, tales que $\alpha y' = y = \alpha y'' + d$. Luego $\alpha(y' - y'') = d$ y por lo tanto $y' - y'' \in \mathbb{R}_+^p \setminus \{0\}$. Luego $y'' \leq y'$, lo cual implica que $y' \notin \mathcal{Y}_N$, una contradicción. Por lo tanto $y \in (\alpha\mathcal{Y})_N$.

Recíprocamente, sea $y \in (\alpha\mathcal{Y})_N$. Entonces $y \in \alpha\mathcal{Y}$ y por lo tanto existe $y' \in \mathcal{Y}$ tal que $y = \alpha y'$. Supongamos que $y' \notin \mathcal{Y}_N$, entonces existen $y'' \in \mathcal{Y}$ y $d \in \mathbb{R}_+^p$, $d \neq 0$, tales que $y' = y'' + d$. Luego $y = \alpha y' = \alpha y'' + \alpha d$, con $\alpha d \in \mathbb{R}_+^p \setminus \{0\}$ (por ser \mathbb{R}_+^p un cono, ver sección 2.1), lo cual implica que $y \notin (\alpha\mathcal{Y})_N$, una contradicción. Por lo tanto $y' \in \mathcal{Y}_N$ y de esta manera $y = \alpha y' \in \alpha\mathcal{Y}_N$. ■

La siguiente proposición muestra que el conjunto \mathcal{Y}_N está contenido en la frontera del conjunto factible \mathcal{Y} .

Proposición 2.2.5. $\mathcal{Y}_N \subset \partial\mathcal{Y}$.

Demostración. Sea $y \in \mathcal{Y}$ tal que $y \notin \partial\mathcal{Y}$. Entonces $y \in \mathcal{Y}^\circ$ y por lo tanto existe un $\varepsilon > 0$ tal que $B(y, \varepsilon) \subset \mathcal{Y}$, donde $B(y, \varepsilon) = y + B(0, \varepsilon)$ es la bola abierta de centro y y radio ε . Por lo tanto, dado $d \in \mathbb{R}_+^p \setminus \{0\}$, existe un $\alpha > 0$ tal que $\alpha d \in B(0, \varepsilon)$. Luego $y - \alpha d \in \mathcal{Y}$, con $\alpha d \in \mathbb{R}_+^p \setminus \{0\}$, lo cual implica que $y \notin \mathcal{Y}_N$. ■

Corolario 2.2.1. Si $\mathcal{Y} \subset \mathbb{R}^p$ es abierto ó $\mathcal{Y} + \mathbb{R}_+^p$ es abierto, entonces $\mathcal{Y}_N = \emptyset$.

Demostración. Se sigue de las proposiciones 2.2.2 y 2.2.5. ■

Los resultados teóricos más básicos en optimización multiobjetivo tratan sobre condiciones suficientes para la existencia de soluciones eficientes y puntos no dominados. En este trabajo presentaremos algunos resultados sobre existencia de puntos (débilmente) no dominados, los cuales naturalmente implican existencia de soluciones (débilmente) eficientes.

Teorema 2.2.1. Si $\mathcal{Y} \subset \mathbb{R}^p$ es no vacío y compacto, entonces $\mathcal{Y}_{wN} \neq \emptyset$.

Demostración. Supongamos que $\mathcal{Y}_{wN} = \emptyset$. Entonces por (2.7) tenemos que para todo $y \in \mathcal{Y}$, existe $y' \in \mathcal{Y}$ tal que $y' < y$, es decir $y \in y' + (\mathbb{R}_+^p)^\circ$. Por lo tanto

$$\mathcal{Y} \subset \bigcup_{y \in \mathcal{Y}} \{y + (\mathbb{R}_+^p)^\circ\}.$$

Dado que cada conjunto $y + (\mathbb{R}_+^p)^\circ$ es abierto, entonces la familia $\{y + (\mathbb{R}_+^p)^\circ\}_{y \in \mathcal{Y}}$ es un cubrimiento por abiertos de \mathcal{Y} . De la compacidad de \mathcal{Y} tenemos que existe un subcubrimiento finito, es decir

$$\mathcal{Y} \subset \bigcup_{i=1}^n \{y^i + (\mathbb{R}_+^p)^\circ\}.$$

De esta manera, para cada $i = 1, \dots, n$ existe algún $j = 1, \dots, n$ tal que $y^i \in y^j + (\mathbb{R}_+^p)^\circ$, es decir $y^j < y^i$. Puesto que el número de índices es finito, esto implica que para cada $i = 1, \dots, n$ existen índices $\{j_1, \dots, j_m\} \subset \{1, \dots, n\}$ tales que

$$y^i > y^{j_1} > \dots > y^{j_m} > y^i$$

lo cual es absurdo. Por lo tanto $\mathcal{Y}_{wN} \neq \emptyset$. ■

La \mathbb{R}_+^p -semicompacidad, definida a continuación, es una propiedad que garantiza la existencia de puntos no dominados.

Definición 2.2.3. *Un conjunto $\mathcal{Y} \subset \mathbb{R}^p$ se dice que es \mathbb{R}_+^p -semicompacto si todo cubrimiento por abiertos de \mathcal{Y} de la forma $\{(y^i - \mathbb{R}_+^p)^c \mid y^i \in \mathcal{Y}, i \in \mathcal{I}\}$, con \mathcal{I} un conjunto de índices, admite un subcubrimiento finito.*

Teorema 2.2.2. *Si $\mathcal{Y} \subset \mathbb{R}^p$ es no vacío y \mathbb{R}_+^p -semicompacto, entonces $\mathcal{Y}_N \neq \emptyset$.*

Demostración. Puesto que un elemento de \mathcal{Y}_N es un elemento minimal respecto al orden de Pareto, deduciremos el teorema al demostrar que \mathcal{Y} es inductivamente ordenado (toda cadena en \mathcal{Y} tiene una cota inferior) y aplicaremos el lema de Zorn para obtener un elemento $y \in \mathcal{Y}_N$.

Supongamos lo contrario, es decir, que \mathcal{Y} no es inductivamente ordenado. Entonces existe una cadena $\mathcal{Y}^{\mathcal{I}} = \{y^i\}_{i \in \mathcal{I}}$ en \mathcal{Y} que no posee cota inferior. Por lo tanto

$$\bigcap_{i \in \mathcal{I}} \{\mathcal{Y} \cap (y^i - \mathbb{R}_+^p)\} = \emptyset$$

pues de existir un elemento \hat{y} en la intersección tendríamos que $\hat{y} \leq y^i$ para todo $i \in \mathcal{I}$, y por lo tanto \hat{y} sería una cota inferior de $\mathcal{Y}^{\mathcal{I}}$. Luego, para cada $y \in \mathcal{Y}$ existe $y^i \in \mathcal{Y}^{\mathcal{I}}$ tal que $y \notin y^i - \mathbb{R}_+^p$, por lo que $\{(y^i - \mathbb{R}_+^p)^c\}_{i \in \mathcal{I}}$ es un cubrimiento por abiertos de \mathcal{Y} . Puesto que $y^i - \mathbb{R}_+^p \subset y^j - \mathbb{R}_+^p$ si y sólo si $y^i \leq y^j$, entonces estos conjuntos están totalmente ordenados por la inclusión. Dado que por hipótesis \mathcal{Y} es \mathbb{R}_+^p -semicompacto, entonces este cubrimiento por abiertos admite un subcubrimiento finito y por estar totalmente ordenados existe $y^{i_0} \in \mathcal{Y}^{\mathcal{I}}$ tal que

$$\mathcal{Y} \subset (y^{i_0} - \mathbb{R}_+^p)^c$$

lo cual contradice el hecho de que $y^{i_0} \in \mathcal{Y}$. Luego, \mathcal{Y} es inductivamente ordenado y por el lema de Zorn existe y elemento minimal de \mathcal{Y} , es decir, $y \in \mathcal{Y}_N$. ■

Puesto que la condición de \mathbb{R}_+^p -semicompacidad es difícil de establecer, utilizaremos la propiedad más fuerte de \mathbb{R}_+^p -compacidad definida a continuación.

Definición 2.2.4. Un conjunto $\mathcal{Y} \subset \mathbb{R}^p$ se dice que es \mathbb{R}_+^p -compacto si para todo $y \in \mathcal{Y}$ el conjunto $\mathcal{Y} \cap (y - \mathbb{R}_+^p)$ es compacto.

Proposición 2.2.6. Si $\mathcal{Y} \subset \mathbb{R}^p$ es \mathbb{R}_+^p -compacto, entonces también es \mathbb{R}_+^p -semicompacto.

Demostración. Sea $\{(y^i - \mathbb{R}_+^p)^c \mid y^i \in \mathcal{Y}, i \in \mathcal{I}\}$ un cubrimiento por abiertos de \mathcal{Y} . Para un índice $i_0 \in \mathcal{I}$ arbitrario, tenemos que $\{(y^i - \mathbb{R}_+^p)^c \mid y^i \in \mathcal{Y}, i \in \mathcal{I}, y^i \neq y^{i_0}\}$ es un cubrimiento por abiertos de $\mathcal{Y} \cap (y^{i_0} - \mathbb{R}_+^p)$, el cual es compacto debido a la \mathbb{R}_+^p -compacidad de \mathcal{Y} . Luego éste admite un subcubrimiento finito, el cual conjuntamente con $(y^{i_0} - \mathbb{R}_+^p)^c$ forma un subcubrimiento finito de \mathcal{Y} . Por lo tanto \mathcal{Y} es \mathbb{R}_+^p -semicompacto. ■

Corolario 2.2.2. Si \mathcal{Y} es \mathbb{R}_+^p -compacto, entonces $\mathcal{Y}_N \neq \emptyset$.

Demostración. Se sigue inmediatamente del teorema 2.2.2 y la proposición 2.2.6. ■

Puesto que $\mathcal{Y}_N \subset \mathcal{Y}_{wN}$, entonces $\mathcal{Y}_{wN} \neq \emptyset$ si $\mathcal{Y}_N \neq \emptyset$. La recíproca, sin embargo, no es cierta, como puede verse del ejemplo 7 tomando $\mathcal{Y} = \{[0, 1] \times [0, 1]\} \setminus \{(0, 0)\}$.

2.2.1. Cotas del conjunto no dominado

En esta sección definimos los puntos *ideal* y *nadir* como cotas inferior y superior del conjunto de puntos no dominados. Estos puntos son usados principalmente como puntos de referencia para obtener soluciones de compromiso.

Definición 2.2.5.

a) Un punto $y^I = (y_1^I, \dots, y_p^I) \in \mathbb{R}^p$ se dice que es un punto ideal del problema de optimización multiobjetivo (2.3) si se cumple que

$$y_j^I = \min_{x \in \mathcal{X}} f_j(x) = \min_{y \in \mathcal{Y}} y_j, \quad j = 1, \dots, p.$$

b) Un punto $y^N = (y_1^N, \dots, y_p^N) \in \mathbb{R}^p$ se dice que es un punto nadir del problema de optimización multiobjetivo (2.3) si se cumple que

$$y_j^N = \max_{x \in \mathcal{X}_E} f_j(x) = \max_{y \in \mathcal{Y}_N} y_j, \quad j = 1, \dots, p.$$

En la figura 2.16 se muestran los puntos y^I e y^N del conjunto $\mathcal{Y} = F(\mathcal{X})$. Obsérvese que en la definición de punto nadir, el máximo se toma sobre el conjunto eficiente \mathcal{X}_E . Esto se debe a que el punto $y' = (y'_1, \dots, y'_p) \in \mathbb{R}^p$ definido por $y'_j = \max_{x \in \mathcal{X}} f_j(x) = \max_{y \in \mathcal{Y}} y_j$ puede no ser una cota superior “ajustada” del conjunto \mathcal{Y}_N , como se muestra en la figura 2.16.

También podemos notar de la figura 2.16 que $y^I \notin \mathcal{Y}$. Los puntos ideal y nadir no tienen por qué pertenecer al conjunto factible, pero si ocurre que $y^I \in \mathcal{Y}$, entonces este punto sería claramente solución al problema de optimización multiobjetivo (2.3) y en este caso tendríamos que $\mathcal{Y}_N = \{y^I\}$. Esto generalmente no ocurre, pues como ya hemos mencionado, en la gran mayoría de los casos los objetivos están en conflicto y por lo tanto no es posible encontrar una única solución al problema (2.3).

De la definición 2.2.5 vemos que el punto ideal puede ser hallado al resolver p problemas de optimización de un objetivo, lo cual, en términos de optimización multiobjetivo, resulta

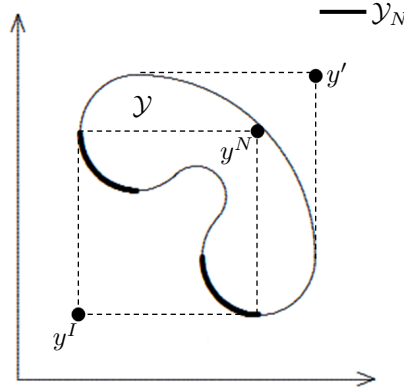


Figura 2.16: Puntos ideal y nadir.

relativamente simple. Sin embargo, el cálculo del punto nadir requiere resolver p problemas de optimización restringidos al conjunto eficiente, algo que resulta mucho más difícil, si uno considera la complejidad que puede tener el cálculo del conjunto \mathcal{X}_E . De hecho, no se conoce todavía un método efectivo para obtener el punto y^N . Por este motivo, diferentes métodos heurísticos son usados para obtener aproximaciones al punto nadir, como el uso de *tablas payoff* (Ehrgott [7], pág. 35), aunque el resultado de utilizar tal tendencia puede llevar a soluciones que sobreestiman o subestiman en gran medida el valor de y^N . No obstante, existe un caso en el cual el punto nadir puede ser hallado con exactitud, y es cuando $p = 2$, es decir, en el caso de problemas de optimización bi-objetivos. Las etapas para hallar $y^N = (y_1^N, y_2^N)$ en tales casos son descriptas a continuación:

- 1) Resolver los problemas de optimización de un objetivo

$$y_j^I = \min_{x \in \mathcal{X}} f_j(x), \quad j = 1, 2,$$

y obtener de esta manera el punto ideal $y^I = (y_1^I, y_2^I)$.

- 2) Resolver los problemas de optimización de un objetivo con restricciones

$$y_1^N = \min_{x \in \mathcal{X}} f_1(x) \quad \text{sujeto a} \quad f_2(x) \leq y_2^I.$$

$$y_2^N = \min_{x \in \mathcal{X}} f_2(x) \quad \text{sujeto a} \quad f_1(x) \leq y_1^I.$$

El punto $y^N = (y_1^N, y_2^N)$ obtenido de esta manera es el punto nadir del problema de optimización bi-objetivo.

De la definición de punto nadir puede verse que este procedimiento para el caso bi-objetivo funciona (esto puede ser chequeado fácilmente en la figura 2.16). Sin embargo, éste no puede ser generalizado para el caso $p > 2$, pues en tales casos no conocemos cuáles objetivos deben ser fijados en la etapa 2 del procedimiento anterior.

2.2.2. Caracterización geométrica de las soluciones eficientes

Los conjuntos \mathcal{X}_E y \mathcal{X}_{wE} pueden ser caracterizados geoméricamente. Para tal fin introducimos los conceptos de conjunto de nivel y curva de nivel de una función.

Definición 2.2.6. Un conjunto de nivel de una función $f : \mathcal{X} \subset \mathbb{R}^n \rightarrow \mathbb{R}$ en $\hat{x} \in \mathcal{X}$ es el conjunto

$$\{x \in \mathcal{X} \mid f(x) \leq f(\hat{x})\}$$

y una curva de nivel de f en $\hat{x} \in \mathcal{X}$ es el conjunto

$$\{x \in \mathcal{X} \mid f(x) = f(\hat{x})\}$$

Finalmente, el conjunto de nivel estricto de f en $\hat{x} \in \mathcal{X}$ es definido como el conjunto

$$\{x \in \mathcal{X} \mid f(x) < f(\hat{x})\}$$

Denotaremos al conjunto de nivel, curva de nivel y conjunto de nivel estricto de f en $\hat{x} \in \mathcal{X}$ por $\{f(x) \leq f(\hat{x})\}$, $\{f(x) = f(\hat{x})\}$ y $\{f(x) < f(\hat{x})\}$ respectivamente. Obsérvese que $\hat{x} \in \{f(x) \leq f(\hat{x})\}$ y $\hat{x} \in \{f(x) = f(\hat{x})\}$, pero $\hat{x} \notin \{f(x) < f(\hat{x})\}$.

Consideremos como ejemplo un problema de optimización multiobjetivo (2.3), con $n = p = 2$, y consideremos también los conjuntos y curvas de nivel en el espacio variable \mathbb{R}^2 de las funciones objetivos hipotéticas f_1 y f_2 en \hat{x} , como se muestra en la figura 2.17.

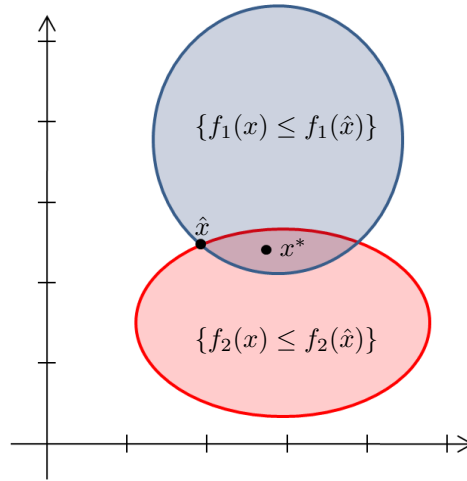


Figura 2.17: Curvas y conjuntos de nivel de f_1 y f_2 en \hat{x} .

Como podemos observar, \hat{x} no puede ser eficiente, pues es posible seleccionar un punto $x^* \in \{f_1(x) < f_1(\hat{x})\} \cap \{f_2(x) < f_2(\hat{x})\}$, es decir, x^* es tal que $f_1(x^*) < f_1(\hat{x})$ y $f_2(x^*) < f_2(\hat{x})$. Mas aún, de lo anterior se desprende que \hat{x} tampoco puede ser débilmente eficiente. Para que esto suceda deberíamos tener $\{f_1(x) < f_1(\hat{x})\} \cap \{f_2(x) < f_2(\hat{x})\} = \emptyset$. El siguiente teorema expone estas ideas más claramente.

Teorema 2.2.3. Dado un problema de optimización multiobjetivo (2.3), sea $\hat{x} \in \mathcal{X}$ y consideremos las curvas y conjuntos de nivel de las funciones objetivos f_1, \dots, f_p en \hat{x} . Entonces

1) \hat{x} es una solución eficiente si y sólo si

$$\bigcap_{i=1}^p \{f_i(x) \leq f_i(\hat{x})\} = \bigcap_{i=1}^p \{f_i(x) = f_i(\hat{x})\}.$$

2) \hat{x} es una solución débilmente eficiente si y sólo si

$$\bigcap_{i=1}^p \{f_i(x) < f_i(\hat{x})\} = \emptyset.$$

Demostración.

- 1) \hat{x} es una solución eficiente si y sólo si no existe $x \in \mathcal{X}$ tal que $f_i(x) \leq f_i(\hat{x}) \forall i = 1, \dots, p$ y $f_j(x) < f_j(\hat{x})$ para algún $j = 1, \dots, p$, si y sólo si no existe $x \in \mathcal{X}$ tal que $x \in \bigcap_{i=1}^p \{f_i(x) \leq f_i(\hat{x})\}$ y $x \in \{f_j(x) < f_j(\hat{x})\}$ para algún $j = 1, \dots, p$, si y sólo si $\bigcap_{i=1}^p \{f_i(x) \leq f_i(\hat{x})\} = \bigcap_{i=1}^p \{f_i(x) = f_i(\hat{x})\}$.
- 2) \hat{x} es una solución débilmente eficiente si y sólo si no existe $x \in \mathcal{X}$ tal que $f_i(x) < f_i(\hat{x}) \forall i = 1, \dots, p$, si y sólo si no existe $x \in \mathcal{X}$ tal que $x \in \bigcap_{i=1}^p \{f_i(x) < f_i(\hat{x})\}$, si y sólo si $\bigcap_{i=1}^p \{f_i(x) < f_i(\hat{x})\} = \emptyset$. ■

Como podemos ver del teorema 2.2.3, si \hat{x} es una solución (débilmente) eficiente, entonces las curvas de nivel son tangentes en el punto \hat{x} .

Una última observación, la cual involucra los gradientes de las funciones objetivos, motivará la introducción de un método importante de caracterización de soluciones eficientes. Siguiendo con el ejemplo bi-objetivo de la figura 2.17, supongamos el caso en que el punto \hat{x} resulta ser una solución débilmente eficiente, es decir, el caso en donde $\{f_1(x) < f_1(\hat{x})\} \cap \{f_2(x) < f_2(\hat{x})\} = \emptyset$, como se muestra en la figura 2.18. Supongamos además que $\nabla f_i(\hat{x}) \neq 0, i = 1, 2$.

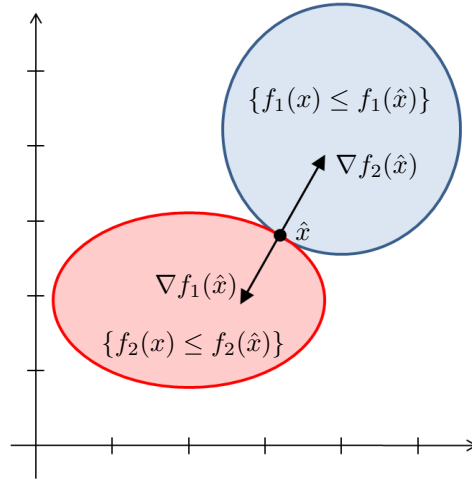


Figura 2.18: \hat{x} es una solución (débilmente) eficiente.

Puesto que $\nabla f_i(\hat{x})$ apunta en la dirección de máximo crecimiento de f_i y por lo tanto es normal a la curva de nivel $\{f_i(x) = f_i(\hat{x})\}, i = 1, 2$, entonces podemos observar que \hat{x} es una solución débilmente eficiente si existe un número real $\lambda > 0$ tal que

$$\nabla f_1(\hat{x}) = -\lambda \nabla f_2(\hat{x}).$$

Una generalización de esta idea nos muestra que \hat{x} es una solución débilmente eficiente si existen números reales $\lambda_1 \geq 0$ y $\lambda_2 \geq 0$ tales que

$$\lambda_1 \nabla f_1(\hat{x}) = -\lambda_2 \nabla f_2(\hat{x}),$$

o equivalentemente

$$\lambda_1 \nabla f_1(\hat{x}) + \lambda_2 \nabla f_2(\hat{x}) = 0. \tag{2.8}$$

Esta generalización es realizada para incluir el caso $\nabla f_1(\hat{x}) = 0$ ó $\nabla f_2(\hat{x}) = 0$. Obsérvese que el lado izquierdo de la igualdad (2.8) es la derivada de la función $f = \lambda_1 f_1 + \lambda_2 f_2$ valuada en \hat{x} . Por la teoría de optimización simple (teorema 1.2.2) sabemos que si \hat{x} es un punto de mínimo de f , entonces la igualdad (2.8) es satisfecha (esto es un “si y sólo si” para el caso en que f es convexa, ver teorema 1.2.4). Esto sugiere que, en principio, podría obtenerse las soluciones (débilmente) eficientes de un problema de optimización multiobjetivo al obtener los puntos de mínimo de ciertas combinaciones lineales de las funciones objetivos. Estas ideas son la motivación para el método de escalarización que se introduce en la sección 2.4.1.

2.3. Eficiencia y no dominancia propia

Como podemos ver de la definición dada en la sección 2.2, una solución eficiente x^* es tal que no es posible lograr una disminución en el valor de un objetivo f_i sin provocar el aumento en al menos otro objetivo f_j . Estos balances entre valores de los objetivos pueden resultar ser no acotados, es decir, una pequeña disminución en el objetivo f_i puede provocar un gran aumento en el objetivo f_j . Para eliminar estos balances no acotados se introduce un concepto de eficiencia más fuerte, el de *eficiencia propia*. En esta sección daremos cuatro tipos de definiciones de soluciones propiamente eficientes: según Borwein, Benson, Geoffrion y Kuhn-Tucker. Tanto Geoffrion como Kuhn y Tucker definen eficiencia propia en el sentido de Pareto ($\mathcal{C} = \mathbb{R}_+^p \setminus \{0\}$), mientras que Borwein y Benson tratan con eficiencia propia respecto de conos cerrados y convexos $\mathcal{C} \subset \mathbb{R}^p$ más generales. Antes de introducir éstos necesitamos las siguientes definiciones.

Definición 2.3.1. Sea $\mathcal{Y} \subset \mathbb{R}^p$. Entonces

a) Dado $y \in \mathcal{Y}$, el cono tangente a \mathcal{Y} en el punto y es el conjunto

$$T(\mathcal{Y}, y) = \left\{ h \in \mathbb{R}^p \mid h = \lim_{k \rightarrow \infty} \frac{y_k - y}{t_k} \right\}$$

para cualquier sucesión $\{t_k\} \subset \mathbb{R}$ tal que $t_k \xrightarrow{k \rightarrow \infty} 0$ y para cualquier sucesión $\{y_k\} \subset \mathcal{Y}$ tal que $y_k \xrightarrow{k \rightarrow \infty} y$.

b) El cono generado por \mathcal{Y} es el conjunto

$$P(\mathcal{Y}) = \{h \in \mathbb{R}^p \mid h = \alpha y, \alpha \geq 0, y \in \mathcal{Y}\}.$$

De la definición es inmediato que $P(\mathcal{Y})$ es un cono que contiene al 0. El hecho de que $T(\mathcal{Y}, y)$ es un cono se prueba en la proposición 2.3.1. Cada vector $h \in T(\mathcal{Y}, y)$ es llamado un vector tangente secuencial al conjunto \mathcal{Y} en el punto y . Esta es una generalización de la definición de vector tangente curvilíneo dado en la definición 1.2.3(a). En efecto, si $h = x'(0)$ para alguna curva suave $x(t)$ ($0 \leq t \leq \delta$) contenida en \mathcal{Y} tal que $y = x(0)$, entonces $h = \lim_{k \rightarrow \infty} \frac{y_k - y}{t_k}$ tomando por ejemplo $t_k = \frac{\delta}{k}$ e $y_k = x(t_k)$; sin embargo, un vector tangente secuencial puede no ser un vector tangente curvilíneo. Un ejemplo de cono tangente y cono generado por un conjunto es mostrado en la figura 2.19.

La siguiente proposición muestra algunas propiedades del cono tangente y el cono generado por un conjunto. La notación $cl(S)$ denota la clausura del conjunto S .

Proposición 2.3.1. Sea $\mathcal{Y} \subset \mathbb{R}^p$ e $y \in \mathcal{Y}$. Entonces

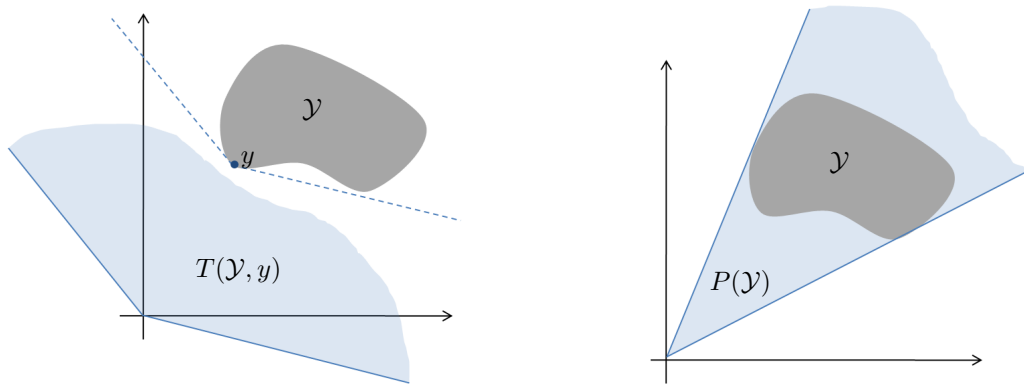


Figura 2.19: Cono tangente y cono generado por \mathcal{Y} .

- 1) El cono tangente $T(\mathcal{Y}, y)$ es un cono cerrado que contiene al 0.
- 2) $T(\mathcal{Y}, y) \subset cl(P(\mathcal{Y} - y))$. Si además el conjunto \mathcal{Y} es convexo, entonces

$$T(\mathcal{Y}, y) = cl(P(\mathcal{Y} - y))$$

Demostración.

- 1) Sea $\frac{y_k - y}{t_k} \rightarrow h \in T(\mathcal{Y}, y)$, con $t_k \rightarrow 0$ e $y_k \rightarrow y$. Si $\alpha > 0$, entonces $\frac{t_k}{\alpha} \rightarrow 0$ y por lo tanto $\frac{\alpha(y_k - y)}{t_k} \rightarrow \alpha h \in T(\mathcal{Y}, y)$, luego $T(\mathcal{Y}, y)$ es un cono. Para ver que es cerrado, sea $\{h_k\}$ una sucesión en $T(\mathcal{Y}, y)$ tal que $h_k \rightarrow h \in \mathbb{R}^p$. Entonces para cada índice k fijo existen sucesiones $\{t_j^k\} \subset \mathbb{R}$ e $\{y_j^k\} \subset \mathcal{Y}$ tales que $t_j^k \xrightarrow{j \rightarrow \infty} 0$, $y_j^k \xrightarrow{j \rightarrow \infty} y$ y $\frac{y_j^k - y}{t_j^k} \xrightarrow{j \rightarrow \infty} h_k$.

Por lo tanto

$$\left\| \frac{y_j^k - y}{t_j^k} - h \right\| \leq \left\| \frac{y_j^k - y}{t_j^k} - h_k \right\| + \|h_k - h\| \xrightarrow{j, k \rightarrow \infty} 0$$

Luego $\frac{y_j^k - y}{t_j^k} \xrightarrow{j, k \rightarrow \infty} h \in T(\mathcal{Y}, y)$ y por lo tanto $T(\mathcal{Y}, y)$ es cerrado.

Es claro que $0 \in T(\mathcal{Y}, y)$, tomando, por ejemplo, $y_k = y$ para todo k .

- 2) Para probar que $T(\mathcal{Y}, y) \subset cl(P(\mathcal{Y} - y))$, sea $h \in T(\mathcal{Y}, y)$. Entonces existen sucesiones $\{t_k\} \subset \mathbb{R}$ e $\{y_k\} \subset \mathcal{Y}$ tales que $t_k \rightarrow 0$, $y_k \rightarrow y$ y $\frac{y_k - y}{t_k} \rightarrow h$. Puesto que para cada índice k fijo tenemos que $y_k - y \in (\mathcal{Y} - y)$, entonces $\frac{1}{t_k}(y_k - y) \in P(\mathcal{Y} - y)$ y por lo tanto $\frac{y_k - y}{t_k} \rightarrow h \in cl(P(\mathcal{Y} - y))$.

Supongamos ahora que \mathcal{Y} es convexo y sean $h, h' \in P(\mathcal{Y})$. Entonces $h = \alpha y$, $h' = \alpha' y'$, con $y, y' \in \mathcal{Y}$ y $\alpha, \alpha' \geq 0$. Puesto que \mathcal{Y} es convexo y $\frac{\alpha}{\alpha + \alpha'} + \frac{\alpha'}{\alpha + \alpha'} = 1$, entonces $\frac{\alpha y}{\alpha + \alpha'} + \frac{\alpha' y'}{\alpha + \alpha'} = \frac{\alpha y + \alpha' y'}{\alpha + \alpha'} \in \mathcal{Y}$. Dado que $\alpha + \alpha' \geq 0$, entonces de la definición de cono generado por un conjunto resulta que $(\alpha + \alpha') \frac{\alpha y + \alpha' y'}{\alpha + \alpha'} = \alpha y + \alpha' y' = h + h' \in P(\mathcal{Y})$ y por lo tanto $P(\mathcal{Y})$ es convexo.

De lo anterior es claro que $cl(P(\mathcal{Y} - y))$ es un cono cerrado y convexo (pues $\mathcal{Y} - y$ es convexo si \mathcal{Y} es convexo). Para probar que $cl(P(\mathcal{Y} - y)) \subset T(\mathcal{Y}, y)$, basta probar que

$P(\mathcal{Y}-y) \subset T(\mathcal{Y}, y)$, pues por la parte 1 de la proposición sabemos que $T(\mathcal{Y}, y)$ es cerrado. Sea entonces $h \in P(\mathcal{Y}-y)$, $h \neq 0$. Luego $h = \alpha(y' - y)$ para algún $y' \in \mathcal{Y}$ y $\alpha > 0$. Definamos la sucesión $\{y_k\}$ por

$$y_k = \left(1 - \frac{1}{k}\right) y + \frac{1}{k} y'$$

Entonces de la convexidad del conjunto \mathcal{Y} tenemos que $\{y_k\} \subset \mathcal{Y}$. Sea la sucesión de números reales $\{t_k\}$ definida por $t_k = \frac{1}{\alpha k}$. Entonces $\frac{1}{t_k}(y_k - y) = \alpha(y' - y)$. Puesto que $t_k \rightarrow 0$ y $y_k \rightarrow y$, entonces

$$\frac{y_k - y}{t_k} \rightarrow \alpha(y' - y) = h \in T(\mathcal{Y}, y)$$

como queríamos demostrar. ■

A continuación definimos eficiencia propia según Borwein.

Definición 2.3.2. (*Eficiencia propia de Borwein*) Un punto $x^* \in \mathcal{X}$ se dice que es una solución propiamente eficiente de (2.3) si

$$T(\mathcal{Y} + \mathbb{R}_+^p, F(x^*)) \cap (-\mathbb{R}_+^p) = \{0\}.$$

En la definición es posible reemplazar el cono \mathbb{R}_+^p por un cono cerrado y convexo arbitrario $\mathcal{C} \subset \mathbb{R}^p$ en el caso de considerar eficiencia respecto del cono \mathcal{C} . La siguiente proposición muestra que las soluciones propiamente eficientes (según Borwein) son también soluciones eficientes.

Proposición 2.3.2. Si $x^* \in \mathcal{X}$ es una solución propiamente eficiente de (2.3) en el sentido de Borwein, entonces x^* es también una solución eficiente.

Demostración. Supongamos que x^* no es una solución eficiente, entonces existen $y \in \mathcal{Y}$ y $d \in \mathbb{R}_+^p$, $d \neq 0$, tales que $F(x^*) = y + d$. Sea la sucesión $\{d_k\} \subset \mathbb{R}_+^p$ definida por

$$d_k = \left(1 - \frac{1}{k}\right) d$$

y sean las sucesiones $\{y_k\} \subset (\mathcal{Y} + \mathbb{R}_+^p)$ y $\{t_k\} \subset \mathbb{R}$ definidas por

$$y_k = y + d_k$$

$$t_k = \frac{1}{k}$$

entonces

$$t_k \rightarrow 0$$

$$y_k = y + d_k = F(x^*) - d + \left(1 - \frac{1}{k}\right) d \rightarrow F(x^*)$$

$$\frac{y_k - F(x^*)}{t_k} = -d \rightarrow -d$$

De esta manera $T(\mathcal{Y} + \mathbb{R}_+^p, F(x^*)) \cap (-\mathbb{R}_+^p) \neq \{0\}$ y por lo tanto x^* no es propiamente eficiente en el sentido de Borwein. ■

La recíproca de la proposición 2.3.2 no es cierta, como puede verse en el siguiente ejemplo.

Ejemplo. Sea $\mathcal{X} = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\}$ y consideremos las funciones objetivos $f_1(x) = x_1$ y $f_2(x) = x_2$ (por lo tanto $\mathcal{X} = \mathcal{Y}$). Entonces los puntos $y_1 = (-1, 0)$ e $y_2 = (0, -1)$ son soluciones eficientes (y también puntos no dominados) pero no son propiamente eficientes en el sentido de Borwein pues $T(\mathcal{Y} + \mathbb{R}_+^2, y_i) \cap (-\mathbb{R}_+^2) = \{(x_1, x_2) \in \mathbb{R}_+^2 \mid x_i = 0\} \neq \{0\}$, $i = 1, 2$, (ver figura 2.20). □

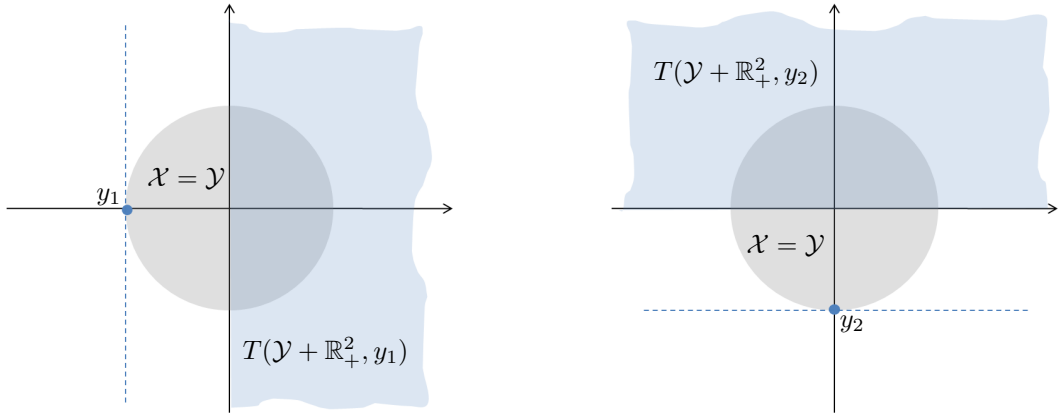


Figura 2.20: Soluciones y_1 e y_2 eficientes pero no propiamente eficientes según Borwein.

A continuación definimos eficiencia propia en el sentido de Benson.

Definición 2.3.3. (Eficiencia propia de Benson) Un punto $x^* \in \mathcal{X}$ se dice que es una solución propiamente eficiente de (2.3) si

$$cl(P(\mathcal{Y} + \mathbb{R}_+^p - F(x^*))) \cap (-\mathbb{R}_+^p) = \{0\}$$

Al igual que en el caso de la eficiencia propia de Borwein, es posible reemplazar el cono \mathbb{R}_+^p por un cono cerrado y convexo arbitrario $\mathcal{C} \subset \mathbb{R}^p$ en el caso de considerar eficiencia en un sentido más amplio.

Las siguientes definiciones resultarán de utilidad.

Definición 2.3.4.

- 1) Un conjunto $\mathcal{Y} \subset \mathbb{R}^p$ se dice que es \mathbb{R}_+^p -convexo si el conjunto $\mathcal{Y} + \mathbb{R}_+^p$ es convexo.
- 2) Un conjunto $\mathcal{Y} \subset \mathbb{R}^p$ se dice que es \mathbb{R}_+^p -cerrado si el conjunto $\mathcal{Y} + \mathbb{R}_+^p$ es cerrado.

Claramente un conjunto convexo es \mathbb{R}_+^p -convexo y un conjunto cerrado es \mathbb{R}_+^p -cerrado. En la figura 2.15 se muestra un conjunto no convexo y no cerrado \mathcal{Y} el cual es \mathbb{R}_+^p -convexo y \mathbb{R}_+^p -cerrado.

El siguiente teorema da una relación entre las eficiencias propias de Borwein y Benson.

Teorema 2.3.1. *Si $x^* \in \mathcal{X}$ es una solución propiamente eficiente de (2.3) en el sentido de Benson, entonces también es una solución propiamente eficiente en el sentido de Borwein. Si \mathcal{X} es un conjunto convexo y los objetivos f_j son funciones convexas (o equivalentemente si el conjunto \mathcal{Y} es \mathbb{R}_+^p -convexo), entonces las definiciones de eficiencia propia de Benson y Borwein coinciden.*

Demostración. Por la proposición 2.3.1 tenemos que $T(\mathcal{Y} + \mathbb{R}_+^p, F(x^*)) \subset cl(P(\mathcal{Y} + \mathbb{R}_+^p - F(x^*)))$ y por lo tanto la eficiencia propia de Benson es más fuerte que la eficiencia propia de Borwein. Si \mathcal{X} es un conjunto convexo y los objetivos f_j son funciones convexas, entonces el conjunto $\mathcal{Y} = F(\mathcal{X})$ es \mathbb{R}_+^p -convexo (Sawaragi et al. [19], proposición 2.1.22) y nuevamente por proposición 2.3.1 resulta $T(\mathcal{Y} + \mathbb{R}_+^p, F(x^*)) = cl(P(\mathcal{Y} + \mathbb{R}_+^p - F(x^*)))$, en cuyo caso ambas definiciones coinciden. ■

A continuación definimos eficiencia propia según Geoffrion.

Definición 2.3.5. (Eficiencia propia de Geoffrion) *Un punto $x^* \in \mathcal{X}$ se dice que es una solución propiamente eficiente de (2.3) si éste es una solución eficiente y si existe un número real $M > 0$ tal que para cada índice i y cada $x \in \mathcal{X}$ satisfaciendo $f_i(x) < f_i(x^*)$, existe un índice j tal que $f_j(x^*) < f_j(x)$ y*

$$\frac{f_i(x^*) - f_i(x)}{f_j(x) - f_j(x^*)} \leq M$$

En la definición de Geoffrion puede verse más claramente el concepto de soluciones eficientes con balance acotado entre los objetivos. El siguiente teorema muestra que en el caso del orden de Pareto, las definiciones de eficiencia propia de Benson y Geoffrion coinciden.

Teorema 2.3.2. *Un punto $x^* \in \mathcal{X}$ es una solución propiamente eficiente de (2.3) según Geoffrion si y sólo si éste es una solución propiamente eficiente según Benson.*

Demostración. Supongamos que x^* es una solución eficiente pero no es propiamente eficiente según Benson. Luego, existe un vector $d \neq 0$ tal que

$$d \in cl(P(\mathcal{Y} + \mathbb{R}_+^p - F(x^*))) \cap (-\mathbb{R}_+^p).$$

Sin pérdida de generalidad, podemos suponer que $d_1 < -1$ y $d_i \leq 0$, $i = 2, \dots, p$ (de lo contrario podemos reordenar los objetivos f_j y reescalar d). Sean las sucesiones $\{t_k\} \subset \mathbb{R}$, $\{d_k\} \subset \mathbb{R}_+^p$ y $\{x_k\} \subset \mathcal{X}$ tales que

$$t_k(F(x_k) + d_k - F(x^*)) \longrightarrow d$$

es decir

$$t_k(f_j(x_k) + (d_k)_j - f_j(x^*)) \longrightarrow d_j \quad \forall j = 1, \dots, p \quad (2.9)$$

con $t_k > 0$. Al tomar una subsucesión de ser necesario podemos suponer que el conjunto de índices

$$I = \{2 \leq i \leq p \mid f_i(x_k) > f_i(x^*)\}$$

sea el mismo para todo índice k . Puesto que x^* es eficiente, tenemos que $I \neq \emptyset$.

Tomemos un número positivo M . Entonces de la convergencia (2.9) existe un número k_0 tal que para todo $k \geq k_0$ se cumple que

$$f_1(x_k) - f_1(x^*) < -\frac{1}{2t_k},$$

$$f_i(x_k) - f_i(x^*) \leq \frac{1}{2Mt_k}, \quad i = 2, \dots, p.$$

Luego, para todo $i \in I$ y $k \geq k_0$ tenemos que

$$0 < f_i(x_k) - f_i(x^*) \leq \frac{1}{2Mt_k},$$

y entonces

$$\frac{f_1(x^*) - f_1(x_k)}{f_i(x_k) - f_i(x^*)} > \frac{2Mt_k}{2t_k} = M.$$

Por lo tanto x^* no es propiamente eficiente en el sentido de Geoffrion.

Recíprocamente, supongamos que x^* es una solución eficiente pero no es propiamente eficiente en el sentido de Geoffrion. Sea $\{M_k\}$ una sucesión no acotada de números reales positivos. Entonces, al reordenar los objetivos f_j de ser necesario, podemos suponer que para cada índice k existe un punto $x_k \in \mathcal{X}$ tal que $f_1(x_k) < f_1(x^*)$ y

$$\frac{f_1(x^*) - f_1(x_k)}{f_i(x_k) - f_i(x^*)} > M_k$$

para todo índice $i = 2, \dots, p$ tal que $f_i(x^*) < f_i(x_k)$. Al tomar una subsucesión de $\{M_k\}$ de ser necesario podemos asumir que el conjunto de índices

$$I = \{2 \leq i \leq p \mid f_i(x^*) < f_i(x_k)\}$$

es el mismo para todo índice k . Puesto que x^* es eficiente, entonces $I \neq \emptyset$.

Definimos la sucesión $\{t_k\} \subset \mathbb{R}$ por

$$t_k = \frac{1}{f_1(x^*) - f_1(x_k)},$$

entonces $t_k > 0$ para todo k . Definimos ahora la sucesión $\{d_k^i\} \subset \mathbb{R}_+^p$ por

$$d_k^i = \begin{cases} 0 & \text{para } i = 1 \text{ ó } i \in I \\ f_i(x^*) - f_i(x_k) & \text{para } i \neq 1, i \notin I \end{cases}$$

entonces

$$t_k(f_i(x_k) + d_k^i - f_i(x^*)) = \begin{cases} -1 & \text{para } i = 1 \\ 0 & \text{para } i \neq 1, i \notin I \end{cases}$$

$$0 < t_k(f_i(x_k) + d_k^i - f_i(x^*)) = t_k(f_i(x_k) - f_i(x^*)) < \frac{1}{M_k}, \quad \forall i \in I.$$

Sea

$$d_i = \lim_{k \rightarrow \infty} [t_k(f_i(x_k) + d_k^i - f_i(x^*))].$$

Claramente $d_1 = -1$ y $d_i = 0$ cuando $i \neq 1$ e $i \notin I$. Puesto que $\{M_k\}$ es una sucesión no acotada de números reales positivos, entonces

$$d_i = 0, \quad \forall i \in I.$$

Por lo tanto

$$d = (-1, 0, \dots, 0) \in cl(P(\mathcal{Y} + \mathbb{R}_+^p - F(x^*))) \cap (-\mathbb{R}_+^p)$$

y x^* no es propiamente eficiente en el sentido de Benson. ■

Finalmente introducimos el concepto de eficiencia propia según Kuhn-Tucker. Para tal fin consideremos el problema de optimización multiobjetivo con restricciones

$$\min_{x \in \mathcal{X}} F(x) = \min_{x \in \mathcal{X}} (f_1(x), \dots, f_p(x)) \quad (2.10)$$

donde el conjunto \mathcal{X} está definido por las restricciones de igualdad y desigualdad

$$\begin{aligned} g_\alpha(x) &\leq 0, \quad \alpha = 1, \dots, l, \\ g_\beta(x) &= 0, \quad \beta = l + 1, \dots, m. \end{aligned}$$

Sea $x^* \in \mathcal{X}$. Al igual que en la sección 1.2.2 consideremos las restricciones tangenciales

$$\langle \nabla g_\alpha(x^*), h \rangle \leq 0, \quad \alpha = \alpha_1, \dots, \alpha_r \quad (\alpha_j \leq l), \quad (2.11a)$$

$$\langle \nabla g_\beta(x^*), h \rangle = 0, \quad \beta = l + 1, \dots, m, \quad (2.11b)$$

donde $\alpha_1, \dots, \alpha_r$ son los índices activos menores o iguales a l . Asumimos que tanto las funciones objetivos f_i como las funciones de restricción g_j son continuamente diferenciables.

Definición 2.3.6. (*Eficiencia propia de Kuhn-Tucker*) Un punto $x^* \in \mathcal{X}$ se dice que es una solución propiamente eficiente del problema multiobjetivo con restricciones (2.10) si éste es eficiente y no existe $h \in \mathbb{R}^p$ tal que satisfaga las restricciones tangenciales (2.11) y que además cumpla

$$\begin{aligned} \langle \nabla f_i(x^*), h \rangle &\leq 0 \quad \forall i = 1, \dots, p, \\ \langle \nabla f_j(x^*), h \rangle &< 0 \quad \text{para algún } j. \end{aligned}$$

Teorema 2.3.3. *Supongamos que el problema de optimización multiobjetivo (2.10) satisface la condición de regularidad de Kuhn-Tucker en el punto $x^* \in \mathcal{X}$ (definición 1.2.3(b)). Si x^* es una solución propiamente eficiente de (2.10) según Geoffrion, entonces es también propiamente eficiente según Kuhn-Tucker.*

Demostración. Supongamos que x^* es una solución eficiente pero no es propiamente eficiente según Kuhn-Tucker. Entonces existe un vector $h \in \mathbb{R}^p$ que satisface las restricciones tangenciales (2.11) y tal que cumple (reordenando los objetivos f_j de ser necesario)

$$\langle \nabla f_i(x^*), h \rangle \leq 0 \quad \forall i = 2, \dots, p, \quad (2.12a)$$

$$\langle \nabla f_1(x^*), h \rangle < 0. \quad (2.12b)$$

Sea $x(t)$ ($0 \leq t \leq \delta$) una curva suave contenida en \mathcal{X} tal que $x^* = x(0)$ y $h = x'(0)$ y consideremos una sucesión $\{t_k\} \subset \mathbb{R}$ de números positivos tal que $t_k \rightarrow 0$. Tomando una subsucesión de $\{t_k\}$ de ser necesario podemos asumir que el conjunto de índices

$$I = \{2 \leq i \leq p \mid f_i(x^*) < f_i(x(t_k))\}$$

es el mismo para todo k . Puesto que x^* es eficiente, entonces $I \neq \emptyset$. Considerando el desarrollo de Taylor de primer orden alrededor de x^* tenemos que

$$f_i(x(t_k)) - f_i(x^*) = t_k \langle \nabla f_i(x^*), h \rangle + o(t_k) > 0, \quad \forall i \in I \quad (2.13)$$

donde el término $o(t_k)$ es tal que $\frac{o(t_k)}{t_k} \rightarrow 0$. Combinando (2.13) con la desigualdad (2.12a) obtenemos

$$\langle \nabla f_i(x^*), h \rangle = 0, \quad \forall i \in I.$$

Luego, de (2.12b) tenemos que $-\langle \nabla f_1(x^*), h \rangle > 0$ y por lo tanto

$$\frac{f_1(x^*) - f_1(x_k)}{f_i(x_k) - f_i(x^*)} = \frac{-\langle \nabla f_1(x^*), h \rangle + \frac{o(t_k)}{t_k}}{\langle \nabla f_i(x^*), h \rangle + \frac{o(t_k)}{t_k}} \rightarrow \infty$$

es decir, x^* no es propiamente eficiente según Geoffrion. ■

Por los teoremas 2.3.1 y 2.3.2 dados anteriormente, tenemos que si x^* es propiamente eficiente en el sentido de Benson, entonces también es propiamente eficiente en el sentido de Geoffrion y Borwein. Si además estamos considerando el problema de optimización multiobjetivo con restricciones (2.10) tal que satisface la condición de regularidad de Kuhn-Tucker en x^* , entonces este punto es también propiamente eficiente según Kuhn-Tucker de acuerdo a los teoremas 2.3.2 y 2.3.3. Por lo tanto, de aquí en adelante siempre nos referiremos a la eficiencia propia en el sentido de Benson, a menos que sea enunciada otra cosa. Denotaremos al conjunto de soluciones propiamente eficientes por \mathcal{X}_{pE} . Si $y = F(x)$, con $x \in \mathcal{X}_{pE}$, entonces llamaremos a y un punto propiamente no dominado. El conjunto de puntos propiamente no dominados es denotado por $\mathcal{Y}_{pN} = F(\mathcal{X}_{pE})$. Claramente tenemos que $\mathcal{X}_{pE} \subset \mathcal{X}_E$ e $\mathcal{Y}_{pN} \subset \mathcal{Y}_N$.

2.4. Caracterización de soluciones eficientes

A continuación introducimos dos tipos de caracterización de soluciones eficientes, las cuales son la base para una gran cantidad de métodos desarrollados para encontrar tales soluciones.

2.4.1. Caracterización por escalarización

Consideremos el problema de optimización multiobjetivo

$$\min_{x \in \mathcal{X}} F(x) = \min_{x \in \mathcal{X}} (f_1(x), \dots, f_p(x)), \quad p \geq 2, \quad (2.14)$$

y para un vector $\lambda = (\lambda_1, \dots, \lambda_p) \in \mathbb{R}_+^p \setminus \{0\}$, definamos la función $f_\lambda = \sum_{i=1}^p \lambda_i f_i$. Veremos cómo están relacionadas las soluciones (débilmente) eficientes y propiamente eficientes del problema de optimización multiobjetivo (2.14) y las soluciones de los problemas de optimización simple

$$\min_{x \in \mathcal{X}} f_\lambda(x) = \min_{x \in \mathcal{X}} \sum_{i=1}^p \lambda_i f_i(x), \quad \lambda \in \mathbb{R}_+^p \setminus \{0\}. \quad (2.15)$$

Estos problemas de optimización de un objetivo también pueden ser escritos como

$$\min_{y \in \mathcal{Y}} \langle \lambda, y \rangle, \quad \lambda \in \mathbb{R}_+^p \setminus \{0\},$$

donde el mínimo se toma sobre el conjunto factible $\mathcal{Y} = F(\mathcal{X})$.

Comenzamos definiendo el conjunto

$$S(\lambda, \mathcal{Y}) = \{\hat{y} \in \mathcal{Y} \mid \langle \lambda, \hat{y} \rangle = \min_{y \in \mathcal{Y}} \langle \lambda, y \rangle\}.$$

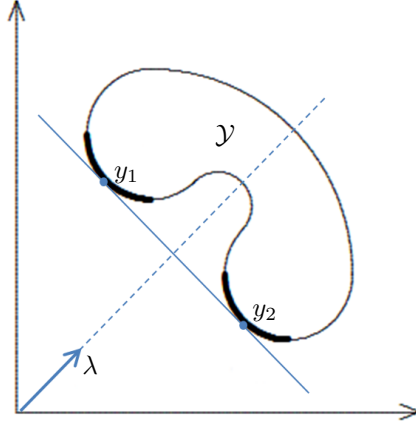


Figura 2.21: Representación geométrica del conjunto $S(\lambda, \mathcal{Y})$.

Un ejemplo es dado en la figura 2.21, donde tenemos que $\lambda \in (\mathbb{R}_+^p)^\circ$ y que $S(\lambda, \mathcal{Y}) = \{y_1, y_2\}$ (un punto $y \in \mathcal{Y}$ está en el conjunto $S(\lambda, \mathcal{Y})$ si éste da la menor proyección sobre la dirección de λ). Nótese también que $S(\lambda, \mathcal{Y}) \subset \mathcal{Y}_N$. Probar esta relación será uno de los objetivos de esta sección.

Es claro que el conjunto $S(\lambda, \mathcal{Y})$ no es modificado si normalizamos el vector λ para obtener $\|\lambda\|_1 = \lambda_1 + \dots + \lambda_p = 1$. Definamos los conjuntos

$$\mathcal{D}(\mathcal{Y}) = \bigcup_{\substack{\lambda \in \mathbb{R}_+^p \setminus \{0\} \\ \|\lambda\|_1 = 1}} S(\lambda, \mathcal{Y}), \quad \mathring{\mathcal{D}}(\mathcal{Y}) = \bigcup_{\substack{\lambda \in (\mathbb{R}_+^p)^\circ \\ \|\lambda\|_1 = 1}} S(\lambda, \mathcal{Y})$$

Claramente tenemos que $\mathring{\mathcal{D}}(\mathcal{Y}) \subset \mathcal{D}(\mathcal{Y})$.

Proposición 2.4.1. *Si el conjunto $\mathcal{Y} \subset \mathbb{R}^p$ es cerrado, entonces $cl(\mathring{\mathcal{D}}(\mathcal{Y})) \subset \mathcal{D}(\mathcal{Y})$.*

Demostración. Sean las sucesiones $\{\lambda_k\} \subset (\mathbb{R}_+^p)^\circ$ e $\{y_k\} \subset \mathring{\mathcal{D}}(\mathcal{Y})$ tales que $\|\lambda_k\|_1 = 1$, $y_k \in S(\lambda_k, \mathcal{Y})$ y además $y_k \rightarrow \hat{y}$. Puesto que la sucesión $\{\lambda_k\}$ está acotada, entonces por el teorema de Bolzano–Weierstrass (Hestenes [9], pag. 407) podemos suponer (tomando una subsucesión de ser necesario) que $\lambda_k \rightarrow \lambda$, con $\|\lambda\|_1 = 1$. Entonces $\lambda \in \mathbb{R}_+^p \setminus \{0\}$.

Puesto que $y_k \in S(\lambda_k, \mathcal{Y})$ para todo k tenemos que

$$\langle \lambda_k, y_k \rangle \leq \langle \lambda_k, y \rangle \quad \forall k, \forall y \in \mathcal{Y}$$

o equivalentemente

$$\langle \lambda_k, y_k - y \rangle \leq 0 \quad \forall k, \forall y \in \mathcal{Y}$$

tomando límite cuando $k \rightarrow \infty$, resulta

$$\langle \lambda, \hat{y} - y \rangle \leq 0 \quad \forall y \in \mathcal{Y}$$

o equivalentemente

$$\langle \lambda, \hat{y} \rangle \leq \langle \lambda, y \rangle \quad \forall y \in \mathcal{Y}$$

y por lo tanto $\hat{y} \in S(\lambda, \mathcal{Y}) \subset \mathcal{D}(\mathcal{Y})$. ■

Una aclaración sobre el producto escalar resultará de utilidad. Sean $x, y \in \mathbb{R}^p$ y consideremos su producto escalar

$$\langle x, y \rangle = \sum_{i=1}^p x_i y_i$$

Supongamos que $y \in \mathbb{R}_+^p \setminus \{0\}$. Como podemos ver directamente de la igualdad anterior, si $x \in \mathbb{R}_+^p$ entonces $\langle x, y \rangle \geq 0$. Por otra parte, si $x \in (\mathbb{R}_+^p)^\circ$, entonces $\langle x, y \rangle > 0$.

A continuación expondremos relaciones entre los conjuntos \mathcal{Y}_N e \mathcal{Y}_{wN} y los conjuntos $\mathring{\mathcal{D}}(\mathcal{Y})$ y $\mathcal{D}(\mathcal{Y})$.

Teorema 2.4.1. *Sea $\mathcal{Y} \subset \mathbb{R}^p$. Entonces $\mathcal{D}(\mathcal{Y}) \subset \mathcal{Y}_{wN}$.*

Demostración. Sea $\hat{y} \notin \mathcal{Y}_{wN}$. Si $\hat{y} \notin \mathcal{Y}$, entonces es claro que $\hat{y} \notin \mathcal{D}(\mathcal{Y})$. Por lo tanto supongamos que $\hat{y} \in \mathcal{Y} \setminus \mathcal{Y}_{wN}$. Luego existe $y \in \mathcal{Y}$ tal que $y < \hat{y}$, es decir, $\hat{y} - y \in (\mathbb{R}_+^p)^\circ$. De esta forma, para cualquier vector $\lambda \in \mathbb{R}_+^p \setminus \{0\}$ se cumple que

$$\langle \lambda, \hat{y} - y \rangle > 0$$

o equivalentemente

$$\langle \lambda, \hat{y} \rangle > \langle \lambda, y \rangle$$

y por lo tanto $\hat{y} \notin \mathcal{D}(\mathcal{Y})$. ■

Teorema 2.4.2. *Sea $\mathcal{Y} \subset \mathbb{R}^p$. Entonces $\mathring{\mathcal{D}}(\mathcal{Y}) \subset \mathcal{Y}_N$.*

Demostración. Sea $\hat{y} \notin \mathcal{Y}_N$. Si $\hat{y} \notin \mathcal{Y}$, entonces es claro que $\hat{y} \notin \mathring{\mathcal{D}}(\mathcal{Y})$. Por lo tanto supongamos que $\hat{y} \in \mathcal{Y} \setminus \mathcal{Y}_N$. Luego existe $y \in \mathcal{Y}$ tal que $y \leq \hat{y}$, es decir, $\hat{y} - y \in \mathbb{R}_+^p \setminus \{0\}$. De esta forma, para cualquier vector $\lambda \in (\mathbb{R}_+^p)^\circ$ se cumple que

$$\langle \lambda, \hat{y} - y \rangle > 0$$

o equivalentemente

$$\langle \lambda, \hat{y} \rangle > \langle \lambda, y \rangle$$

y por lo tanto $\hat{y} \notin \mathring{\mathcal{D}}(\mathcal{Y})$. ■

Algunos de los resultados siguientes requerirán las hipótesis de convexidad. Una propiedad interesante de los conjuntos convexos disjuntos es que pueden ser separados por un hiperplano. En el siguiente teorema denotamos el interior relativo de un conjunto $\mathcal{Y} \subset \mathbb{R}^p$ por $ri(\mathcal{Y})$.

Teorema 2.4.3. *Sean $\mathcal{Y}, \mathcal{Y}'$ conjuntos convexos no vacíos de \mathbb{R}^p . Entonces*

- 1) $ri(\mathcal{Y}) \cap ri(\mathcal{Y}') = \emptyset$ si y sólo si existe un vector $b \neq 0$ y un número real β tal que el hiperplano

$$\mathcal{H} = \{x \in \mathbb{R}^p \mid \langle x, b \rangle = \beta\}$$

separa propiamente a los conjuntos \mathcal{Y} e \mathcal{Y}' , es decir

$$\begin{aligned} \langle y, b \rangle &\leq \beta \quad \forall y \in \mathcal{Y}, \\ \langle y', b \rangle &\geq \beta \quad \forall y' \in \mathcal{Y}'. \end{aligned}$$

2) $0 \notin cl(\mathcal{Y} - \mathcal{Y}')$ si y sólo si existe un vector $b \neq 0$ y un número real β tal que el hiperplano

$$\mathcal{H} = \{x \in \mathbb{R}^p \mid \langle x, b \rangle = \beta\}$$

separa estrictamente a los conjuntos \mathcal{Y} e \mathcal{Y}' , es decir

$$\begin{aligned} \langle y, b \rangle &< \beta \quad \forall y \in \mathcal{Y}, \\ \langle y', b \rangle &> \beta \quad \forall y' \in \mathcal{Y}'. \end{aligned}$$

Una prueba del teorema 2.4.3 es dada en *Rockafellar* [18] (teoremas 11.3 y 11.4).

Teorema 2.4.4. Si \mathcal{Y} es \mathbb{R}_+^p -convexo, entonces $\mathcal{Y}_N \subset \mathcal{D}(\mathcal{Y})$.

Demostración. Sea $\hat{y} \in \mathcal{Y}_N$. Por proposición 2.2.2 tenemos que $\hat{y} \in (\mathcal{Y} + \mathbb{R}_+^p)_N$ y por (2.5) resulta

$$(\mathcal{Y} + \mathbb{R}_+^p) \cap (\hat{y} - \mathbb{R}_+^p) = \{\hat{y}\}$$

o equivalentemente

$$(\mathcal{Y} + \mathbb{R}_+^p - \hat{y}) \cap (-\mathbb{R}_+^p) = \{0\}$$

Luego $ri(\mathcal{Y} + \mathbb{R}_+^p - \hat{y}) \cap ri(-\mathbb{R}_+^p) = \emptyset$ y por teorema 2.4.3 existe un vector $\lambda \neq 0$ tal que el hiperplano

$$\mathcal{H} = \{x \in \mathbb{R}^p \mid \langle x, \lambda \rangle = 0\}$$

separa propiamente ambos conjuntos convexos, es decir

$$\langle y + d - \hat{y}, \lambda \rangle \geq 0 \quad \text{para todo } y \in \mathcal{Y} \text{ y } d \in \mathbb{R}_+^p,$$

$$\langle -d, \lambda \rangle \leq 0 \quad \text{para todo } d \in \mathbb{R}_+^p.$$

La última desigualdad nos dice que $\lambda \in \mathbb{R}_+^p \setminus \{0\}$. De la primera desigualdad (tomando $d = 0$) deducimos que

$$\langle y, \lambda \rangle \geq \langle \hat{y}, \lambda \rangle \quad \text{para todo } y \in \mathcal{Y}$$

y por lo tanto $\hat{y} \in S(\lambda, \mathcal{Y}) \subset \mathcal{D}(\mathcal{Y})$. ■

Teorema 2.4.5. Si \mathcal{Y} es \mathbb{R}_+^p -convexo, entonces $\mathcal{Y}_{wN} = \mathcal{D}(\mathcal{Y})$.

Demostración. Por teorema 2.4.1 basta probar que $\mathcal{Y}_{wN} \subset \mathcal{D}(\mathcal{Y})$. Sea entonces $\hat{y} \in \mathcal{Y}_{wN}$. Por proposición 2.2.1 tenemos que $\mathcal{Y}_{wN} \subset (\mathcal{Y} + (\mathbb{R}_+^p)^\circ)_{wN}$ y por (2.7) resulta

$$\{\mathcal{Y} + (\mathbb{R}_+^p)^\circ\} \cap \{\hat{y} - (\mathbb{R}_+^p)^\circ\} = \emptyset$$

o equivalentemente

$$\{\mathcal{Y} + (\mathbb{R}_+^p)^\circ - \hat{y}\} \cap \{-(\mathbb{R}_+^p)^\circ\} = \emptyset$$

Luego $ri\{\mathcal{Y} + (\mathbb{R}_+^p)^\circ - \hat{y}\} \cap ri\{-(\mathbb{R}_+^p)^\circ\} = \emptyset$ y por teorema 2.4.3 existe un vector $\lambda \neq 0$ tal que el hiperplano

$$\mathcal{H} = \{x \in \mathbb{R}^p \mid \langle x, \lambda \rangle = 0\}$$

separa propiamente ambos conjuntos convexos, es decir

$$\langle y + d - \hat{y}, \lambda \rangle \geq 0 \quad \text{para todo } y \in \mathcal{Y} \text{ y } d \in (\mathbb{R}_+^p)^\circ,$$

$$\langle -d, \lambda \rangle \leq 0 \quad \text{para todo } d \in (\mathbb{R}_+^p)^\circ.$$

La última desigualdad nos dice que $\lambda \in \mathbb{R}_+^p \setminus \{0\}$. De la primera desigualdad (tomando $d = 0$) deducimos que

$$\langle y, \lambda \rangle \geq \langle \hat{y}, \lambda \rangle \quad \text{para todo } y \in \mathcal{Y}$$

y por lo tanto $\hat{y} \in S(\lambda, \mathcal{Y}) \subset \mathcal{D}(\mathcal{Y})$. ■

Proposición 2.4.2. *Si $S(\lambda, \mathcal{Y}) = \{\hat{y}\}$ para algún $\lambda \in \mathbb{R}_+^p \setminus \{0\}$, entonces $\hat{y} \in \mathcal{Y}_N$.*

Demostración. Supongamos que $\hat{y} \notin \mathcal{Y}_N$. Entonces existe $y \in \mathcal{Y}$ tal que $y \leq \hat{y}$, es decir $\hat{y} - y \in \mathbb{R}_+^p \setminus \{0\}$. Por lo tanto

$$\langle \lambda, \hat{y} - y \rangle \geq 0$$

Si $\langle \lambda, \hat{y} - y \rangle = 0$, entonces $\langle \lambda, \hat{y} \rangle = \langle \lambda, y \rangle$, lo cual implica que $y \in S(\lambda, \mathcal{Y})$, con $y \neq \hat{y}$, una contradicción de la hipótesis. Por otra parte, si $\langle \lambda, \hat{y} - y \rangle > 0$, entonces $\langle \lambda, \hat{y} \rangle > \langle \lambda, y \rangle$, lo cual implica que $\hat{y} \notin S(\lambda, \mathcal{Y})$, otra contradicción. Luego $\hat{y} \in \mathcal{Y}_N$. ■

Finalmente exploraremos relaciones entre el conjunto \mathcal{Y}_{pN} y los conjuntos $\mathring{\mathcal{D}}(\mathcal{Y})$ y $\mathcal{D}(\mathcal{Y})$.

Teorema 2.4.6. *Sea $\mathcal{Y} \subset \mathbb{R}^p$. Entonces $\mathring{\mathcal{D}}(\mathcal{Y}) \subset \mathcal{Y}_{pN}$.*

Demostración. Sea $\lambda \in (\mathbb{R}_+^p)^\circ$, $\hat{y} \in S(\lambda, \mathcal{Y})$ y $d \in \mathbb{R}_+^p$ tal que $-d \in cl(P(\mathcal{Y} + \mathbb{R}_+^p - \hat{y}))$. Entonces existen sucesiones $\{y_k\} \subset \mathcal{Y}$, $\{d_k\} \subset \mathbb{R}_+^p$ y $\{\beta_k\} \subset \mathbb{R}$ tales que $\beta_k \geq 0$ y

$$\beta_k(y_k + d_k - \hat{y}) \longrightarrow -d$$

Tomando el producto interno con λ tenemos que

$$\beta_k(\langle \lambda, y_k \rangle + \langle \lambda, d_k \rangle - \langle \lambda, \hat{y} \rangle) \longrightarrow -\langle \lambda, d \rangle$$

Puesto que $\lambda \in (\mathbb{R}_+^p)^\circ$, $\hat{y} \in S(\lambda, \mathcal{Y})$ y $d_k \in \mathbb{R}_+^p$, entonces el lado izquierdo es no negativo y por lo tanto

$$\langle \lambda, d \rangle \leq 0$$

lo cual nos conduce a $d = 0$, es decir

$$cl(P(\mathcal{Y} + \mathbb{R}_+^p - \hat{y})) \cap (-\mathbb{R}_+^p) = \{0\}.$$

Por lo tanto \hat{y} es propiamente eficiente. ■

Antes de dar el siguiente teorema necesitamos la siguiente definición.

Definición 2.4.1. *Sea $\mathcal{Y} \subset \mathbb{R}^p$. El polar de \mathcal{Y} es el conjunto definido por*

$$\mathcal{Y}^* = \{\lambda \in \mathbb{R}^p \mid \langle \lambda, y \rangle \geq 0 \quad \forall y \in \mathcal{Y}\}.$$

*Denotamos por \mathcal{Y}^{**} al polar del conjunto polar \mathcal{Y}^* , es decir*

$$\mathcal{Y}^{**} = \{\mu \in \mathbb{R}^p \mid \langle \mu, \lambda \rangle \geq 0 \quad \forall \lambda \in \mathcal{Y}^*\}.$$

Las siguientes proposiciones serán de utilidad.

Proposición 2.4.3. Sea $\mathcal{Y} \subset \mathbb{R}^p$. Entonces el polar \mathcal{Y}^* es un cono convexo cerrado.

Demostración. Si $\lambda \in \mathcal{Y}^*$, entonces es claro que para todo $\alpha > 0$ se tiene que $\langle \alpha\lambda, y \rangle = \alpha\langle \lambda, y \rangle \geq 0$ para todo $y \in \mathcal{Y}$, por lo tanto \mathcal{Y}^* es un cono. Si $\lambda, \lambda' \in \mathcal{Y}^*$, entonces

$$\langle \lambda + \lambda', y \rangle = \langle \lambda, y \rangle + \langle \lambda', y \rangle \geq 0 \quad \forall y \in \mathcal{Y}$$

y por lo tanto \mathcal{Y}^* es convexo. Para ver que \mathcal{Y}^* es cerrado sea $\{\lambda_k\} \subset \mathcal{Y}^*$ tal que $\lambda_k \rightarrow \hat{\lambda}$, entonces

$$\langle \lambda_k, y \rangle \rightarrow \langle \hat{\lambda}, y \rangle \quad \forall y \in \mathcal{Y}$$

Puesto que $\langle \lambda_k, y \rangle \geq 0$ para todo k y para todo $y \in \mathcal{Y}$, entonces $\langle \hat{\lambda}, y \rangle \geq 0$ para todo $y \in \mathcal{Y}$. Luego $\hat{\lambda} \in \mathcal{Y}^*$. ■

Proposición 2.4.4. Sea $\mathcal{C} \subset \mathbb{R}^p$ un cono convexo cerrado. Entonces $\mathcal{C}^{**} \subset \mathcal{C}$.

Demostración. Sea $\hat{y} \in \mathbb{R}^p$ tal que $\hat{y} \notin \mathcal{C}$. Puesto que \mathcal{C} es cerrado, entonces $0 \notin cl(\mathcal{C} - \{\hat{y}\})$ y por teorema 2.4.3 tenemos que existen $\lambda \neq 0$ en \mathbb{R}^p y $\beta \in \mathbb{R}$ tales que

$$\langle \hat{y}, \lambda \rangle < \beta, \tag{2.16a}$$

$$\langle d, \lambda \rangle > \beta \quad \forall d \in \mathcal{C}. \tag{2.16b}$$

Puesto que \mathcal{C} es un cono y $\langle \alpha d, \lambda \rangle \xrightarrow{\alpha \rightarrow 0} 0$, entonces por (2.16b) tenemos que $\beta < 0$ y por lo tanto $\langle \hat{y}, \lambda \rangle < 0$ por (2.16a). Por otra parte, no puede suceder que $\langle d', \lambda \rangle < 0$ para algún $d' \in \mathcal{C}$, pues de lo contrario tendríamos que $\langle \alpha d', \lambda \rangle \xrightarrow{\alpha \rightarrow \infty} -\infty$, contradiciendo (2.16b). Luego debe ser

$$\langle d, \lambda \rangle \geq 0 \quad \forall d \in \mathcal{C}$$

y por lo tanto $\lambda \in \mathcal{C}^*$. Dado que $\langle \hat{y}, \lambda \rangle < 0$, entonces $\hat{y} \notin \mathcal{C}^{**}$. ■

Teorema 2.4.7. Sea $\mathcal{Y} \subset \mathbb{R}^p$ un conjunto \mathbb{R}_+^p -convexo. Entonces $\mathcal{Y}_{pN} = \mathring{\mathcal{D}}(\mathcal{Y})$.

Demostración. La inclusión $\mathring{\mathcal{D}}(\mathcal{Y}) \subset \mathcal{Y}_{pN}$ es dada en el teorema 2.4.6. Probaremos la inclusión recíproca. Sea $\hat{y} \in \mathcal{Y}_{pN}$, es decir

$$cl(P(\mathcal{Y} + \mathbb{R}_+^p - \hat{y})) \cap (-\mathbb{R}_+^p) = \{0\}.$$

Para simplificar denotemos $\mathcal{C} = cl(P(\mathcal{Y} + \mathbb{R}_+^p - \hat{y}))$. Por proposición 2.3.1, \mathcal{C} es un cono convexo cerrado.

Supongamos que $\mathcal{C}^* \cap (\mathbb{R}_+^p)^\circ = \emptyset$. Entonces por teorema 2.4.3 existen $\mu \neq 0$ en \mathbb{R}^p y $\beta \in \mathbb{R}$ tales que

$$\langle d, \mu \rangle \leq \beta \quad \forall d \in (\mathbb{R}_+^p)^\circ,$$

$$\langle \nu, \mu \rangle \geq \beta \quad \forall \nu \in \mathcal{C}^*.$$

Puesto que $(\mathbb{R}_+^p)^\circ$ y \mathcal{C}^* son conos (proposición 2.4.3) y además $\langle \alpha \nu, \mu \rangle \xrightarrow{\alpha \rightarrow 0} 0$ y $\langle \alpha d, \mu \rangle \xrightarrow{\alpha \rightarrow 0} 0$, entonces $\beta = 0$ y por lo tanto

$$\langle d, \mu \rangle \leq 0 \quad \forall d \in (\mathbb{R}_+^p)^\circ, \tag{2.17a}$$

$$\langle \nu, \mu \rangle \geq 0 \quad \forall \nu \in \mathcal{C}^*. \tag{2.17b}$$

Dado $\varepsilon > 0$, para cada $j = 1, \dots, p$ definimos el vector $d^j = (d_1^j, \dots, d_p^j) \in (\mathbb{R}_+^p)^\circ$ por

$$d_k^j = \begin{cases} 1 & \text{si } k = j \\ \varepsilon & \text{si } k \neq j \end{cases}$$

Puesto que $\langle d^j, \mu \rangle \xrightarrow{\varepsilon \rightarrow 0} \mu_j$ para todo $j = 1, \dots, p$, entonces por (2.17a) tenemos que $\mu_j \leq 0$ y por lo tanto $\mu \in -\mathbb{R}_+^p \setminus \{0\}$. Por (2.17b) y proposición 2.4.4 resulta que $\mu \in \mathcal{C}^{**} \subset \mathcal{C}$. De esta manera

$$\mu \in \mathcal{C} \cap (-\mathbb{R}_+^p), \quad \mu \neq 0,$$

lo cual contradice el hecho de que \hat{y} es propiamente no dominado. Luego existe $\lambda \in (-\mathbb{R}_+^p)^\circ$ tal que $\lambda \in \mathcal{C}^*$. Dado que claramente $(\mathcal{Y} - \hat{y}) \subset \mathcal{C}$, entonces

$$\langle \lambda, y - \hat{y} \rangle \geq 0 \quad \forall y \in \mathcal{Y}$$

o equivalentemente

$$\langle \lambda, \hat{y} \rangle \leq \langle \lambda, y \rangle \quad \forall y \in \mathcal{Y}$$

es decir, $\hat{y} \in S(\lambda, \mathcal{Y}) \subset \mathring{\mathcal{D}}(\mathcal{Y})$. ■

La siguiente proposición, cuya prueba es dada en *Rockafellar* [18] (corolario 37.3.2) será de utilidad.

Proposición 2.4.5. Sean C y D conjuntos convexos cerrados no vacíos en \mathbb{R}^m y \mathbb{R}^n respectivamente, y sea K una función continua cóncava-convexa sobre $C \times D$. Si C ó D es acotado, entonces

$$\inf_{v \in D} \sup_{u \in C} K(u, v) = \sup_{u \in C} \inf_{v \in D} K(u, v).$$

Si en la proposición anterior ambos conjuntos C y D resultan ser acotados (es decir, compactos), entonces obtenemos el teorema de minimax de *Sion-Kakutani*:

$$\min_{v \in D} \max_{u \in C} K(u, v) = \max_{u \in C} \min_{v \in D} K(u, v).$$

Este resultado será usado en la demostración del siguiente teorema.

Teorema 2.4.8. Si $\mathcal{Y} \subset \mathbb{R}^p$ es \mathbb{R}_+^p -convexo y \mathbb{R}_+^p -cerrado, entonces $\mathcal{Y}_N \subset cl(\mathring{\mathcal{D}}(\mathcal{Y})) = cl(\mathcal{Y}_{pN})$.

Demostración. La igualdad $cl(\mathring{\mathcal{D}}(\mathcal{Y})) = cl(\mathcal{Y}_{pN})$ es una consecuencia del teorema 2.4.7. Puesto que $\mathring{\mathcal{D}}(\mathcal{Y}) = \mathring{\mathcal{D}}(\mathcal{Y} + \mathbb{R}_+^p)$ e $\mathcal{Y}_N = (\mathcal{Y} + \mathbb{R}_+^p)_N$ (proposición 2.2.2), basta suponer que \mathcal{Y} es convexo y cerrado.

Primero empezaremos probando que si $y \in \mathcal{Y}$, entonces $(\mathcal{Y} - y)_N = \mathcal{Y}_N - y$. En efecto, la inclusión $(\mathcal{Y} - y)_N \subset \mathcal{Y}_N - y$ es una consecuencia de la proposición 2.2.3. Para probar la recíproca, sea $y' \in \mathcal{Y}_N - y$. Entonces existe $\hat{y} \in \mathcal{Y}_N$ tal que $y' = \hat{y} - y$. Si $y' \notin (\mathcal{Y} - y)_N$, entonces existe $y'' \in (\mathcal{Y} - y)$ tal que $y' = y'' + d$, con $d \neq 0$ en \mathbb{R}_+^p . Por otra parte, existe $y''' \in \mathcal{Y}$ tal que $y'' = y''' - y$. Luego $\hat{y} = y' + y = y'' + d + y = y''' - y + d + y = y''' + d$, lo cual implica que $\hat{y} \notin \mathcal{Y}_N$. Puesto que esto es un absurdo, debemos tener que $\mathcal{Y}_N - y \subset (\mathcal{Y} - y)_N$.

Probaremos entonces que $\mathcal{Y}_N \subset cl(\mathring{\mathcal{D}}(\mathcal{Y}))$. Sea $\hat{y} \in \mathcal{Y}_N$. Por lo probado anteriormente, podemos suponer sin pérdida de generalidad que $\hat{y} = 0$. Dividiremos la prueba en dos etapas:

- 1) Supongamos en primer lugar que \mathcal{Y} es compacto y convexo. Dado $\mu \in (\mathbb{R}_+^p)^\circ$, para cada $\varepsilon > 0$ definimos el conjunto

$$C(\varepsilon) = \varepsilon\mu + \mathbb{R}_+^p.$$

Si $\mathcal{B} = cl(B(0,1))$ denota la bola cerrada de centro 0 y radio 1, entonces para ε suficientemente pequeño tenemos que los conjuntos \mathcal{Y} y $C(\varepsilon) \cap \mathcal{B}$ son no vacíos, convexos y compactos (figura 2.22(a)). Aplicando el teorema de minimax de *Sion-Kakutani* al producto escalar $\langle \cdot, \cdot \rangle : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ sobre el conjunto $(C(\varepsilon) \cap \mathcal{B}) \times \mathcal{Y}$, existen $y(\varepsilon) \in \mathcal{Y}$ y $\mu(\varepsilon) \in C(\varepsilon) \cap \mathcal{B}$ tales que

$$\langle \mu, y(\varepsilon) \rangle \leq \langle \mu(\varepsilon), y(\varepsilon) \rangle \leq \langle \mu(\varepsilon), y \rangle \quad \forall y \in \mathcal{Y}, \forall \mu \in C(\varepsilon) \cap \mathcal{B}. \quad (2.18)$$

Puesto que $\hat{y} = 0 \in \mathcal{Y}$, por (2.18) resulta

$$\langle \mu, y(\varepsilon) \rangle \leq 0 \quad \forall \mu \in C(\varepsilon) \cap \mathcal{B}. \quad (2.19)$$

Dado que el conjunto \mathcal{Y} es compacto, existe una sucesión $\{\varepsilon_k\} \subset \mathbb{R}$, con $\varepsilon_k \rightarrow 0$, tal que la sucesión $\{y_k\} = \{y(\varepsilon_k)\}$ converge a un punto de \mathcal{Y} , es decir, $y_k \rightarrow y' \in \mathcal{Y}$. Para cualquier $\mu \in (\mathbb{R}_+^p)^\circ \cap \mathcal{B}$ es claro que existe un $\bar{\varepsilon} > 0$ tal que $\mu \in C(\varepsilon) \cap \mathcal{B}$ para todo $0 < \varepsilon \leq \bar{\varepsilon}$ (figura 2.22(b)), y por lo tanto (2.19) implica que $\langle \mu, y_k \rangle \leq 0$ para k suficientemente grande. Tomando límite cuando $k \rightarrow \infty$ tenemos que $\langle \mu, y' \rangle \leq 0$ y por lo tanto $\langle \alpha\mu, y' \rangle \leq 0$ para todo $\alpha > 0$. Puesto que $(\mathbb{R}_+^p)^\circ$ es un cono, esto implica que

$$\langle \mu, y' \rangle \leq 0 \quad \forall \mu \in (\mathbb{R}_+^p)^\circ,$$

y por lo tanto

$$-y' \in ((\mathbb{R}_+^p)^\circ)^* = \mathbb{R}_+^p.$$

Dado que $\hat{y} = 0 \in \mathcal{Y}_N$, entonces debe ser $y' = 0$ (de lo contrario tendríamos que $y' < \hat{y}$, contradiciendo el hecho de que \hat{y} es un punto no dominado).

Definimos la sucesión $\{\mu_k\}$ por

$$\mu_k = \frac{\mu(\varepsilon_k)}{\|\mu(\varepsilon_k)\|} \in (\mathbb{R}^p)^\circ \cap \partial\mathcal{B}$$

donde $\mu(\varepsilon_k)$ es el μ asociado a ε_k tal que cumple (2.18). Por lo tanto tenemos que

$$\langle \mu_k, y_k \rangle \leq \langle \mu_k, y \rangle \quad \forall y \in \mathcal{Y},$$

y entonces $y_k \in S(\mu_k, \mathcal{Y}) \subset \mathring{\mathcal{D}}(\mathcal{Y})$. Tomando límite resulta $y_k \rightarrow y' = \hat{y} = 0 \in cl(\mathring{\mathcal{D}}(\mathcal{Y}))$.

- 2) Sea \mathcal{Y} cerrado y convexo, pero no necesariamente compacto. Dado que $\mathcal{Y} \cap \mathcal{B}$ es no vacío, convexo, compacto y además $\hat{y} = 0 \in (\mathcal{Y} \cap \mathcal{B})_N$, por la parte 1 existen sucesiones $\{\mu_k\} \subset (\mathbb{R}_+^p)^\circ$ y $\{y_k\} \subset \mathcal{Y}$ tales que $y_k \rightarrow 0$ y $y_k \in S(\mu_k, \mathcal{Y} \cap \mathcal{B})$, con $\|\mu_k\| = 1$. Si k es suficientemente grande, entonces $y_k \in (\mathcal{B})^\circ$ y por lo tanto $y_k \in S(\mu_k, \mathcal{Y}) \subset \mathring{\mathcal{D}}(\mathcal{Y})$. En efecto, si existe $y' \in \mathcal{Y}$ tal que

$$\langle \mu_k, y' \rangle < \langle \mu_k, y_k \rangle, \quad (2.20)$$

entonces $\alpha y' + (1 - \alpha)y_k \in \mathcal{Y} \cap \mathcal{B}$ para $\alpha > 0$ suficientemente pequeño y por (2.20) resulta

$$\langle \mu_k, \alpha y' + (1 - \alpha)y_k \rangle < \langle \mu_k, y_k \rangle,$$

lo cual contradice el hecho de que $y_k \in S(\mu_k, \mathcal{Y} \cap \mathcal{B})$. Luego $y_k \rightarrow \hat{y} = 0 \in cl(\mathring{\mathcal{D}}(\mathcal{Y}))$. ■

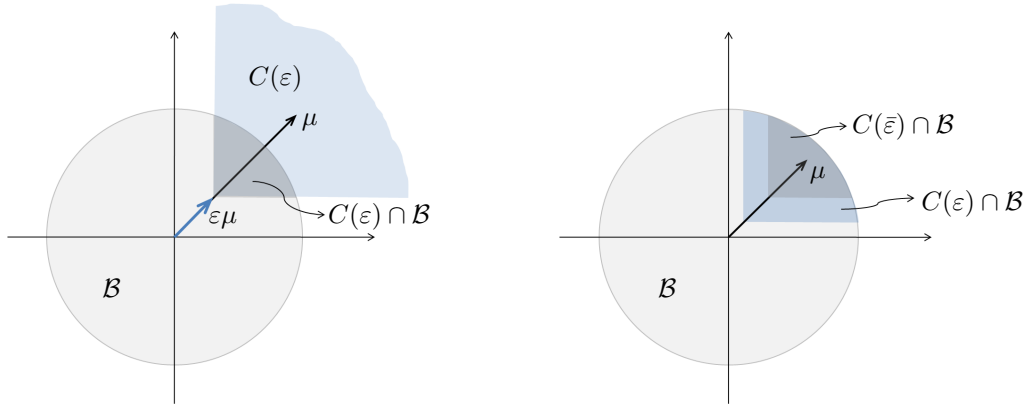


Figura 2.22: (a) Ilustración del conjunto $C(\varepsilon) \cap \mathcal{B}$. (b) $\mu \in C(\varepsilon) \cap \mathcal{B}$ para todo $0 < \varepsilon \leq \bar{\varepsilon}$.

Resumiendo, para conjuntos arbitrarios $\mathcal{Y} \subset \mathbb{R}^p$ tenemos que

$$\mathcal{Y}_{pN} \subset \mathcal{Y}_N \subset \mathcal{Y}_{wN}$$

y además se cumplen las siguientes relaciones con los conjuntos $\mathring{\mathcal{D}}(\mathcal{Y})$ y $\mathcal{D}(\mathcal{Y})$

$$\mathring{\mathcal{D}}(\mathcal{Y}) \subset \mathcal{Y}_{pN} \subset \mathcal{Y}_N \quad \text{y} \quad \mathcal{D}(\mathcal{Y}) \subset \mathcal{Y}_{wN} \quad (2.21)$$

Si \mathcal{Y} es \mathbb{R}_+^p -convexo, entonces

$$\mathring{\mathcal{D}}(\mathcal{Y}) = \mathcal{Y}_{pN} \subset \mathcal{Y}_N \subset \mathcal{Y}_{wN} = \mathcal{D}(\mathcal{Y}) \quad (2.22)$$

La tendencia tradicional para resolver problemas de optimización multiobjetivo en el sentido de Pareto es por técnicas de escalarización, es decir, encontrando las soluciones propiamente o débilmente eficientes del problema de optimización multiobjetivo (2.14) al resolver el problema de optimización simple (2.15). El siguiente teorema muestra la importancia de este método para el caso en que el conjunto \mathcal{Y} es \mathbb{R}_+^p -convexo.

Teorema 2.4.9. *Consideremos el problema de optimización multiobjetivo (2.14) y sea \hat{x} una solución del problema de optimización simple (2.15) para un vector $\lambda \in \mathbb{R}_+^p \setminus \{0\}$. Entonces se cumple*

- 1) Si $\lambda \in (\mathbb{R}_+^p)^\circ$, entonces $\hat{x} \in \mathcal{X}_{pE}$.
- 2) Si $\lambda \in \mathbb{R}_+^p \setminus \{0\}$, entonces $\hat{x} \in \mathcal{X}_{wE}$.
- 3) Si $\lambda \in \mathbb{R}_+^p \setminus \{0\}$ y \hat{x} es la única solución óptima al problema (2.15), entonces $\hat{x} \in \mathcal{X}_E$.

Si \mathcal{X} es un conjunto convexo y los objetivos f_j son funciones convexas (o equivalentemente, si el conjunto \mathcal{Y} es \mathbb{R}_+^p -convexo), entonces además se cumple

- 1') Si $\hat{x} \in \mathcal{X}_{pE}$, entonces existe algún $\lambda \in (\mathbb{R}_+^p)^\circ$ tal que \hat{x} es una solución óptima de (2.15).
- 2') Si $\hat{x} \in \mathcal{X}_{wE}$, entonces existe algún $\lambda \in \mathbb{R}_+^p \setminus \{0\}$ tal que \hat{x} es una solución óptima de (2.15).

Demostración. Los puntos 1 y 2 son una consecuencia inmediata de (2.21) y los puntos 1' y 2' son una consecuencia inmediata de (2.22). El punto 3 es una consecuencia de la proposición 2.4.2. ■

Como puede verse de (2.22), si \mathcal{Y} es \mathbb{R}_+^p -convexo, entonces tomando $\lambda \in \mathbb{R}_+^p \setminus \{0\}$ y resolviendo el problema de optimización simple (2.15), podemos encontrar las soluciones débilmente eficientes del problema de optimización multiobjetivo (2.14). De igual manera, tomando $\lambda \in (\mathbb{R}_+^p)^\circ$ y resolviendo (2.15), podemos encontrar las soluciones propiamente eficientes de (2.14). Sin embargo, encontrar \mathcal{Y}_{wN} no es garantía de tener una aproximación del conjunto \mathcal{Y}_N , pues como se muestra en el ejemplo 7 de la página 25, puede existir una gran diferencia entre un conjunto y otro. Por otra parte, el teorema 2.4.8 nos dice que si \mathcal{Y} es \mathbb{R}_+^p -cerrado y \mathbb{R}_+^p -convexo, entonces

$$\mathcal{Y}_{pN} \subset \mathcal{Y}_N \subset cl(\mathcal{Y}_{pN}),$$

y por lo tanto \mathcal{Y}_{pN} es denso en \mathcal{Y}_N . Esto nos dice que, en este caso, el conjunto \mathcal{Y}_{pN} es una buena aproximación del conjunto \mathcal{Y}_N . La mejor situación se da, no obstante, cuando los objetivos f_j son funciones estrictamente convexas. En este caso la combinación convexa (2.15) también resulta estrictamente convexa (*Rockafellar* [18], pág. 33). Si para cada $\lambda \in \mathbb{R}_+^p \setminus \{0\}$ el problema de optimización simple (2.15) tiene una solución óptima \hat{x} , entonces esta es única (teorema 1.2.3) y por el teorema 2.4.9, punto 3, sabemos que \hat{x} es una solución eficiente de (2.14). Por lo tanto, en esta situación tenemos que

$$\mathcal{Y}_N = \mathcal{Y}_{wN}.$$

En el capítulo 3 usaremos este hecho para derivar el método de continuación global desarrollado en *Pereyra* [16].

Para problemas no convexos, sin embargo, el método de escalarización puede no ser bueno para hallar el conjunto de soluciones eficientes, como se muestra en el siguiente ejemplo.

Ejemplo. Sea $\mathcal{X} = \{(x_1, x_2) \in \mathbb{R}_+^2 \mid x_1^2 + x_2^2 \geq 1\}$ y consideremos los objetivos $f_1(x) = x_1$ y $f_2(x) = x_2$ (por lo tanto $\mathcal{X} = \mathcal{Y}$). Como puede observarse en la figura 2.23, el conjunto de soluciones eficientes y puntos no dominados es $\mathcal{X}_E = \mathcal{Y}_N = \{(y_1, y_2) \in \mathbb{R}_+^2 \mid y_1^2 + y_2^2 = 1\}$. Sin embargo, para cualquier $\lambda \in \mathbb{R}_+^p \setminus \{0\}$ tenemos que $S(\lambda, \mathcal{Y})$ sólo puede contener los puntos $(1, 0)$ ó $(0, 1)$. □

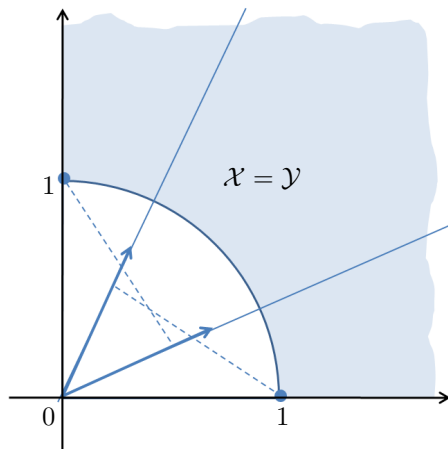


Figura 2.23: Ejemplo no convexo.

2.4.2. Caracterización como solución de problemas con restricciones (ε -Constraint)

El método ε -Constraint es probablemente, después del método de escalarización, la técnica más conocida para resolver problemas de optimización multiobjetivo. En este caso sustituimos el problema (2.14) por p problemas con restricciones del tipo:

$$\min_{x \in \mathcal{X}} f_k(x),$$

sujeto a

$$f_i(x) \leq \varepsilon_i, \quad i = 1, \dots, p, \quad i \neq k,$$

para cada $k = 1, \dots, p$ y $\varepsilon = (\varepsilon_1, \dots, \varepsilon_p) \in \mathbb{R}^p$. El problema con restricciones anterior es mencionado comúnmente en la literatura como el problema $P_k(\varepsilon)$.

Los siguientes teoremas muestran las relaciones entre el problema de optimización multiobjetivo (2.14) y los problemas de optimización con restricciones $P_k(\varepsilon)$.

Teorema 2.4.10. *Un punto $\hat{x} \in \mathcal{X}$ es una solución eficiente del problema (2.14) si y sólo si \hat{x} es una solución del problema $P_k(\hat{\varepsilon})$ para todo $k = 1, \dots, p$, donde $\hat{\varepsilon} = F(\hat{x})$.*

Demostración. Si \hat{x} no es una solución eficiente del problema (2.14), entonces existe $x \in \mathcal{X}$ tal que $f_i(x) \leq f_i(\hat{x}) = \hat{\varepsilon}_i$, donde se cumple la desigualdad estricta para al menos algún $1 \leq k \leq p$. Por lo tanto \hat{x} no es una solución del problema $P_k(\hat{\varepsilon})$.

Recíprocamente, si \hat{x} no es una solución del problema $P_k(\hat{\varepsilon})$ para algún $1 \leq k \leq p$, entonces existe algún $x \in \mathcal{X}$ tal que $f_k(x) < f_k(\hat{x})$ y $f_i(x) \leq f_i(\hat{x})$ para $i \neq k$, por lo que \hat{x} no es una solución eficiente de (2.14). ■

Teorema 2.4.11. *Sea \hat{x} la única solución del problema con restricciones $P_k(\varepsilon)$ para algún $1 \leq k \leq p$ y $\varepsilon \in \mathbb{R}^p$. Entonces \hat{x} es una solución eficiente del problema (2.14).*

Demostración. Sea $x \in \mathcal{X}$, $x \neq \hat{x}$, satisfaciendo $f_i(x) \leq f_i(\hat{x}) \leq \varepsilon_i$ para todo $i \neq k$. Puesto que \hat{x} es la única solución del problema $P_k(\varepsilon)$, entonces debemos tener que $f_k(x) > f_k(\hat{x})$. Por lo tanto \hat{x} es una solución eficiente de (2.14). ■

En la figura 2.24 se muestra un ejemplo del método ε -Constraint para un problema de optimización bi-objetivo. En este caso se ilustra el problema $P_2(\varepsilon)$:

$$\min_{x \in \mathcal{X}} f_2(x),$$

sujeto a

$$f_1(x) \leq \varepsilon_1.$$

El punto \hat{x} es la solución a tal problema. Puesto que éste es la única solución al problema $P_2(\varepsilon)$, entonces el teorema 2.4.11 asegura que $\hat{x} \in \mathcal{X}_E$. En la figura se puede observar directamente que $(f_1(\hat{x}), f_2(\hat{x})) \in \mathcal{Y}_N$.

2.5. Conectividad de conjuntos no dominados

En esta sección estudiaremos una propiedad topológica importante para analizar el conjunto \mathcal{Y}_N , la propiedad de *conectividad*. Este concepto es importante porque si el conjunto de puntos no dominados es conectado, entonces éste puede ser explorado usando ideas de búsqueda local partiendo de un punto no dominado simple. La siguiente definición introduce este concepto.

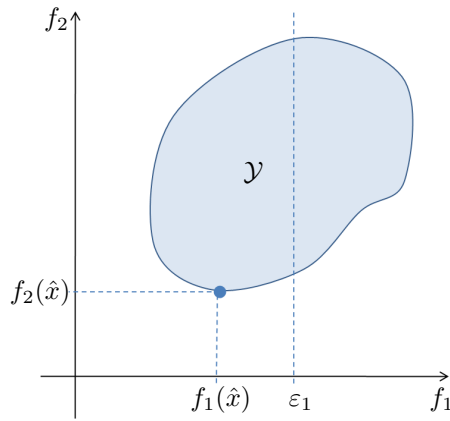


Figura 2.24: Ilustración del método ε -Constraint para un problema bi-objetivo.

Definición 2.5.1. Un conjunto A se dice separable si puede ser escrito como

$$A = A_1 \cup A_2$$

donde

$$cl(A_1) \cap A_2 = A_1 \cap cl(A_2) = \emptyset$$

para conjuntos A_1 y A_2 no vacíos. Equivalentemente, A es separable si existen conjuntos abiertos \mathcal{O}_1 y \mathcal{O}_2 tales que

$$A \subset \mathcal{O}_1 \cup \mathcal{O}_2, \quad A \cap \mathcal{O}_i \neq \emptyset \quad (i = 1, 2), \quad A \cap \mathcal{O}_1 \cap \mathcal{O}_2 = \emptyset.$$

Si el conjunto A no es separable, entonces se dice conectado.

En la figura 2.25 se muestra un ejemplo de un conjunto no dominado conectado y otro que es separable.

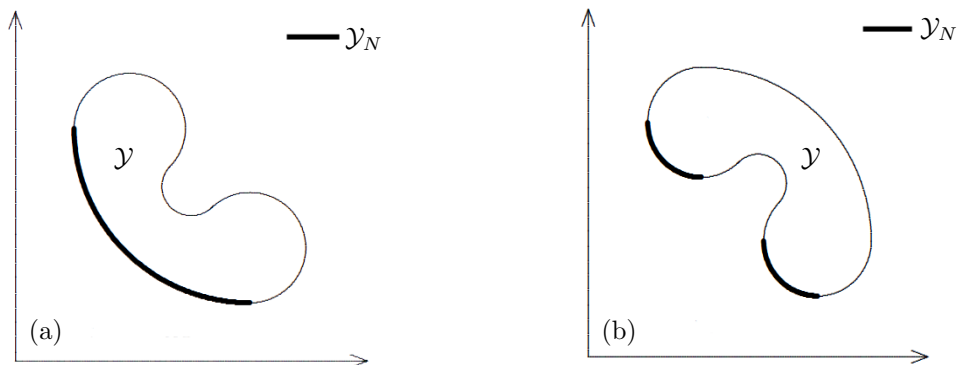


Figura 2.25: (a) Conjunto no dominado conectado. (b) Conjunto no dominado separable.

Los siguientes lemas dados a continuación (sin demostración) serán de utilidad.

Lema 2.5.1. Sean A y B dos conjuntos. Si A es conectado y se cumple que

$$A \subset B \subset cl(A)$$

entonces B es conectado.

Lema 2.5.2. Si $\{A_i \mid i \in \mathcal{I}\}$ es una familia de conjuntos conectados tal que $\bigcap_{i \in \mathcal{I}} A_i \neq \emptyset$, entonces $\bigcup_{i \in \mathcal{I}} A_i$ es conectado.

Los siguientes teoremas establecen que bajo ciertas hipótesis sobre el conjunto $\mathcal{Y} \subset \mathbb{R}^p$, el conjunto no dominado \mathcal{Y}_N resulta ser conectado.

Teorema 2.5.1. Si $\mathcal{Y} \subset \mathbb{R}^p$ es convexo y compacto, entonces $\mathring{\mathcal{D}}(\mathcal{Y})$ es conectado.

Demostración. Supongamos que el conjunto $\mathring{\mathcal{D}}(\mathcal{Y})$ es separable. Entonces existen conjuntos abiertos \mathcal{O}_1 y \mathcal{O}_2 tales que

$$\mathring{\mathcal{D}}(\mathcal{Y}) \subset \mathcal{O}_1 \cup \mathcal{O}_2, \quad \mathring{\mathcal{D}}(\mathcal{Y}) \cap \mathcal{O}_i \neq \emptyset \quad (i = 1, 2), \quad \mathring{\mathcal{D}}(\mathcal{Y}) \cap \mathcal{O}_1 \cap \mathcal{O}_2 = \emptyset.$$

Definamos los conjuntos

$$\mathcal{M}_i = \{\lambda \in (\mathbb{R}_+^p)^\circ \mid S(\lambda, \mathcal{Y}) \cap \mathcal{O}_i \neq \emptyset\} \quad (i = 1, 2).$$

Debido a la convexidad de \mathcal{Y} , el teorema 1.2.3 nos dice que el conjunto $S(\lambda, \mathcal{Y})$ es convexo y por lo tanto conectado. Luego

$$\mathcal{M}_i = \{\lambda \in (\mathbb{R}_+^p)^\circ \mid S(\lambda, \mathcal{Y}) \subset \mathcal{O}_i\} \quad (i = 1, 2),$$

y por lo tanto

$$\mathcal{M}_1 \cap \mathcal{M}_2 = \emptyset. \tag{2.23}$$

Puesto que $\mathring{\mathcal{D}}(\mathcal{Y}) \cap \mathcal{O}_i \neq \emptyset$, entonces $\mathcal{M}_i \neq \emptyset$ y por lo tanto

$$\mathcal{M}_i \cap (\mathbb{R}_+^p)^\circ \neq \emptyset \quad (i = 1, 2). \tag{2.24}$$

Dado que $\mathring{\mathcal{D}}(\mathcal{Y}) \subset \mathcal{O}_1 \cup \mathcal{O}_2$, entonces

$$(\mathbb{R}_+^p)^\circ \subset \mathcal{M}_1 \cup \mathcal{M}_2. \tag{2.25}$$

Mas aún, se cumple que $(\mathbb{R}_+^p)^\circ = \mathcal{M}_1 \cup \mathcal{M}_2$. A continuación probaremos que cada conjunto \mathcal{M}_i es abierto. Consideremos \mathcal{M}_1 y supongamos que éste no es abierto. Entonces su complemento \mathcal{M}_1^c no es cerrado y por lo tanto éste no contiene todos sus puntos de acumulación, es decir, existe $\hat{\lambda} \in \mathcal{M}_1$ y una sucesión $\{\lambda_k\} \subset (\mathbb{R}_+^p)^\circ \setminus \mathcal{M}_1 = \mathcal{M}_2$ tal que $\lambda_k \rightarrow \hat{\lambda}$. Sea la sucesión $\{y^k\} \subset \mathcal{Y}$ tal que $y^k \in S(\lambda_k, \mathcal{Y})$. Puesto que \mathcal{Y} es compacto, podemos suponer sin pérdida de generalidad que $y^k \rightarrow \hat{y} \in \mathcal{Y}$. Mas aún, $\hat{y} \in S(\hat{\lambda}, \mathcal{Y})$, pues de lo contrario existiría un $y' \in \mathcal{Y}$ tal que

$$\langle \hat{\lambda}, y' \rangle < \langle \hat{\lambda}, \hat{y} \rangle$$

y de la continuidad del producto escalar tendríamos que para k suficientemente grande

$$\langle \lambda_k, y' \rangle < \langle \lambda_k, y^k \rangle$$

lo cual contradice el hecho de que $y^k \in S(\lambda_k, \mathcal{Y})$. Puesto que $y^k \in S(\lambda_k, \mathcal{Y}) \subset \mathring{\mathcal{D}}(\mathcal{Y}) \cap \mathcal{O}_2$ y $\mathring{\mathcal{D}}(\mathcal{Y}) \cap \mathcal{O}_1 \cap \mathcal{O}_2 = \emptyset$, entonces $y^k \in \mathcal{O}_1^c$ para todo k , y como \mathcal{O}_1^c es cerrado, entonces $y^k \rightarrow \hat{y} \in \mathcal{O}_1^c$, por lo que $\hat{y} \notin \mathcal{O}_1$, una contradicción. Luego \mathcal{M}_1 es abierto y de igual manera se prueba que \mathcal{M}_2 también lo es.

Finalmente, de (2.23), (2.24) y (2.25) concluimos que $(\mathbb{R}_+^p)^\circ$ es separable, lo que es claramente una contradicción. Luego $\mathring{\mathcal{D}}(\mathcal{Y})$ debe ser conectado. ■

Corolario 2.5.1. Si $\mathcal{Y} \subset \mathbb{R}^p$ es convexo y compacto, entonces \mathcal{Y}_N es conectado.

Demostración. Puesto que en particular tenemos que \mathcal{Y} es \mathbb{R}_+^p -convexo y \mathbb{R}_+^p -cerrado, entonces por teoremas 2.4.2 y 2.4.8 resulta

$$\mathring{\mathcal{D}}(\mathcal{Y}) \subset \mathcal{Y}_N \subset cl(\mathring{\mathcal{D}}(\mathcal{Y})).$$

Por teorema 2.5.1 tenemos que $\mathring{\mathcal{D}}(\mathcal{Y})$ es conectado, por lo tanto \mathcal{Y}_N resulta conectado por lema 2.5.1. ■

El siguiente teorema es una generalización del corolario 2.5.1.

Teorema 2.5.2. Si $\mathcal{Y} \subset \mathbb{R}^p$ es convexo y \mathbb{R}_+^p -compacto, entonces \mathcal{Y}_N es conectado.

Demostración. Dado $d \in (\mathbb{R}_+^p)^\circ$, para cada $\alpha \in \mathbb{R}$ definimos $y(\alpha) = \alpha d$. Entonces para cada $y \in \mathbb{R}^p$ es claro que existe un $\alpha > 0$ tal que $y \in y(\alpha) - \mathbb{R}_+^p$. Luego, dado $\hat{y} \in \mathcal{Y}_N$, existe $\hat{\alpha} > 0$ tal que $\hat{y} \in y(\hat{\alpha}) - \mathbb{R}_+^p$, es decir

$$(y(\hat{\alpha}) - \mathbb{R}_+^p) \cap \mathcal{Y}_N \neq \emptyset.$$

Para cada $\alpha \in \mathbb{R}$, definimos el conjunto

$$\mathcal{Y}(\alpha) = (y(\alpha) - \mathbb{R}_+^p) \cap \mathcal{Y}$$

y por lo tanto tenemos que

$$\mathcal{Y}_N = \bigcup_{\alpha \geq \hat{\alpha}} \mathcal{Y}(\alpha)_N.$$

Puesto que por hipótesis \mathcal{Y} es convexo y \mathbb{R}_+^p -compacto, entonces para cada $\alpha \in \mathbb{R}$ el conjunto $\mathcal{Y}(\alpha)$ es convexo y compacto y por corolario 2.5.1 resulta $\mathcal{Y}(\alpha)_N$ conectado. Por otra parte tenemos que $\mathcal{Y}(\alpha)_N \supset \mathcal{Y}(\hat{\alpha})_N$ para todo $\alpha > \hat{\alpha}$, es decir

$$\bigcap_{\alpha \geq \hat{\alpha}} \mathcal{Y}(\alpha)_N = \mathcal{Y}(\hat{\alpha})_N \neq \emptyset.$$

Por lema 2.5.2 resulta \mathcal{Y}_N conectado. ■

Corolario 2.5.2. Si $\mathcal{Y} \subset \mathbb{R}^p$ es \mathbb{R}_+^p -convexo y el conjunto $\mathcal{Y} + \mathbb{R}_+^p$ es \mathbb{R}_+^p -compacto, entonces \mathcal{Y}_N es conectado.

Demostración. Por hipótesis el conjunto $\mathcal{Y} + \mathbb{R}_+^p$ es convexo y \mathbb{R}_+^p -compacto, por lo tanto aplicando el teorema 2.5.2 y la proposición 2.2.2 resulta $(\mathcal{Y} + \mathbb{R}_+^p)_N = \mathcal{Y}_N$ conectado. ■

En la figura 2.15 puede verse la utilidad del corolario 2.5.2. En efecto, allí tenemos que \mathcal{Y} no es convexo ni \mathbb{R}_+^p -compacto, pero es \mathbb{R}_+^p -convexo y el conjunto $\mathcal{Y} + \mathbb{R}_+^p$ es \mathbb{R}_+^p -compacto. Como puede verse en la figura 2.15, el conjunto \mathcal{Y}_N es conectado.

2.6. Condiciones de optimalidad multiobjetivo

En esta sección probaremos condiciones necesarias y suficientes de Kuhn-Tucker para la existencia de soluciones débilmente y propiamente eficientes del problema de optimización multiobjetivo con restricciones

$$\min_{x \in \mathcal{X}} F(x) = \min_{x \in \mathcal{X}} (f_1(x), \dots, f_p(x)), \quad p \geq 2, \quad (2.26)$$

donde el conjunto \mathcal{X} es dado por restricciones de igualdad y desigualdad de la forma

$$g_\alpha(x) \leq 0, \quad \alpha = 1, \dots, l, \quad (2.27a)$$

$$g_\beta(x) = 0, \quad \beta = l + 1, \dots, m. \quad (2.27b)$$

Estas condiciones de optimalidad serán derivadas bajo la suposición de que tanto las funciones objetivos f_i como las funciones de restricción g_j son continuamente diferenciables. Con estas hipótesis, consideremos las restricciones tangenciales en el punto $\hat{x} \in \mathcal{X}$

$$\langle \nabla g_\alpha(\hat{x}), h \rangle \leq 0, \quad \alpha = \alpha_1, \dots, \alpha_r \quad (\alpha_j \leq l), \quad (2.28a)$$

$$\langle \nabla g_\beta(\hat{x}), h \rangle = 0, \quad \beta = l + 1, \dots, m, \quad (2.28b)$$

donde $\alpha_1, \dots, \alpha_r$ son los índices activos menores o iguales a l .

Antes de continuar introducimos dos teoremas de alternativas que resultarán de utilidad. Aquí denotamos el orden \leq en \mathbb{R}^p por

$$x \leq y \iff x_i \leq y_i, \quad \forall i = 1, \dots, p.$$

Teorema 2.6.1. *(Teorema de la alternativa de Motzkin) Sean las matrices A , B y C , con A no nula. Entonces se cumple una y sólo una de las siguientes alternativas:*

- a) $Ax < 0$, $Bx \leq 0$, $Cx = 0$, tiene solución x .
- b) $A^T y_1 + B^T y_2 + C^T y_3 = 0$ tiene solución y_1, y_2, y_3 , con $y_1 > 0$, $y_2 \geq 0$.

Teorema 2.6.2. *(Teorema de la alternativa de Tucker) Sean las matrices A , B y C , con A no nula. Entonces se cumple una y sólo una de las siguientes alternativas:*

- a) $Ax \leq 0$, $Bx \leq 0$, $Cx = 0$, tiene solución x .
- b) $A^T y_1 + B^T y_2 + C^T y_3 = 0$ tiene solución y_1, y_2, y_3 , con $y_1 > 0$, $y_2 \geq 0$.

Una prueba de los teoremas 2.6.1 y 2.6.2 es dada en Mangasarian [12], pág. 28 y 29.

Teorema 2.6.3. *Una condición necesaria para que el punto \hat{x} sea una solución propiamente eficiente de (2.26) en el sentido de Kuhn-Tucker, es que existan $\hat{\mu} \in \mathbb{R}^p$ y $\hat{\lambda} \in \mathbb{R}^m$ tales que*

$$\sum_{i=1}^p \hat{\mu}_i \nabla f_i(\hat{x}) + \sum_{j=1}^m \hat{\lambda}_j \nabla g_j(\hat{x}) = 0, \quad (2.29a)$$

$$\sum_{j=1}^m \hat{\lambda}_j g_j(\hat{x}) = 0, \quad (2.29b)$$

$$\hat{\mu} > 0, \quad \hat{\lambda} \geq 0. \quad (2.29c)$$

Demostración. Por la definición de eficiencia propia en el sentido de Kuhn-Tucker (definición 2.3.6), no existe un vector $h \in \mathbb{R}^p$ tal que satisfaga las restricciones tangenciales (2.28) y que además cumpla

$$\begin{aligned}\langle \nabla f_i(\hat{x}), h \rangle &\leq 0, \quad \forall i = 1, \dots, p, \\ \langle \nabla f_j(\hat{x}), h \rangle &< 0, \quad \text{para algún } j.\end{aligned}$$

Definimos las matrices

$$A^T = (\nabla f_1(\hat{x}) \quad \cdots \quad \nabla f_p(\hat{x}))$$

$$B^T = (\nabla g_{\alpha_1}(\hat{x}) \quad \cdots \quad \nabla g_{\alpha_r}(\hat{x}) \quad \nabla g_{l+1}(\hat{x}) \quad \cdots \quad \nabla g_m(\hat{x}))$$

es decir, la matriz A^T tiene a los gradientes $\nabla f_i(\hat{x})$ como columnas y la matriz B^T tiene a los gradientes $\nabla g_j(\hat{x})$ como columnas, para aquellos índices j activos. Aplicando el teorema 2.6.2 a las matrices A y B (con $C = 0$), tenemos que la alternativa (a) no puede cumplirse (pues \hat{x} es propiamente eficiente en el sentido de Kuhn-Tucker) y por lo tanto debe cumplirse la alternativa (b), es decir, existen vectores $\hat{\mu} \in \mathbb{R}^p$ y $\hat{\lambda} \in \mathbb{R}^{r+m-l}$ tales que

$$\begin{aligned}A^T \hat{\mu} + B^T \hat{\lambda} &= \sum_{i=1}^p \hat{\mu}_i \nabla f_i(\hat{x}) + \sum_{j=1}^r \hat{\lambda}_{\alpha_j} \nabla g_{\alpha_j}(\hat{x}) + \sum_{k=l+1}^m \hat{\lambda}_k \nabla g_k(\hat{x}) = 0, \\ \hat{\mu} &> 0, \quad \hat{\lambda} \geq 0.\end{aligned}$$

Definiendo $\hat{\lambda}_j = 0$ para aquellos índices j inactivos obtenemos $\hat{\lambda} \in \mathbb{R}^m$ tal que se cumple (2.29). ■

Corolario 2.6.1. *Si la condición de regularidad de Kuhn-Tucker (definición 1.2.3(b)) es satisfecha en un punto $\hat{x} \in \mathcal{X}$, entonces una condición necesaria para que \hat{x} sea una solución propiamente eficiente de (2.26) en el sentido de Geoffrion (o equivalentemente en el sentido de Benson) es que se cumplan las condiciones del teorema 2.6.3.*

Demostración. Es inmediato del teorema 2.6.3 y del teorema 2.3.3. ■

Teorema 2.6.4. *Si las funciones objetivos f_i y las funciones de restricción g_j del problema de optimización multiobjetivo (2.26) son convexas, entonces las condiciones del teorema 2.6.3 son también suficientes para que el punto $\hat{x} \in \mathcal{X}$ sea propiamente eficiente.*

Demostración. Dado $\hat{x} \in \mathcal{X}$, supongamos que existen $\hat{\mu} \in \mathbb{R}^p$ y $\hat{\lambda} \in \mathbb{R}^m$ tales que se cumplen las condiciones (2.29). Luego, la función

$$h = \sum_{i=1}^p \hat{\mu}_i f_i + \sum_{j=1}^m \hat{\lambda}_j g_j$$

es convexa. Por (2.29a) tenemos que $\nabla h(\hat{x}) = 0$ y por el teorema 1.2.4 resulta que \hat{x} es una solución del problema de optimización simple

$$\min_{x \in \mathcal{X}} h(x).$$

Puesto que se cumple (2.29b), entonces las restricciones (2.27) junto con (2.29c) implican que \hat{x} es también una solución del problema de optimización simple

$$\min_{x \in \mathcal{X}} \sum_{i=1}^p \hat{\mu}_i f_i(x).$$

Dado que $\hat{\mu} \in (\mathbb{R}_+^p)^\circ$, entonces el teorema 2.4.9 nos dice que $\hat{x} \in \mathcal{X}_{pE}$. ■

Teorema 2.6.5. *Supongamos que se cumple la condición de regularidad de Kuhn-Tucker (definición 1.2.3(b)) en el punto $\hat{x} \in \mathcal{X}$. Entonces una condición necesaria para que \hat{x} sea una solución débilmente eficiente del problema de optimización multiobjetivo (2.26) es que existan $\hat{\mu} \in \mathbb{R}^p$ y $\hat{\lambda} \in \mathbb{R}^m$ tales que*

$$\sum_{i=1}^p \hat{\mu}_i \nabla f_i(\hat{x}) + \sum_{j=1}^m \hat{\lambda}_j \nabla g_j(\hat{x}) = 0, \quad (2.30a)$$

$$\sum_{j=1}^m \hat{\lambda}_j g_j(\hat{x}) = 0, \quad (2.30b)$$

$$\hat{\mu} \geq 0, \quad \hat{\lambda} \geq 0. \quad (2.30c)$$

Demostración. Sea \hat{x} débilmente eficiente. Probaremos que no existe un vector $h \in \mathbb{R}^p$ tal que satisfaga las restricciones tangenciales (2.28) y que además cumpla

$$\langle \nabla f_i(\hat{x}), h \rangle < 0, \quad \forall i = 1, \dots, p. \quad (2.31)$$

En efecto, supongamos lo contrario, es decir, que existe un vector h tal que satisface las restricciones tangenciales (2.28) y también (2.31). Puesto que se cumple la condición de regularidad de Kuhn-Tucker en \hat{x} , entonces existe una curva suave $x(t)$ ($0 \leq t \leq \delta$) contenida en \mathcal{X} tal que $x(0) = \hat{x}$ y $x'(0) = h$. Tomando el desarrollo de Taylor de primer orden de f_i ($i = 1, \dots, p$) alrededor de \hat{x} obtenemos que

$$f_i(x(t)) = f_i(\hat{x}) + t \langle \nabla f_i(\hat{x}), h \rangle + o(t)$$

donde se cumple que $\frac{o(t)}{t} \xrightarrow{t \rightarrow 0} 0$. Por lo tanto, por (2.31) tenemos que $f_i(x(t)) < f_i(\hat{x})$ para t suficientemente pequeño, contradiciendo la eficiencia débil de \hat{x} . Luego, no existe h tal que satisfaga (2.28) y (2.31).

Definamos las matrices

$$A^T = (\nabla f_1(\hat{x}) \quad \cdots \quad \nabla f_p(\hat{x}))$$

$$B^T = (\nabla g_{\alpha_1}(\hat{x}) \quad \cdots \quad \nabla g_{\alpha_r}(\hat{x}) \quad \nabla g_{l+1}(\hat{x}) \quad \cdots \quad \nabla g_m(\hat{x}))$$

Aplicando el teorema 2.6.1 a las matrices A y B (con $C = 0$), de lo anterior se desprende que la alternativa (a) no puede cumplirse y por lo tanto se cumple la alternativa (b), es decir, existen $\hat{\mu} \in \mathbb{R}^p$ y $\hat{\lambda} \in \mathbb{R}^{r+m-l}$ tales que

$$A^T \hat{\mu} + B^T \hat{\lambda} = \sum_{i=1}^p \hat{\mu}_i \nabla f_i(\hat{x}) + \sum_{j=1}^r \hat{\lambda}_{\alpha_j} \nabla g_{\alpha_j}(\hat{x}) + \sum_{k=l+1}^m \hat{\lambda}_k \nabla g_k(\hat{x}) = 0,$$

$$\hat{\mu} \geq 0, \quad \hat{\lambda} \geq 0.$$

Definiendo $\hat{\lambda}_j = 0$ para los índices j inactivos obtenemos $\hat{\lambda} \in \mathbb{R}^m$ tal que se cumple (2.30). ■

Teorema 2.6.6. *Si las funciones objetivos f_i y las funciones de restricción g_j del problema de optimización multiobjetivo (2.26) son convexas, entonces las condiciones del teorema 2.6.5 son también suficientes para que el punto $\hat{x} \in \mathcal{X}$ sea débilmente eficiente.*

Demostración. Dado $\hat{x} \in \mathcal{X}$, supongamos que existen $\hat{\mu} \in \mathbb{R}^p$ y $\hat{\lambda} \in \mathbb{R}^m$ tales que se cumple (2.30). Luego, la función

$$h = \sum_{i=1}^p \hat{\mu}_i f_i + \sum_{j=1}^m \hat{\lambda}_j g_j$$

es convexa. Por (2.30a) tenemos que $\nabla h(\hat{x}) = 0$ y por el teorema 1.2.4 resulta que \hat{x} es una solución del problema de optimización simple

$$\min_{x \in \mathcal{X}} h(x).$$

Puesto que se cumple (2.30b), entonces las restricciones (2.27) junto con (2.30c) implican que \hat{x} es también una solución del problema de optimización simple

$$\min_{x \in \mathcal{X}} \sum_{i=1}^p \hat{\mu}_i f_i(x).$$

Dado que $\hat{\mu} \in \mathbb{R}_+^p \setminus \{0\}$, entonces el teorema 2.4.9 nos dice que $\hat{x} \in \mathcal{X}_{wE}$. ■

Los resultados anteriores para problemas multiobjetivos son los análogos a las condiciones KKT para el caso de un objetivo (ver sección 1.2.2). Nótese que las condiciones necesarias y suficientes probadas anteriormente son para soluciones propiamente y débilmente eficientes. Sin embargo, puesto que $\mathcal{X}_E \subset \mathcal{X}_{wE}$, entonces las condiciones del teorema 2.6.5 son también condiciones necesarias para soluciones eficientes. De manera análoga, puesto que $\mathcal{X}_{pE} \subset \mathcal{X}_E$, entonces las condiciones del teorema 2.6.4 son también condiciones suficientes para soluciones eficientes. Puesto que la diferencia entre las condiciones de optimalidad para soluciones propiamente eficientes y débilmente eficientes es la de considerar $\hat{\mu} > 0$ (teorema 2.6.3) ó $\hat{\mu} \geq 0$ (teorema 2.6.5), entonces no esperamos otros resultados de este tipo para el caso de soluciones eficientes.

Capítulo 3

Método de continuación global

En esta sección presentaremos un método de continuación global desarrollado en *Pereyra* [16] que nos permitirá obtener una aproximación numérica equiespaciada del conjunto de soluciones eficientes en el espacio variable o del conjunto de puntos no dominados en el espacio objetivo para un problema de optimización multiobjetivo convexo irrestricto, donde las funciones objetivos son continuamente diferenciables. En la actualidad existen varios métodos para lograr tal fin, algunos de ellos basados en técnicas de escalarización (sección 2.4.1). La principal dificultad que afrontan tales métodos es que el mapa $\lambda \rightarrow f_\lambda(x)$ que asocia los pesos (normalizados) con las combinaciones convexas de las funciones objetivos puede ser altamente no lineal, y por lo tanto diferentes estrategias deben ser aplicadas para poder obtener un buen muestreo de las soluciones. Un tipo de método muy usado son algoritmos genéticos, los cuales resultan ser muy robustos pero muy costosos. Como veremos, un método de continuación secuencial (también desarrollado en [16]), donde, al igual que en el método de continuación global, se consideran restricciones adicionales que permiten obtener una representación equiespaciada de las soluciones eficientes o de los puntos no dominados, puede ser implementado con un bajo costo. Sin embargo, el método de continuación global ofrece ventajas, respecto de los algoritmos genéticos por su bajo costo de implementación y la representación uniforme de las soluciones, y respecto del método de continuación secuencial por su posible paralelización.

Obtendremos el método de continuación global al considerar un problema bi-objetivo convexo irrestricto, proponiéndose *a posteriori* su implementación al caso multiobjetivo convexo irrestricto con más de dos objetivos.

Consideremos entonces el problema de optimización bi-objetivo irrestricto

$$\min_{x \in \mathbb{R}^n} (f_1(x), f_2(x)), \quad (3.1)$$

donde supondremos que las funciones objetivos f_1 y f_2 son estrictamente convexas y continuamente diferenciables. Sean x_1^* y x_2^* puntos de mínimo de f_1 y f_2 respectivamente. Definimos entonces la *variedad de Pareto* como la curva paramétrica $x^*(\lambda)$ en \mathbb{R}^n que une x_1^* con x_2^* y que satisface la ecuación

$$G(x(\lambda), \lambda) = (1 - \lambda) \nabla f_1(x) + \lambda \nabla f_2(x) = 0, \quad 0 \leq \lambda \leq 1. \quad (3.2)$$

Tenemos que $G : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$. La imagen por (f_1, f_2) de la variedad de Pareto se conoce como el *frente de Pareto*. Nótese que la parte izquierda de la igualdad (3.2) es la derivada de la función $f_\lambda = (1 - \lambda)f_1 + \lambda f_2$, la cual resulta ser estrictamente convexa por ser una combinación convexa de funciones estrictamente convexas (*Rockafellar* [18], pag. 33). Por teorema 1.2.4, $x^*(\lambda)$ es un punto de mínimo de f_λ si y sólo si se cumple la igualdad (3.2). Por lo tanto, del teorema 2.4.9 resulta que $x^*(\lambda)$ ($0 \leq \lambda \leq 1$) es solución de (3.2) si y sólo si es una solución

eficiente del problema (3.1). Puesto que f_λ es estrictamente convexa, el teorema 1.2.3 garantiza que cada solución $x^*(\lambda)$ de la igualdad (3.2) es única.

Vemos entonces que la variedad de Pareto es

$$\mathcal{X}_E = \{x^*(\lambda)\}_{0 \leq \lambda \leq 1}$$

y por lo tanto el frente de Pareto es

$$\mathcal{Y}_N = \{(y_1, y_2) \in \mathbb{R}^2 \mid y_1 = f_1(x^*(\lambda)), y_2 = f_2(x^*(\lambda)), 0 \leq \lambda \leq 1\}.$$

Obsérvese de la ecuación (3.2) y del teorema 1.2.4 que $x_1^* = x^*(0)$ y $x_2^* = x^*(1)$. En la figura 3.1 se muestra un ejemplo de la variedad de Pareto para $n = 2$. De lo expuesto en la sección 2.2.2 tenemos que para cada $0 \leq \lambda \leq 1$, las curvas de nivel $\{f_1(x) = f_1(x^*(\lambda))\}$ y $\{f_2(x) = f_2(x^*(\lambda))\}$ son tangentes en el punto $x^*(\lambda)$.

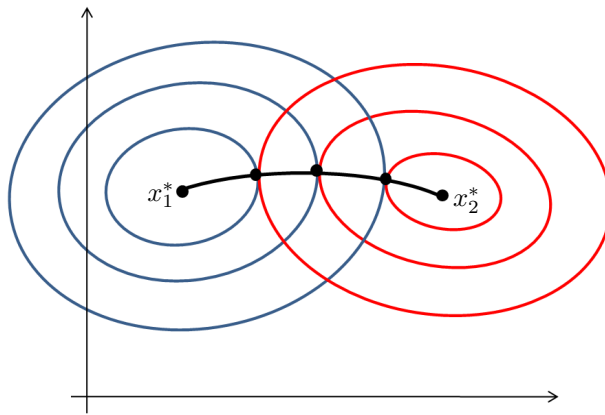


Figura 3.1: Variedad de Pareto en \mathbb{R}^2 .

3.1. Cálculo de la variedad y el frente de Pareto

Al diferenciar $G(x(\lambda), \lambda)$ respecto de λ obtenemos

$$x'(\lambda) = -G_x(x, \lambda)^{-1}G_\lambda(x) \quad \text{sujeto a} \quad x(0) = x_1^*, x(1) = x_2^*, \quad (3.3)$$

donde

$$G_x(x, \lambda) = (1 - \lambda)\nabla^2 f_1(x) + \lambda\nabla^2 f_2(x),$$

$$G_\lambda(x) = \nabla f_2(x) - \nabla f_1(x).$$

Este es un problema de valores de frontera de dos puntos para un conjunto de n ecuaciones diferenciales ordinarias en la variable independiente λ . Aquí suponemos que la matriz $G_x(x, \lambda)$ es no singular a lo largo de la variedad de Pareto.

Con el fin de obtener una representación discreta de la variedad de Pareto podemos resolver (3.3) numéricamente. Una aproximación inicial puede ser dada por el segmento de recta que une los puntos x_1^* y x_2^*

$$x^{(0)}(\lambda_i) = (1 - \lambda_i)x_1^* + \lambda_i x_2^*, \quad \lambda_i = i \cdot \delta\lambda, \quad i = 0, \dots, l + 1, \quad (3.4)$$

donde el intervalo $[0, 1]$ ha sido dividido en $l + 1$ subintervalos de longitud $\delta\lambda$.

Desafortunadamente, para resolver (3.3) aplicando el método de Newton necesitamos derivadas de tercer orden, lo cual resulta poco práctico. Por lo tanto resolveremos (3.2) siguiendo una tendencia secuencial. Aplicando el método de Newton (*Polak* [17], pag. 71) a la ecuación (3.2) obtenemos

$$x^{(j+1)}(\lambda_i) = x^{(j)}(\lambda_i) - G_x^{-1}(x^{(j)}(\lambda_i), \lambda_i) G(x^{(j)}(\lambda_i), \lambda_i), \quad i = 1, \dots, l, \quad j = 0, 1, 2, \dots \quad (3.5)$$

Si $x^*(\lambda_i)$ es la solución aproximada en la i -ésima etapa, entonces usamos $x^{(0)}(\lambda_{i+1}) = x^*(\lambda_i)$ como valor inicial para la etapa siguiente, lo que convierte a (3.5) en un método esencialmente secuencial. Los índices $i = 0$ e $i = l + 1$ han sido descartados en (3.5) debido a que x_1^* y x_2^* son soluciones de (3.2) y por lo tanto $x_1^* = x^{(j)}(\lambda_0)$ y $x_2^* = x^{(j)}(\lambda_{l+1})$ para todo j .

Aunque el planteo del problema (3.3) no resulta muy práctico, aplicando ideas de soluciones numéricas a tales problemas por diferencias finitas es posible desarrollar un método de continuación novedoso para problemas convexos siguiendo una tendencia *global* en vez de *secuencial*, es decir, en lugar de considerar el método de continuación secuencial (3.5), nuestro objetivo será deformar la malla completa de soluciones iniciales (3.4) hasta que ésta aproxime la variedad de Pareto con la precisión deseada (ver figura 3.2).

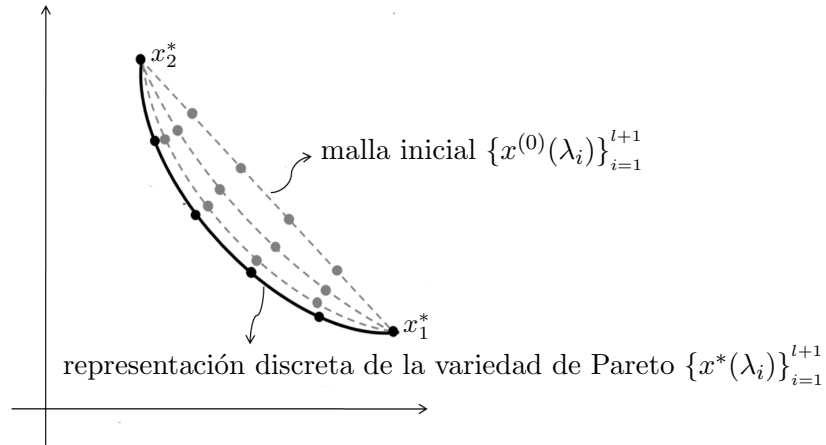


Figura 3.2: Aproximación a la variedad de Pareto por el método de continuación global.

Una característica deseable es obtener una representación equiespaciada de la variedad de Pareto. Aún comenzando con una malla inicial como en (3.4), con puntos equidistantes, eso no garantiza la obtención de un espaciado uniforme en la variedad. Con este fin consideramos las restricciones adicionales:

$$L_i(x(\lambda_1), \dots, x(\lambda_l), \lambda_1, \dots, \lambda_l) = \|x(\lambda_i) - x(\lambda_{i-1})\|^2 - \frac{P}{l+1} = 0, \quad i = 1, \dots, l+1, \quad (3.6)$$

donde

$$P = \sum_{j=1}^{l+1} \|x(\lambda_j) - x(\lambda_{j-1})\|^2.$$

Tenemos que $L_i : \mathbb{R}^{(n+1)l} \rightarrow \mathbb{R}$. Como ya mencionamos, los puntos $x_1^* = x^*(\lambda_0)$ y $x_2^* = x^*(\lambda_{l+1})$ son soluciones de (3.2) y por lo tanto permanecen fijos. Es inmediato ver que la ecuación (3.6) se cumple si y sólo si $\|x(\lambda_i) - x(\lambda_{i-1})\| = \sqrt{\frac{P}{l+1}} = cte$ para todo $i = 1, \dots, l+1$, es decir, si y sólo si las soluciones están igualmente espaciadas.

Definiendo $\bar{\lambda} = (\lambda_1, \dots, \lambda_l) \in \mathbb{R}^l$ y usando la abreviación $x(\bar{\lambda}) = (x(\lambda_1), \dots, x(\lambda_l)) \in \mathbb{R}^{nl}$, introducimos el operador $\mathcal{H} : \mathbb{R}^{(n+1)l} \rightarrow \mathbb{R}^{(n+1)l}$ definido por

$$\mathcal{H}(x(\bar{\lambda}), \bar{\lambda}) = \mathcal{H}(x(\lambda_1), \dots, x(\lambda_l), \lambda_1, \dots, \lambda_l) = \begin{pmatrix} G(x(\lambda_1), \lambda_1) \\ \vdots \\ G(x(\lambda_l), \lambda_l) \\ L_1(x(\bar{\lambda}), \bar{\lambda}) \\ \vdots \\ L_l(x(\bar{\lambda}), \bar{\lambda}) \end{pmatrix} = \begin{pmatrix} G_1 \\ \vdots \\ G_l \\ L_1 \\ \vdots \\ L_l \end{pmatrix} \quad (3.7)$$

En (3.7) hemos usado la abreviación $G(x(\lambda_i), \lambda_i) = G_i$. Si $x(\lambda_i) = (x_1(\lambda_i), \dots, x_n(\lambda_i))$, entonces la matriz jacobiana del operador \mathcal{H} es una matriz de dimensión $(n+1)l \times (n+1)l$ que está dada por

$$H(x(\bar{\lambda}), \bar{\lambda}) = \left(\begin{array}{ccc|ccc} G_{1,x(\lambda_1)} & \cdots & 0 & G_{1,\lambda_1} & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & G_{l,x(\lambda_l)} & 0 & \cdots & G_{l,\lambda_l} \\ \hline L_{1,x(\lambda_1)} & \cdots & L_{1,x(\lambda_l)} & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ L_{l,x(\lambda_1)} & \cdots & L_{l,x(\lambda_l)} & 0 & \cdots & 0 \end{array} \right)$$

donde hemos usado las abreviaciones

$$G_{i,x(\lambda_i)} = \left(\frac{\partial G_i}{\partial x_1(\lambda_i)} \cdots \frac{\partial G_i}{\partial x_n(\lambda_i)} \right) = (1 - \lambda_i) \nabla^2 f_1(x(\lambda_i)) + \lambda_i \nabla^2 f_2(x(\lambda_i)) \quad (3.8)$$

$$G_{i,\lambda_i} = \nabla f_2(x(\lambda_i)) - \nabla f_1(x(\lambda_i)) \quad (\text{derivada total respecto de } \lambda_i) \quad (3.9)$$

$$L_{i,x(\lambda_j)} = \left(\frac{\partial L_i}{\partial x_1(\lambda_j)} \cdots \frac{\partial L_i}{\partial x_n(\lambda_j)} \right) = \begin{cases} 2[x(\lambda_j) - x(\lambda_{j-1})] + \eta_j, & \text{si } j = i \\ -2[x(\lambda_{j+1}) - x(\lambda_j)] + \eta_j, & \text{si } j = i - 1 \\ \eta_j, & \text{para cualquier otro } j \end{cases} \quad (3.10)$$

con $\eta_j = -\frac{2}{l+1}[2x(\lambda_j) - x(\lambda_{j-1}) - x(\lambda_{j+1})]$. La matriz H está compuesta por lo tanto de un bloque superior izquierdo de dimensión $nl \times nl$, un bloque superior derecho de dimensión $nl \times l$ y un bloque inferior izquierdo de dimensión $l \times nl$. Denotando estos bloques por $G_{x(\bar{\lambda})}$, $G_{\bar{\lambda}}$ y C respectivamente podemos escribir abreviadamente

$$H = \begin{pmatrix} G_{x(\bar{\lambda})} & G_{\bar{\lambda}} \\ C & 0 \end{pmatrix} \quad (3.11)$$

Nótese que $x(\lambda_1), \dots, x(\lambda_l)$ satisfacen (3.2) y (3.6), es decir, son puntos equiespaciados sobre la variedad de Pareto, si y sólo si $\mathcal{H}(x(\bar{\lambda}), \bar{\lambda}) = 0$. Aplicando el método de Newton a esta igualdad y resolviendo para $(x(\bar{\lambda}), \bar{\lambda})$ obtenemos

$$\begin{pmatrix} x^{(j+1)}(\lambda_1) \\ \vdots \\ x^{(j+1)}(\lambda_l) \\ \lambda_1^{(j+1)} \\ \vdots \\ \lambda_l^{(j+1)} \end{pmatrix} = \begin{pmatrix} x^{(j)}(\lambda_1) \\ \vdots \\ x^{(j)}(\lambda_l) \\ \lambda_1^{(j)} \\ \vdots \\ \lambda_l^{(j)} \end{pmatrix} - H^{-1}(x(\bar{\lambda}), \bar{\lambda}) \mathcal{H}(x(\bar{\lambda}), \bar{\lambda})$$

o equivalentemente

$$H(x(\bar{\lambda}), \bar{\lambda}) \begin{pmatrix} \Delta x(\bar{\lambda}) \\ \Delta \bar{\lambda} \end{pmatrix} = -\mathcal{H}(x(\bar{\lambda}), \bar{\lambda}) = \begin{pmatrix} -G_1 \\ \vdots \\ -G_l \\ -L_1 \\ \vdots \\ -L_l \end{pmatrix}$$

donde

$$\Delta x(\bar{\lambda}) = \begin{pmatrix} x^{(j+1)}(\lambda_1) - x^{(j)}(\lambda_1) \\ \vdots \\ x^{(j+1)}(\lambda_l) - x^{(j)}(\lambda_l) \end{pmatrix} = \begin{pmatrix} \Delta x(\lambda_1) \\ \vdots \\ \Delta x(\lambda_l) \end{pmatrix}$$

$$\Delta \bar{\lambda} = \begin{pmatrix} \lambda_1^{(j+1)} - \lambda_1^{(j)} \\ \vdots \\ \lambda_l^{(j+1)} - \lambda_l^{(j)} \end{pmatrix} = \begin{pmatrix} \Delta \lambda_1 \\ \vdots \\ \Delta \lambda_l \end{pmatrix}$$

Al aplicar el método de Newton suponemos que la matriz H es no singular. Usando (3.11), podemos expresar el sistema anterior en la forma compacta

$$\begin{pmatrix} G_{x(\bar{\lambda})} & G_{\bar{\lambda}} \\ C & 0 \end{pmatrix} \begin{pmatrix} \Delta x(\bar{\lambda}) \\ \Delta \bar{\lambda} \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} \quad (3.12)$$

donde $a = (-G_1 \dots -G_l)^T$ y $b = (-L_1 \dots -L_l)^T$.

El siguiente paso será resolver el sistema en bloque (3.12) por eliminación Gaussiana. Para esto multiplicamos la primera fila por la matriz $C G_{x(\bar{\lambda})}^{-1}$ y modificamos la segunda fila por la diferencia con la primera fila modificada, obteniendo el sistema

$$\begin{pmatrix} G_{x(\bar{\lambda})} & G_{\bar{\lambda}} \\ 0 & D \end{pmatrix} \begin{pmatrix} \Delta x(\bar{\lambda}) \\ \Delta \bar{\lambda} \end{pmatrix} = \begin{pmatrix} a \\ b^* \end{pmatrix}$$

donde $D = -C G_{x(\bar{\lambda})}^{-1} G_{\bar{\lambda}}$ y $b^* = b - C G_{x(\bar{\lambda})}^{-1} a$. Haciendo sustitución hacia atrás obtenemos

$$\Delta \bar{\lambda} = D^{-1} b^*. \quad (3.13)$$

Ahora calculamos $\Delta x(\bar{\lambda})$:

$$G_{x(\bar{\lambda})} \Delta x(\bar{\lambda}) + G_{\bar{\lambda}} \Delta \bar{\lambda} = a \implies \Delta x(\bar{\lambda}) = G_{x(\bar{\lambda})}^{-1} (a - G_{\bar{\lambda}} \Delta \bar{\lambda}) = G_{x(\bar{\lambda})}^{-1} a - G_{x(\bar{\lambda})}^{-1} G_{\bar{\lambda}} \Delta \bar{\lambda}$$

o en notación matricial

$$\begin{pmatrix} \Delta x(\lambda_1) \\ \vdots \\ \Delta x(\lambda_l) \end{pmatrix} = \begin{pmatrix} G_{1,x(\lambda_1)}^{-1} & & 0 \\ & \ddots & \\ 0 & & G_{l,x(\lambda_l)}^{-1} \end{pmatrix} \begin{pmatrix} -G_1 \\ \vdots \\ -G_l \end{pmatrix} - \begin{pmatrix} G_{1,x(\lambda_1)}^{-1} & & 0 \\ & \ddots & \\ 0 & & G_{l,x(\lambda_l)}^{-1} \end{pmatrix} \begin{pmatrix} G_{1,\lambda_1} & & 0 \\ & \ddots & \\ 0 & & G_{l,\lambda_l} \end{pmatrix} \begin{pmatrix} \Delta \lambda_1 \\ \vdots \\ \Delta \lambda_l \end{pmatrix}$$

Por lo tanto

$$\Delta x(\lambda_i) = -G_{i,x(\lambda_i)}^{-1} G_i - G_{i,x(\lambda_i)}^{-1} G_{i,\lambda_i} \Delta \lambda_i, \quad i = 1, \dots, l. \quad (3.14)$$

Para resolver (3.14), factorizamos $G_{i,x(\lambda_i)}$ y resolvemos los sistemas

$$\begin{cases} G_{i,x(\lambda_i)} c_i = -G_i, \\ G_{i,x(\lambda_i)} d_i = G_{i,\lambda_i}, \end{cases} \quad i = 1, \dots, l \quad (3.15)$$

obteniendo

$$\Delta x(\lambda_i) = c_i - d_i \Delta \lambda_i, \quad i = 1, \dots, l. \quad (3.16)$$

Finalmente, con estas correcciones, podemos resolver para $(x(\bar{\lambda}), \bar{\lambda})$ iterativamente

$$\begin{cases} x^{(j+1)}(\lambda_i) = x^{(j)}(\lambda_i) + \Delta x(\lambda_i), \\ \lambda_i^{(j+1)} = \lambda_i^{(j)} + \Delta \lambda_i, \end{cases} \quad i = 1, \dots, l. \quad (3.17)$$

El sistema (3.17) resume el método de continuación global. Es posible incrementar la región de convergencia al introducir un coeficiente de Armijo (*Polak* [17], pag. 82). Puesto que las ecuaciones (3.16) son independientes, el sistema (3.17) puede ser resuelto en paralelo.

A continuación resumimos las etapas del método:

Dado el problema de optimización multiobjetivo (3.1), consideramos la condición de primer orden (3.2) y las restricciones (3.6). Comenzando con la malla inicial (3.4):

- 1 Calculamos G_i y L_i de (3.2) y (3.6) respectivamente, obteniendo los vectores a y b
- 2 Calculamos $G_{i,x(\lambda_i)}$, G_{i,λ_i} y $L_{i,x(\lambda_i)}$ de (3.8), (3.9) y (3.10) respectivamente, obteniendo las matrices $G_{x(\bar{\lambda})}$, $G_{\bar{\lambda}}$ y C
- 3 De (3.13) resolvemos $D\Delta\bar{\lambda} = b^*$ por eliminación Gaussiana y obtenemos $\Delta\bar{\lambda}$
- 4 Resolvemos los sistemas (3.15)
- 5 Calculamos $\Delta x(\lambda_i)$ de (3.16), obteniendo $\Delta x(\bar{\lambda})$
- 6 Calculamos el paso iterativo siguiente a través de (3.17)
- 7 Repetimos los pasos 1 a 6 hasta que se cumpla un criterio de detención respecto del error $(x^{(j+1)}(\lambda_i) - x^{(j)}(\lambda_i))$

Este algoritmo permite obtener una representación discreta uniforme de la variedad de Pareto. Sin embargo, muchas veces es más conveniente obtener el frente de Pareto. Dado que las funciones objetivos f_1 y f_2 no son lineales, una representación equiespaciada en el espacio variable no garantiza un espaciado uniforme en el espacio objetivo. En este caso reemplazamos las restricciones (3.6) por las siguientes

$$L_i(x(\bar{\lambda}), \bar{\lambda}) = \|F(x(\lambda_i)) - F(x(\lambda_{i-1}))\|^2 - \frac{P}{l+1} = 0, \quad i = 1, \dots, l, \quad (3.18)$$

donde

$$P = \sum_{j=1}^{l+1} \|F(x(\lambda_j)) - F(x(\lambda_{j-1}))\|^2,$$

$$F = (f_1, f_2).$$

Naturalmente, esto requiere un nuevo cálculo del bloque C de la jacobiana H usando

$$L_{i,x(\lambda_j)} = \begin{cases} 2 \sum_{k=1}^2 [f_k(x(\lambda_j)) - f_k(x(\lambda_{j-1}))] \nabla f_k(x(\lambda_j)) + \eta_j, & \text{si } j = i \\ -2 \sum_{k=1}^2 [f_k(x(\lambda_{j+1})) - f_k(x(\lambda_j))] \nabla f_k(x(\lambda_j)) + \eta_j, & \text{si } j = i - 1 \\ \eta_j, & \text{para cualquier otro } j \end{cases}$$

donde

$$\eta_j = \eta_{j_1} + \eta_{j_2},$$

$$\eta_{j_1} = -\frac{2}{l+1} [2f_1(x(\lambda_j)) - f_1(x(\lambda_{j-1})) - f_1(x(\lambda_{j+1}))] \nabla f_1(x(\lambda_j)),$$

$$\eta_{j_2} = [2f_2(x(\lambda_j)) - f_2(x(\lambda_{j-1})) - f_2(x(\lambda_{j+1}))] \nabla f_2(x(\lambda_j)).$$

También tiene que ser redefinido el vector $b = (-L_1 \dots - L_l)^T$ en (3.12). Finalmente, resolvemos como antes para $(x(\bar{\lambda}), \bar{\lambda})$ iterativamente a través de (3.17) y tomamos su imagen por F , obteniendo de esta manera una representación discreta uniforme del frente de Pareto.

Como una observación final, es importante aclarar que es posible considerar el método (3.5) junto con restricciones en el espacio objetivo (ver sección 3.3) para obtener de esta manera un método de continuación secuencial que permita calcular una representación equiespaciada del frente de Pareto, aunque es claro que tal método no cuenta con la ventaja de la paralelización del método de continuación global. Sin embargo, en ocasiones la malla (3.4) puede ser una mala aproximación inicial (sobre todo cuando el frente de Pareto resulta ser muy curvado) lo que provoca que el método de continuación global falle. En estos casos el método de continuación secuencial mencionado puede ser de utilidad al proporcionar una aproximación inicial más exacta que (3.4), para luego poder aplicar el método global con éxito.

3.2. Generalización a problemas con más de dos objetivos

El método de continuación global desarrollado anteriormente para el caso bi-objetivo convexo puede ser extendido a problemas con más de dos objetivos, al precio de incluir más restricciones, con el fin de calcular una representación equiespaciada de la variedad de Pareto, obteniendo en el caso $n \geq p$ una variedad de dimensión $p - 1$ para un problema con p funciones

objetivos. Por ejemplo, si consideramos un problema con tres objetivos, debemos comenzar con una malla inicial $(l + 1) \times (l + 1)$ y reemplazar las ecuaciones (3.2) y (3.6) por

$$G(x(\lambda_1, \lambda_2), \lambda_1, \lambda_2) = \lambda_1 \nabla f_1(x) + \lambda_2 \nabla f_2(x) + \lambda_3 \nabla f_3(x) = 0, \quad \lambda_1 + \lambda_2 + \lambda_3 = 1$$

$$\|x(\lambda_i, \lambda_j) - x(\lambda_{i-1}, \lambda_j)\|^2 - \frac{1}{(l+1)^2} \sum_{s=1}^{l+1} \sum_{t=1}^{l+1} P_{s,t} = 0$$

$$\|x(\lambda_i, \lambda_j) - x(\lambda_i, \lambda_{j-1})\|^2 - \frac{1}{(l+1)^2} \sum_{s=1}^{l+1} \sum_{t=1}^{l+1} P_{s,t} = 0$$

donde

$$P_{s,t} = \|x(\lambda_s, \lambda_t) - x(\lambda_{s-1}, \lambda_t)\|^2 + \|x(\lambda_s, \lambda_t) - x(\lambda_s, \lambda_{t-1})\|^2$$

con las distancias a puntos fuera de la malla igualados a cero. Por otra parte, si lo que queremos es obtener una aproximación equiespaciada del frente de Pareto, reemplazamos las ecuaciones de restricción anteriores por las correspondientes para este caso (ver ecuación (3.18)).

3.3. Problemas de testeo

La siguiente lista muestra los métodos desarrollados anteriormente:

- 1) Método de continuación global (3.17) con restricciones de espaciado uniforme (3.6).
- 2) Método de continuación global (3.17) con restricciones de espaciado uniforme (3.18).
- 3) Método de Newton simple (3.5).
- 4) Método de continuación secuencial.

El método de continuación secuencial 4 combina el método de Newton (3.5) con restricciones de espaciado uniforme en el frente de Pareto. Tales restricciones se obtienen al considerar

$$\|F(x(\lambda_i)) - F(x(\lambda_{i-1}))\| = d = cte, \quad i = 1, \dots, l + 1,$$

donde la constante d es igual a la longitud del frente de Pareto dividido por el número de subdivisiones (en nuestro caso $l + 1$). Puesto que no conocemos la longitud del frente *a priori*, estimamos la misma como un valor algo mayor que la distancia $\|F(x_2^*) - F(x_1^*)\|$. A lo largo de esta sección nos referiremos a los métodos enumerados anteriormente como métodos 1, 2, 3 y 4 respectivamente.

Con el fin de analizar la performance de tales métodos, en *Pereyra* [16] (pag. 1941 a 1946) se implementan los mismos en algunas funciones de testeo extraídas principalmente de *Deb et al.* [6], tabla 1 (pag. 187).

Ejemplo 1. Considerar el problema (SCH) de [6]:

$$f_1(x) = x^2, \quad f_2(x) = (x - 2)^2.$$

Este es un problema de optimización multiobjetivo convexo muy simple con $n = 1$ y $p = 2$. En este caso $x_1^* = 0$, $x_2^* = 2$ y por lo tanto es claro que la variedad de Pareto es el intervalo $[0, 2]$. La ecuación para la variedad es obtenida directamente de (3.2):

$$2(1 - \lambda)x(\lambda) + 2\lambda(x(\lambda) - 2) = 0 \implies x(\lambda) = 2\lambda, \quad 0 \leq \lambda \leq 1$$

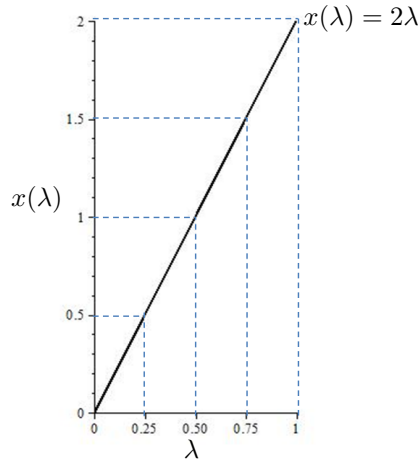


Figura 3.3: Gráfico $x(\lambda)$ para el ejemplo 1.

Como podemos observar en la figura 3.3, una partición uniforme del intervalo $[0, 1]$ permite obtener directamente un espaciado uniforme en la variedad de Pareto. Sin embargo, esto no garantiza el espaciado uniforme en el frente de Pareto.

En la figura 3.4 se muestra la representación discreta del frente de Pareto utilizando 150 puntos, obtenida por el método 3. En este caso los puntos no están equiespaciados, aunque el muestreo obtenido es muy satisfactorio. En la figura 3.5(a) se muestra el frente obtenido con el método 2, usando tan sólo 30 puntos. Como puede verse en 3.5(b), estos puntos están uniformemente espaciados por construcción.

Es interesante comparar estos resultados con aquellos obtenidos por medio de algoritmos genéticos. En la figura 3.6 se muestra el frente de Pareto del ejemplo 1 obtenido con el método NSGA-II descrito en *Deb et al.* [6] usando 100 puntos. Como es posible observar, este método no proporciona un equiespaciado de las soluciones, además de resultar muy costoso (este ejemplo requirió de aproximadamente 25000 evaluaciones de funciones). El mismo es comparado con otro método (PAES) para mostrar el mejoramiento en la capacidad de muestreo. \square

Ejemplo 2. Considerar el problema (FON) de [6]:

$$f_1(x) = 1 - \exp\left(-\sum_{i=1}^3 \left(x_i - \frac{1}{\sqrt{3}}\right)^2\right), \quad f_2(x) = 1 - \exp\left(-\sum_{i=1}^3 \left(x_i + \frac{1}{\sqrt{3}}\right)^2\right).$$

Este es un problema de optimización multiobjetivo con $n = 3$ y $p = 2$. A pesar de ser un problema no convexo, el método 4 es capaz de obtener un muestreo de puntos aún en la parte no convexa del frente de Pareto, como puede observarse en la figura 3.7. Por otra parte, el método 3 sólo cubre la parte convexa inicial (observar la diferencia de escala entre los dos gráficos). \square

Ejemplo 3. Considerar el problema (POL) de [6]:

$$f_1(x) = 1 + (A_1 - B_1)^2 + (A_2 - B_2)^2, \quad f_2(x) = (x_1 + 3)^2 + (x_2 + 1)^2,$$

donde

$$\begin{aligned} A_1 &= 0.5 \sin(1) - 2 \cos(1) + 2 \sin(2) - 1.5 \cos(2) \\ A_2 &= 1.5 \sin(1) - \cos(1) + 2 \sin(2) - 0.5 \cos(2) \\ B_1 &= 0.5 \sin(x_1) - 2 \cos(x_1) + \sin(x_2) - 1.5 \cos(x_2) \\ B_2 &= 1.5 \sin(x_1) - \cos(x_1) + 2 \sin(x_2) - 0.5 \cos(x_2) \end{aligned}$$

Este es un problema de optimización multiobjetivo no convexo con $n = 2$, $p = 2$ y con frente de Pareto separable (ver sección 2.5). Como podemos ver de la figura 3.8, los métodos 3 y 4 muestran sólo un tramo conectado del conjunto. Nuevamente, el método 4 realiza un mejor muestreo de este tramo que el método 3. \square

Ejemplo 4. Considerar el problema de optimización multiobjetivo:

$$f_1(x) = (x_2 - 1)^2 + (x_3 - 1)^2 + (x_1 - 1)^4, \quad f_2(x) = (x_1 + 1)^2 + (x_3 + 1)^2 + (x_2 + 1)^4.$$

Este es un problema convexo con $n = 3$ y $p = 2$. Sin embargo, este problema tiene un frente de Pareto con una gran curvatura, ocasionando que el método 2 falle debido a que (3.4) resulta ser una mala aproximación inicial. Para sortear este inconveniente, se obtuvo una mejor aproximación inicial aplicando el método 4, para luego continuar con el método 2 y obtener así el frente mostrado en la figura 3.9. \square

A continuación damos el último ejemplo de implementación.

Ejemplo 5. Considerar el problema (ZDT4) de [6] (con algunas modificaciones):

$$f_1(x) = x_1^2, \quad f_2(x) = g(x) \left(1 - \sqrt{\frac{x_1}{g(x)}} \right),$$

donde

$$g(x) = 1 + 10(n - 1) + \sum_{i=2}^n (x_i^2 - 10 \cos(4\pi x_i)), \quad n = 10.$$

Este es un problema de optimización multiobjetivo convexo con $n = 10$ y $p = 2$. Como podemos ver en la figura 3.10, el método 3 da un pobre muestreo del frente de Pareto, mientras que el método 4 proporciona una representación excelente con sólo 30 puntos. Tal como se menciona en [16] (tabla 2), no hay un impacto significativo en cuanto a tiempos de ejecución al incrementar la dimensionalidad del problema. \square

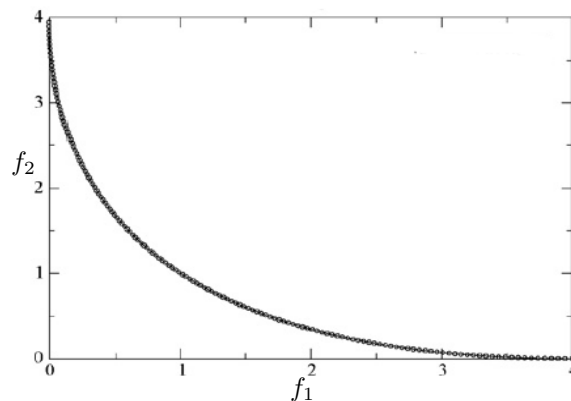


Figura 3.4: Frente de Pareto de ejemplo 1 obtenido por el método 3.

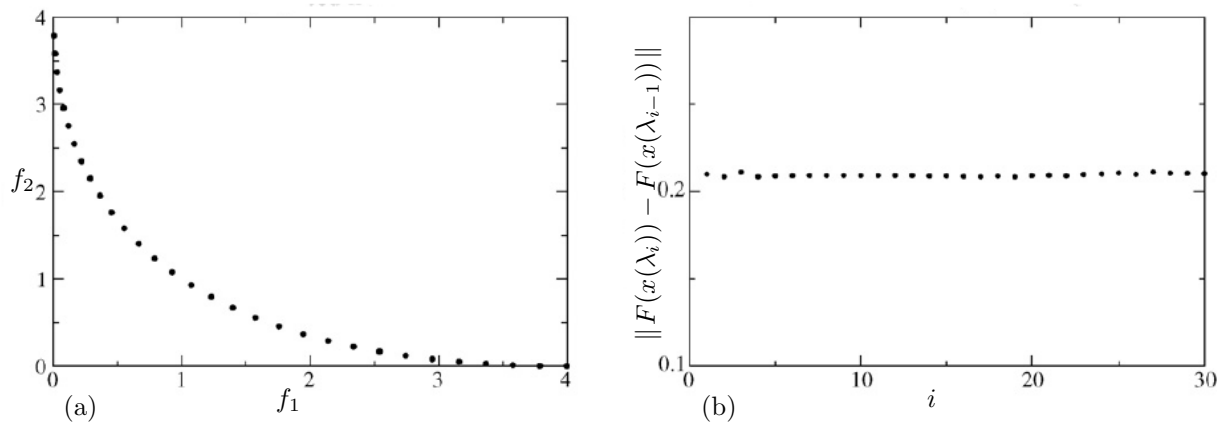


Figura 3.5: (a) Frente de Pareto de ejemplo 1 obtenido por metodo 2. (b) Distancia entre puntos.

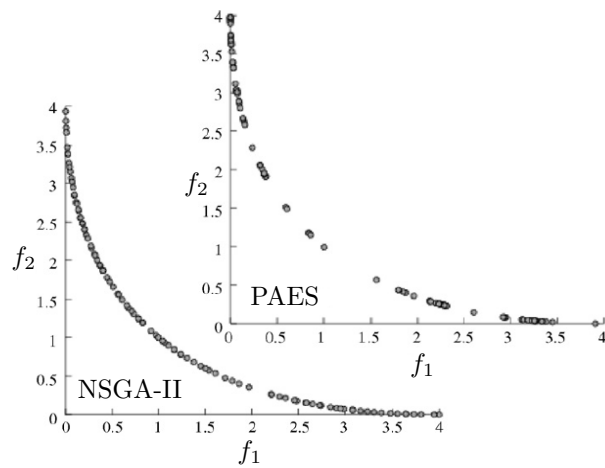


Figura 3.6: Frente de Pareto de ejemplo 1 obtenido por algoritmos genéticos.

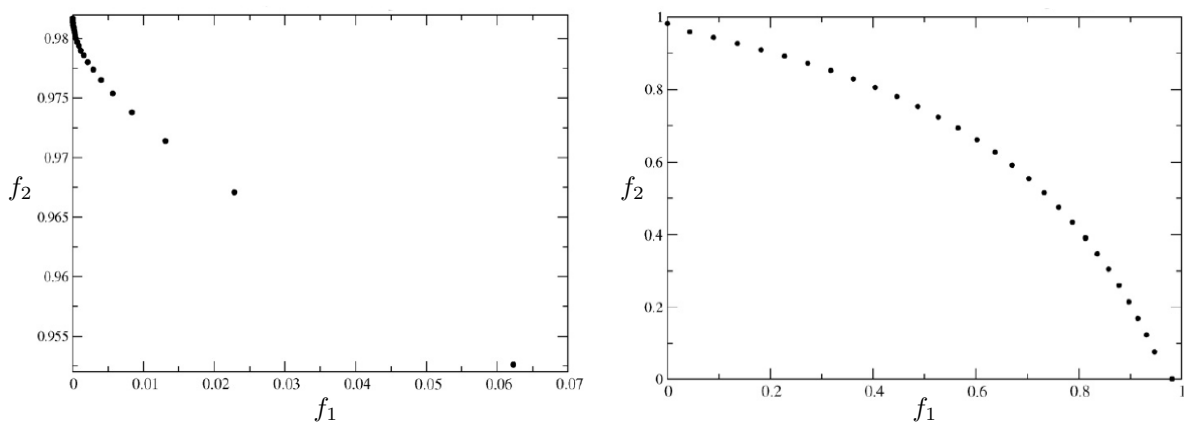


Figura 3.7: Frente de Pareto de ejemplo 2 obtenido por el método 3 (izquierda) y el método 4 (derecha).

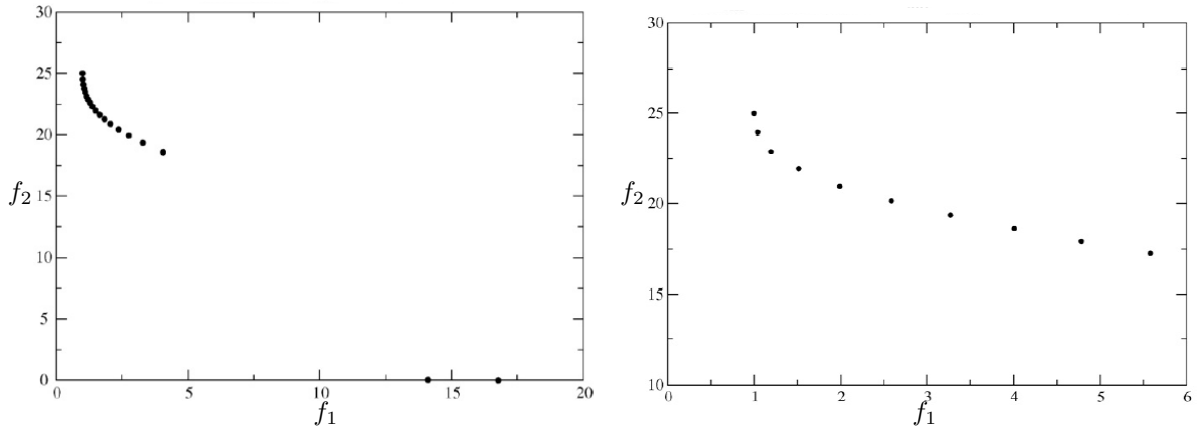


Figura 3.8: Tramo conectado del frente de Pareto de ejemplo 3 obtenido por el método 3 (izquierda) y el método 4 (derecha).

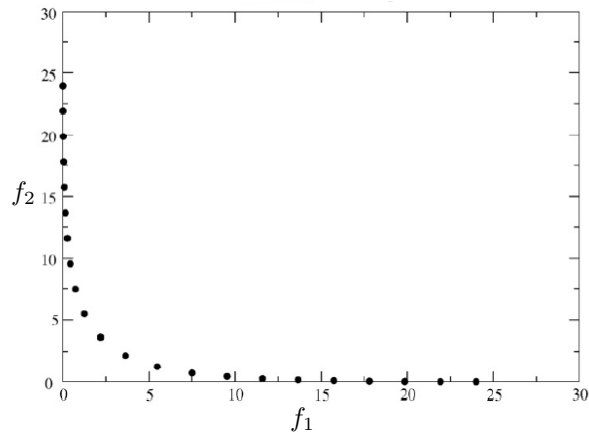


Figura 3.9: Frente de Pareto de ejemplo 4 obtenido por el método 2 con aproximación inicial dada por el método 4.

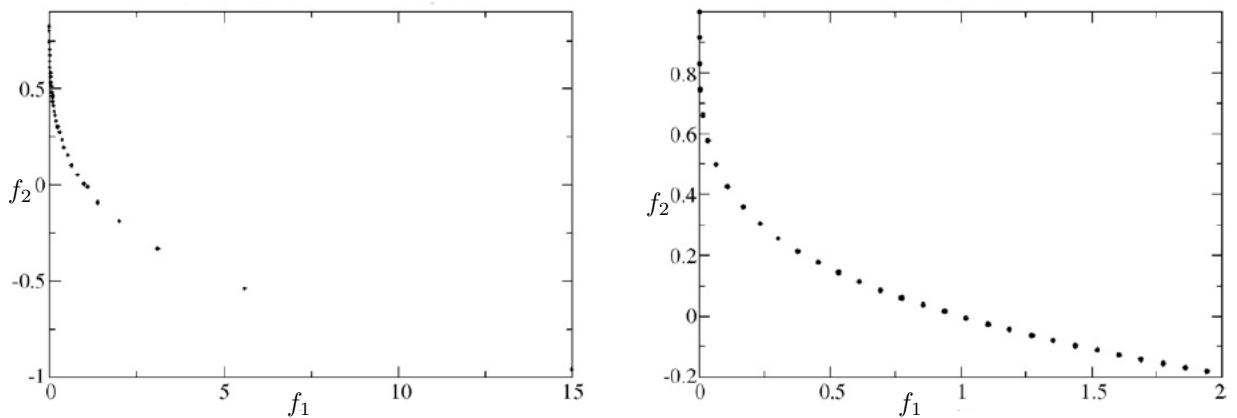


Figura 3.10: Frente de Pareto de ejemplo 5 obtenido por el método 3 (izquierda) y el método 4 (derecha).

En la tabla siguiente se muestra a modo de resumen los métodos considerados en *Pereyra* [16] y los ejemplos en donde han sido aplicados.

Problema	Dimensiones	Características	Método aplicado
Ejemplo 1 (SCH)	$n = 1$, $p = 2$	convexo	1, 2, 3 y 4
Ejemplo 2 (FON)	$n = 3$, $p = 2$	no convexo	3 y 4
Ejemplo 3 (POL)	$n = 2$, $p = 2$	no convexo, frente separable	3 y 4
Ejemplo 4	$n = 3$, $p = 2$	convexo, alta curvatura del frente	1, 2 y 4
Ejemplo 5 (ZDT4)	$n = 10$, $p = 2$	convexo, alta dimensionalidad	3 y 4

En la tabla 2 de [16] (pag. 1945) se muestran los tiempos de ejecución (en ms) sobre una máquina Intel dual Xeon 3.2 GH con SO Linux.

3.4. Conclusiones

En el capítulo 3 se ha mostrado en detalle un método de continuación global desarrollado en [16], además de un método de continuación secuencial con restricciones de equiespaciado, los cuales permiten obtener una aproximación numérica de la variedad o el frente de Pareto para un problema multiobjetivo irrestricto. Para tal fin se presentó en los dos primeros capítulos una revisión de conceptos de optimización, comenzando en primer lugar con optimización simple (capítulo 1) y luego siguiendo con optimización multiobjetivo (capítulo 2), cuyo concepto se introdujo al dar ejemplos de problemas con dos o más objetivos en conflicto. También se presentaron los resultados teóricos sobre el tema necesarios para derivar tales métodos.

Estos métodos de continuación fueron obtenidos al aplicar el método de Newton a ciertas condiciones de optimalidad de primer orden y al considerar condiciones adicionales de equiespaciado. Los mismos han demostrado tener una buena performance al aplicarlos a algunas funciones de testeo, donde se mostró también su capacidad de muestreo uniforme de los puntos óptimos. En el caso del método global, éste fué obtenido en primer lugar al considerar un problema bi-objetivo con funciones objetivos estrictamente convexas, donde obtuvimos el conjunto de soluciones eficientes como una variedad de dimensión uno (variedad de Pareto) y luego se propuso su implementación a problemas multiobjetivos generales, donde en este caso la variedad de Pareto resulta ser de dimensión $p - 1$ para p funciones objetivos. No obstante, los métodos de continuación propuestos pueden ser aplicados con éxito a algunos problemas donde las funciones objetivos pueden no ser estrictamente convexas. Sin embargo, deben tenerse en cuenta algunas consideraciones:

1. En el caso de funciones convexas no estrictas, éstas pueden tener más de un punto de mínimo y por lo tanto puede haber más de una curva paramétrica uniendo los diferentes puntos óptimos. En estas circunstancias el conjunto de soluciones débilmente eficientes \mathcal{X}_{wE} estará formado por la unión de estas curvas. Puesto que en este caso el conjunto eficiente \mathcal{Y} en el espacio objetivo resulta ser \mathbb{R}_+^p -convexo, entonces el conjunto no dominado \mathcal{Y}_N resulta conectado si el conjunto $\mathcal{Y} + \mathbb{R}_+^p$ es \mathbb{R}_+^p -compacto (ver corolario 2.5.2), lo cual facilita la obtención del frente de Pareto por los métodos de continuación.
2. En el caso de funciones no convexas, algunos de los inconvenientes para aplicar los métodos de continuación son:
 - *Conjunto no dominado \mathcal{Y}_N cóncavo.* Como comentamos al final de la sección 2.4.1, la obtención del frente de Pareto usando combinaciones convexas de los objetivos

(escalarización) puede fallar en el caso de funciones objetivos no convexas, como se muestra en el ejemplo del frente cóncavo de la figura 2.23. Sin embargo, el método de continuación secuencial es capaz de obtener puntos en la parte cóncava del frente, como se muestra en el ejemplo 2 (FON).

- *Conjunto no dominado \mathcal{Y}_N separable.* Puesto que el método de Newton es esencialmente local, los métodos de continuación sólo obtienen el muestreo de una parte conectada del frente de Pareto en el caso que éste resulte separable (ver sección 2.5), como se muestra en el ejemplo 3 (POL).

A pesar de tales inconvenientes, ambos métodos de continuación son más económicos de implementar que los algoritmos evolutivos usados generalmente para problemas de optimización multiobjetivo, aunque estos últimos son aplicables a una clase más amplia de problemas. Sin embargo, en problemas de optimización de alta dimensionalidad o que involucran un número considerable de funciones objetivos, los métodos de continuación propuestos pueden ser una alternativa útil.

Bibliografía

- [1] M. Bazaraa, H. Sherali, C. Shetty. *Nonlinear Programming. Theory and Algorithms*. Second edition. John Wiley & Sons, 1993.
- [2] C. A. Coello Coello, G. Lamont, D. Van Veldhuizen. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Second edition. Springer, 2007.
- [3] Y. Collette, P. Siarry. *Multiobjective Optimization: Principles and Case Studies*. Springer, 2003.
- [4] C. Correa Flórez, R. Bolaños, A. Molina Cabrera. *Algoritmo Multiobjetivo NSGA-II Aplicado al Problema de la Mochila*. Scientia et Technica Año XIV, No 39, Septiembre de 2008. Universidad Tecnológica de Pereira.
- [5] G. Dantzig, M. Thapa. *Linear Programming 1: Introduction*. Springer, 1997.
- [6] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan. *A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II*. IEEE Transactions on Evolutionary Computation, Vol. 6, No. 2, April 2002.
- [7] M. Ehrgott. *Multicriteria Optimization*. Second edition. Springer, 2005.
- [8] G. Eichfelder. *Adaptive Scalarization Methods in Multiobjective Optimization*. Springer, 2008.
- [9] M. Hestenes. *Optimization Theory: the finite dimensional case*. John Wiley & Sons, 1975.
- [10] J. Jahn. *Vector Optimization. Theory, Applications, and Extensions*. Second edition. Springer, 2011.
- [11] D. Luenberger, Y. Ye. *Linear and Nonlinear Programming*. Third edition. Springer, 2008.
- [12] O. Mangasarian. *Nonlinear Programming*. McGraw-Hill, 1969.
- [13] M. Martínez, S. Santos, 1995. *Métodos computacionais de otimização*. Departamento de Matemática aplicada, IMECC-UNICAMP.
- [14] K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Boston, 1999 .
- [15] H. Nakayama, Y. Yun, M. Yoon. *Sequential Approximate Multiobjective Optimization Using Computational Intelligence*. Springer, 2009.

- [16] V. Pereyra. *Fast computation of equispaced Pareto manifolds and Pareto fronts for multiobjective optimization problems*. Mathematics and Computers in Simulation 79 (2009) 1935 - 1947. Elsevier.
- [17] E. Polak. *Optimization: Algorithms and Consistent Approximations*. Springer, 1997.
- [18] R. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [19] Y. Sawaragi, H. Nakayama, T. Tanino. *Theory of Multiobjective Optimization*. Mathematics in Science and Engineering, volume 176. Academic Press Inc., 1985.
- [20] G. Strang. *Linear Algebra and its Applications*. Third edition. Thomson Learning, 1988.