

UNIVERSIDAD NACIONAL DEL CALLAO

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

ESCUELA PROFESIONAL DE MATEMÁTICA



**MÉTODOS DE OPTIMIZACIÓN NO LINEAL PARA
PROBLEMAS DE ESTRUCTURAS**

**TESIS PARA OPTAR EL TÍTULO PROFESIONAL
DE LICENCIADO EN MATEMÁTICA**

DARWIN CASTILLO HUAMANÍ

Callao, 2018

PERÚ

UNIVERSIDAD NACIONAL DEL CALLAO
FACULTAD DE CIENCIAS NATURALES Y MATEMATICA

ACTA DE SUSTENTACION DE TESIS

En el Callao, en el Auditorio de la Facultad de Ciencias Naturales y Matemática, sito en la Av. Juan Pablo II N° 306, Bellavista, siendo las 1:25 p.m. del día lunes veintinueve de octubre del año dos mil dieciocho, se reunieron, a fin de proceder en primer término al acto de instalación, el Jurado Evaluador del Trabajo de Tesis presentado por el Señor Bachiller DARWIN CASTILLO HUAMANI, titulado: "MÉTODOS DE OPTIMIZACIÓN NO LINEAL PARA PROBLEMAS DE ESTRUCTURAS" Jurado que está integrado por los siguientes docentes de la Facultad de Ciencias Naturales y Matemática de la Universidad Nacional del Callao:

- Lic. ELMER ALBERTO LEÓN ZÁRATE : Presidente
- Lic. HERMINIA BERTHA TELLO BEDRIÑANA : Secretaria
- Lic. SOFIA IRENA DURAN QUIÑONES : Vocal

Luego de la instalación, se dio lectura, por el Secretario del Jurado, de la Resolución Decanal N° 072-2017-D-FCNM que designa a los miembros del Jurado Evaluador del Trabajo de Tesis.

A continuación, se dio inicio a la exposición del Trabajo de Tesis de acuerdo a lo normado por el Art. 81° del Reglamento de Grados y Títulos de la Universidad Nacional del Callao, aprobado por Resolución N° 309-2017-CU de fecha 24.10.2017, concordante con el Flujoograma de los Procedimientos de Grados y Títulos de Pre y Posgrado de la Universidad Nacional del Callao, aprobado por Resolución N° 1157-2011-R de fecha 18.11.2011.

Culminado el acto de exposición, los señores miembros del Jurado procedieron a formular las preguntas al indicado Bachiller, las mismas que fueron respondidas.

Luego de un cuarto intermedio, para la deliberación en privado del Jurado, y después de calificar el Trabajo de Tesis referido líneas arriba, se ACORDÓ CALIFICAR la Tesis sustentada por el Señor Bachiller DARWIN CASTILLO HUAMANI, para optar el Título Profesional de Licenciado en Matemática, según la puntuación cuantitativa y cualitativa que a continuación se indica:

| Calificación cuantitativa | Calificación cualitativa |
|---------------------------|--------------------------|
| 16 | MUY BUENO |


Finalmente, se procedió a leer en público el acta de sustentación redactada por la secretaria del jurado evaluador del trabajo de tesis.

Siendo las 2:15 p.m. del día lunes veintinueve de octubre del año dos mil dieciocho, el señor Presidente del Jurado dio por concluido el acto de sustentación de tesis.

En señal de conformidad con lo actuado, se levanta la presente acta con las siguientes firmas:


Lic. Elmer Alberto León Zárate
Presidente


Lic. Herminia Bertha Tello Bedriñana
Secretaria


Lic. Sofia Irena Duran Quiñones
Vocal

UNIVERSIDAD NACIONAL DEL CALLAO
OFICINA DE SECRETARIA GENERAL
EL SECRETARIO GENERAL DE LA UNIVERSIDAD NACIONAL DEL CALLAO que suscribe, CERTIFICA: Que la presente es copia fiel del original. Se expide la presente certificación a solicitud del (s) interesado (a) para los fines que juzgue conveniente.
Callao, 14 JUN 2021, del 20.....


UNIVERSIDAD NACIONAL DEL CALLAO
Oficina de Secretaría General
Lic. Ana Rosa Eudria Cisneros
Secretaria General (e)

DEDICATORIA

A mis padres por el apoyo continuo y la confianza.

AGRADECIMIENTO

A mis padres por el apoyo en mi educación y por la confianza en todas las decisiones de mi vida, dando fuerza para lograr mis metas.

A mis hermanos por su apoyo y amor fraterno.

A mi asesor Montoro, le agradezco la enorme confianza depositada en mí, la inmensa disposición y paciencia para resolver mis dudas durante mi carrera universitaria.

A mis amigos de la UNAC, por compartir momentos de felicidad, amistad y sinceridad. Principalmente, Berceli, Carlos, Abraham, Leibniz y Luis.

Índice general

| | |
|--|----------|
| Resumen | 3 |
| Abstract | 4 |
| INTRODUCCIÓN | 5 |
| I. PLANTEAMIENTO DEL PROBLEMA | 6 |
| 1.1. Determinación del problema | 6 |
| 1.2. Formulación del problema | 7 |
| 1.2.1. Problema general | 7 |
| 1.2.2. Problema específico | 7 |
| 1.3. Objetivo de la investigación | 7 |
| 1.4. Justificación | 8 |
| II. MARCO TEÓRICO | 9 |
| 2.1. Antecedentes del estudio | 9 |
| 2.2. Optimización topológica | 10 |
| 2.2.1. Formulación del problema de la minimización de la flexibilidad | 10 |
| 2.3. Método de los elementos finitos | 12 |
| 2.3.1. Relación entre tensión y deformación | 12 |
| 2.3.2. Relación entre desplazamiento y deformación | 12 |
| 2.3.3. Elementos finitos bidimensionales planos | 13 |
| 2.3.4. Interpolación de los desplazamientos | 13 |
| 2.3.5. Matriz de rigidez y el vector de carga del elemento finito . | 15 |
| 2.3.6. Montaje de la matriz de rigidez global | 16 |
| 2.4. Tablero de ajedrez | 18 |
| 2.4.1. Filtro de la media ponderada de las densidades | 18 |
| 2.5. Método de las Asíntotas Móviles | 19 |
| 2.5.1. Secuencia de subproblemas de aproximación convexa | 19 |
| 2.5.2. Método dual para resolver los subproblemas | 22 |
| 2.6. Aspectos prácticos de MMA | 24 |
| 2.6.1. Métodos de puntos interiores para resolver los subproblemas convexos | 26 |
| 2.6.2. Condiciones de optimalidad para el subproblema | 27 |
| 2.6.3. Una dirección de Newton para las condiciones de optimalidad | 28 |
| 2.6.4. Búsqueda lineal en la dirección de Newton | 30 |

| | |
|---|-----------|
| 2.7. Método de gradiente espectral proyectado no monótono | 34 |
| 2.8. Definiciones de términos básicos | 37 |
| III. HIPÓTESIS Y VARIABLES | 38 |
| 3.1. Hipótesis general e hipótesis específicas | 38 |
| 3.2. Operacionalización de variables | 38 |
| IV. METODOLOGÍA DE LA INVESTIGACIÓN | 39 |
| 4.1. Tipo de investigación | 39 |
| 4.2. Diseño de la investigación | 39 |
| 4.3. Población y muestra | 39 |
| 4.4. Técnicas e instrumentos de recolección de datos | 39 |
| 4.5. Plan de análisis estadísticos de datos | 39 |
| V. RESULTADOS | 40 |
| 5.1. Convergencia del método SPG | 40 |
| 5.2. Resultados numéricos | 42 |
| VI. CONCLUSIONES | 45 |
| VII. RECOMENDACIONES | 46 |
| VIII. REFERENCIAS BIBLIOGRÁFICAS | 47 |
| ANEXO | 49 |

Resumen

En este trabajo, estudiamos el método de gradiente espectral proyectado no monótono para resolver problemas de optimización de estructuras. La propuesta del método presente fue por su simple implementación y por la eficiencia en problemas de gran escala. Este método está compuesto por dos ingredientes que son el paso de Barzilai-Borwein y la búsqueda lineal no monótona de Grippo, Lampariello, y Lucidi. El desempeño del método de gradiente espectral proyectado no monótono es estudiado a través de ejemplos numéricos en dos dimensiones.

Abstract

In this work, we studied the nonmonotone spectral projected gradient method to solve structural optimization problems. The proposal of the present method was for its simple implementation and efficiency in large scale problems. This method is composed of two ingredients that are the Barzilai-Borwein step and the nonmonotone linear search of Grippo, Lampariello, and Lucidi. The performance of the nonmonotone spectral projected gradient method is studied through two-dimensional numerical examples.

INTRODUCCIÓN

Un problema clásico de optimización de estructuras es la búsqueda de la distribución óptima de un material dentro de un dominio que minimice la flexibilidad de una estructura considerando condiciones de frontera, una fuerza externa y una restricción de volumen. Bendsøe [4] presentó el método de las densidades, o *Solid Isotropic Material with Penalization* (SIMP), para resolver el presente problema usando un conjunto de métodos computacionales, optimización no lineal y método de los elementos finitos. Después de obtener la discretización del dominio, usamos una función binaria para representar la selección del material. La función toma el valor de uno si colocamos material y cero si retiramos material. Tomando como base esa idea, el problema de optimización se convierte en un problema de optimización con variables discretas. Sin embargo, dada la dificultad de resolver problemas de optimización de gran escala con variables discretas, la función binaria es reemplazada por una función de variable continua entre 0 y 1, y el problema es convertido en un problema de optimización continua con valores intermediarios, $0 < x_e < 1$, que son usualmente empleados para la construcción de la estructura óptima. En el trabajo de Bendsøe [4], se sustituye la función binaria por una función potencia x_e^p continua, con valores adecuados de p [14], para penalizar los valores intermediarios de las variables, denominadas densidades.

En este trabajo, resolvemos el problema de la minimización de la flexibilidad de una estructura usando el método de gradiente espectral proyectado no monótono o *Spectral Project Gradient* (SPG) [6]. El desempeño del método propuesto será estudiado a través de ejemplos numéricos en dos dimensiones y comparado con el método de las asíntotas móviles, o *The Method of Moving Asymptotes* (MMA) que fue introducido por Svanberg [15].

I. PLANTEAMIENTO DEL PROBLEMA

1.1. Determinación del problema

El problema de optimización abordado en este trabajo consiste en obtener la distribución óptima de un material dentro de un dominio tal que la flexibilidad de la estructura sea minimizada considerando una fuerza externa, condición de equilibrio estático y una restricción de volumen.

Para formular el problema de la minimización de la flexibilidad de una estructura, discretizamos el dominio aplicando el método de los elementos finitos. Luego, en cada elemento debemos colocar o retirar un material. Este criterio de selección de material dentro de un elemento e está representado por la función binaria

$$\mathbf{x}_e = \begin{cases} 1, & \text{colocar material} \\ 0, & \text{caso contrario.} \end{cases} \quad (1.1)$$

De este modo, formulamos un problema de optimización de variables enteras. El costo de obtener la solución de un problema de optimización de gran escala con ese tipo de variables es muy alto. Para evitar este inconveniente, usamos el método SIMP que reemplaza la variable binaria (1.1) por una variable continua ($0 \leq x_e \leq 1$) y define la siguiente función para cada elemento e del dominio discretizado:

$$\mathbf{E}_e(\mathbf{x}) = \mathbf{E}_0 + x_e^p(\mathbf{E} - \mathbf{E}_0) \quad (1.2)$$

donde \mathbf{E} es el módulo de *Young* del material, \mathbf{E}_0 un valor positivo muy pequeño (módulo artificial), y x_e representa una pseudodensidad del elemento e del dominio discretizado, observe que para valores adecuados de p ($p \geq 3$) conseguimos disminuir los valores intermedios.

A continuación formulamos el problema de la minimización de la flexibilidad de una estructura

$$\begin{aligned} & \text{minimizar} && \mathbf{C}(\mathbf{x}) = \mathbf{F}^T \mathbf{U} = \mathbf{U}^T \mathbf{K}(\mathbf{x}) \mathbf{U} \\ & \text{s a} && \mathbf{K}(\mathbf{x}) \mathbf{U} = \mathbf{F} \\ & && \sum_{e=1}^{nel} x_e v_e - \eta \mathbf{V} \leq 0 \\ & && 0 \leq x_e \leq 1, \quad \forall e \end{aligned} \quad (1.3)$$

donde \mathbf{C} es la flexibilidad de la estructura, nel es el número de elementos del dominio discretizado, \mathbf{K} es la matriz de rigidez, v_e es el volumen del elemento e , \mathbf{V} es el volumen total de la estructura, η es la fracción del volumen total, \mathbf{F} es el vector de fuerzas nodales asociadas al cargamento externo, \mathbf{E}_0 debe tener un

valor positivo muy pequeño para evitar que la matriz \mathbf{K} se vuelva singular. Se observa que la matriz de rigidez se define como:

$$\mathbf{K}(\mathbf{x}) = \sum_{e=1}^{nel} \mathbf{E}_e(\mathbf{x}) \hat{\mathbf{K}}_e \quad (1.4)$$

donde $E_e(x)$ es la función (1.2) y $\hat{\mathbf{K}}_e$ es la matriz definida en la sección 2.3.6.

1.2. Formulación del problema

1.2.1. Problema general

¿Es posible adaptar el método de gradiente espectral proyectado no monótono para resolver problemas de optimización estructural?

1.2.2. Problema específico

¿Es posible resolver el problema de la minimización de la flexibilidad de una estructura con el método SPG?

¿Es posible que el método SPG obtenga la solución del problema de minimización en menor tiempo que el método de MMA?

1.3. Objetivo de la investigación

Objetivo general

Resolver el problema de la minimización de la flexibilidad de una estructura usando el método de gradiente espectral proyectado no monótono.

Objetivo específicos

1. Estudiar la convergencia del método SPG.
2. Implementar el algoritmo del método SPG en conjunto con el método de los elementos finitos para resolver el problema de minimización.
3. Analizar los resultados numéricos del método propuesto.

1.4. Justificación

En la industria automotriz, se procura la reducción de materias primas en la fabricación de vehículos, pocos gramos por vehículos en la producción de varios millones de unidades supone un ahorro de toneladas de material. Algunos Softwares comerciales que fueron usados para la fabricación de piezas óptimas de los vehículos emplearon en el método de MMA. A pesar de la eficiencia del método de MMA para resolver el problema de minimización de la flexibilidad de una estructura, la solución obtenida por el método carece de justificación teórica, no tiene prueba de convergencia global.

En este trabajo, proponemos el método SPG para resolver el problema de la minimización de la flexibilidad de una estructura. Escogimos el método SPG por su teoría de convergencia global y su fácil implementación.

II. MARCO TEÓRICO

2.1. Antecedentes del estudio

El problema de minimización de la flexibilidad de una estructura pertenece a los problemas de Optimización Topológica (OT), que busca obtener la topología óptima de una estructura satisfaciendo un conjunto de restricciones dadas y un criterio de costo, es un tema reciente en el campo de optimización estructural. Sin embargo, los conceptos básicos que dan soporte teórico al método se establecieron hace más de un siglo.

La gran ventaja de la OT en contraste con los métodos tradicionales de optimización, como la optimización de forma o la optimización paramétrica, es que estos métodos no son capaces de cambiar el layout de la estructura original; así, no auxilian el proyecto conceptual de la estructura. De esta forma, la aplicación de la OT se ha mostrado un camino provechoso de investigaciones vinculadas a proyectos de interés de las industrias, ya que facilita el modelamiento de materiales bajo ciertas condiciones de diseño.

Svanberg [15] presentó el método de MMA para revolver problemas asociados a OT, este método resuelve una secuencia de subproblemas convexos explícitos que son generados de un problema no lineal. Otro método que tuvo éxito para resolver el problema de OT fue el método de programación lineal secuencial [10], es un método iterativo que utiliza programación lineal hasta alcanzar la convergencia deseada. Después de calcular el punto inicial la función objetivo y sus restricciones son aproximadas linealmente en ese punto y el problema puede ser resuelto con programación lineal (PL). La nueva solución obtenida por PL puede ser usada con un nuevo punto entorno del cual la función objetivo y sus restricciones serán nuevamente linealizadas y este proceso continúa hasta que el criterio de convergencia sea alcanzado.

2.2. Optimización topológica

Optimización topológica es un conjunto de métodos computacionales con el propósito de obtener una estructura con máxima rigidez, que sean capaces de soportar cargas externas sin que ocurra grandes desplazamientos y deformaciones, satisfaciendo ciertas condiciones, por ejemplo, disponer de una cantidad máxima de material.

Otras estrategias también son usadas para encontrar la estructura más rígida. En la optimización paramétrica [18], a partir de una estructura conocida, se busca optimizar las variables del proyecto, por ejemplo, la espesura o el área de la sección transversal de una estructura. En la optimización de forma [11], se busca optimizar el contorno de la estructura en cada etapa del proceso de optimización.

A diferencia de las dos estrategias presentadas arriba, la optimización topológica no precisa conocer previamente la forma de la estructura, solo basta definir la dimensión de la estructura, la fuerza aplicada y puntos de apoyo.

2.2.1. Formulación del problema de la minimización de la flexibilidad

Considere una estructura como un cuerpo elástico en un dominio Ω que debe ser mantenido fijo en la región Γ_d y que sea sometido a la aplicación de una fuerza (carga externa) t en la región Γ_t , vea la Figura 2.1 . La carga externa t aplicado en la región Γ_t , genera un campo de desplazamiento $\mathbf{u} = [u_1, u_2]^T$ tal que las componentes de \mathbf{u} son funciones que pertenece a un conjunto de funciones admisibles definido por

$$V = \{v_i \in H^1(\Omega) \mid v_i = 0 \text{ em } \Gamma_d, i = 1, 2\},$$

donde $H^1(\Omega)$ denota el espacio de funciones $f : \Omega \rightarrow \mathbb{R}$ que satisfacen

$$\left[\int_{\Omega} f^2 d\Omega \right]^{1/2} < \infty \text{ y } \left[\int_{\Omega} \|\nabla f\|^2 d\Omega \right]^{1/2} < \infty.$$

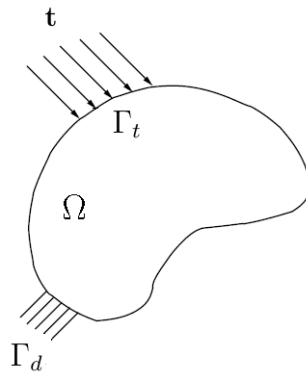


Figura 2.1: Cuerpo sometido a la carga t

El trabajo externo realizado por la carga t es definido por la forma lineal $L : V \rightarrow V$, dada por

$$L(\mathbf{u}) = \int_{\Gamma_t} \mathbf{t} \cdot \mathbf{u} \, d\Gamma \quad \text{para } \mathbf{u} \in V \quad (2.5)$$

El trabajo interno asociado a la deformación cuando aplicamos la carga t es definido por la forma bilineal:

$$a(\mathbf{u}, \mathbf{v}) = \int_{\Omega} \varepsilon(\mathbf{u})^T \mathbf{C} \varepsilon(\mathbf{v}) \, d\Omega, \quad \mathbf{u}, \mathbf{v} \in V, \quad (2.6)$$

donde ε es un tensor de deformaciones y \mathbf{C} es un tensor de elasticidad simétrico relacionado a las propiedades elásticas de la estructura. Usando las definiciones (2.5) y (2.6), la condición de equilibrio estático del cuerpo es establecida por la ecuación

$$a(\mathbf{u}, \mathbf{v}) = L(\mathbf{v}), \quad \forall \mathbf{v} \in V. \quad (2.7)$$

La energía potencial del cuerpo es definida por

$$\Pi = \frac{1}{2}a(\mathbf{u}, \mathbf{u}) - L(\mathbf{u}), \quad \mathbf{u} \in V. \quad (2.8)$$

La condición de equilibrio estático es equivalente a la solución del problema de minimización de la energía potencial del cuerpo. La solución que minimiza la energía potencial es obtenida igualando a cero la primera variación del funcional Π .

Además, al construir la estructura disponemos de una cantidad limitada de volumen. Por lo tanto, imponemos una restricción que establece el volumen máximo $\Omega_{\text{máx}}$ que la estructura debe tener, es decir, exigimos que

$$\int_{\Omega} \rho(x) \, d\Omega \leq \Omega_{\text{máx}}, \quad (2.9)$$

donde $\rho : \Omega \rightarrow [0, 1]$ es la función que representa las densidades del material en cada punto del dominio Ω . En consecuencia, la formulación del problema de la minimización de la flexibilidad de una estructura usando la condición de equilibrio (2.7) y la restricción de volumen (2.9) es dado por

$$\begin{aligned} & \underset{\rho}{\text{mín}} && L(\mathbf{u}) \\ & \text{s a} && a(\mathbf{u}, \mathbf{v}) = L(\mathbf{v}) \quad \forall \mathbf{v} \in V \\ & && \int_{\Omega} \rho(x) \, d\Omega \leq \Omega_{\text{máx}} \\ & && 0 \leq \rho(x) \leq 1, \quad \forall x \in \Omega. \end{aligned} \quad (2.10)$$

donde L es la flexibilidad de la estructura, la ecuación $a(\mathbf{u}, \mathbf{v}) = L(\mathbf{v}), \forall \mathbf{v} \in V$, representa la condición de equilibrio estático del cuerpo, la integral $\int_{\Omega} \rho(x) \, d\Omega$ es el volumen de la estructura y $\Omega_{\text{máx}}$ es el limitante del volumen.

2.3. Método de los elementos finitos

En esta sección presentamos de una forma simple el método de los elementos finitos (MEF) [2], que es usado para resolver numéricamente la ecuación de equilibrio estático de una estructura. En este trabajo, supondremos que el espesor de la estructura es constante y muy pequeño. Se aborda inicialmente dos relaciones que ocurren en cada punto del cuerpo, que son la relación entre tensión y deformación y la relación entre desplazamiento y deformación. En la secuencia se presenta el método de los elementos finitos rectangulares.

2.3.1. Relación entre tensión y deformación

De ahora en adelante, supondremos que tipo de material que compone a la estructura será del tipo elástico, lineal e isotrópico (las propiedades del material no dependen de la dirección). Además, la estructura utilizada debe cumplir la condición de estado plano de tensión, es decir, el espesor del cuerpo debe ser muy pequeña si se compara con sus otras dimensiones. Considerando todas estas propiedades, así como la ley de *Hooke* [9] arriba, tenemos la siguiente ecuación matricial que relaciona la tensión $\boldsymbol{\sigma}$ y la deformación $\boldsymbol{\varepsilon}$ en un punto de la estructura.

$$\begin{bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{bmatrix} = E \frac{1}{1 - \nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1-\nu}{2} \end{bmatrix} \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{bmatrix}$$

o

$$\boldsymbol{\sigma} = E\mathbf{D}\boldsymbol{\varepsilon} = \mathbf{C}\boldsymbol{\varepsilon}, \quad (2.11)$$

donde $E > 0$ es el módulo de *Young* (el módulo de elasticidad) del material de la estructura y $0 < \nu < 1$ es el coeficiente de *Poisson*, que se define como la razón entre las deformaciones lateral y axial del material.

2.3.2. Relación entre desplazamiento y deformación

La relación entre deformación $\boldsymbol{\varepsilon}$ y desplazamiento \mathbf{u} en cada punto de la estructura es expresada por

$$\varepsilon_x = \frac{\partial u}{\partial x}, \quad \varepsilon_y = \frac{\partial v}{\partial y}, \quad \gamma_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}$$

o

$$\begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial y} & \frac{\partial}{\partial x} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

o todavía

$$\boldsymbol{\varepsilon} = \mathbf{L}\mathbf{u} \quad (2.12)$$

2.3.3. Elementos finitos bidimensionales planos

El MEF divide el dominio continuo Ω en un número finito de subdominios Ω_e denominados elementos finitos. En este trabajo, usaremos elemento finito rectangular en la construcción de la malla que contendrá 4 nodos localizados en sus vértices, con dos grados de libertad por nodo, esto es desplazamiento nodal en la horizontal y en la vertical. De este modo, el MEF genera una malla rectangular donde los rectángulos son interconectados por los nodos, como se ve en la Figura 2.2b. Note que en cada punto del elemento Ω_e , ocurren desplazamientos debido a la carga externa aplicado a la estructura. Esos desplazamientos deben ser aproximados por una función interpoladora. Cuanto mayor sea el número de elementos finitos, mejor será la aproximación obtenida para el problema estático, aunque el costo computacional también aumente.

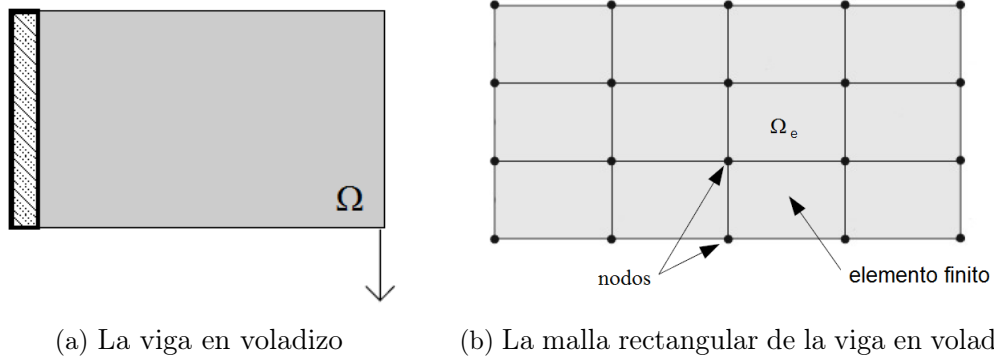


Figura 2.2: Discretización de la viga en voladizo usando el método de los elementos finitos con elementos rectangulares.

2.3.4. Interpolación de los desplazamientos

Los desplazamientos de los nodos de un elemento finito que surgen por la aplicación de fuerzas externas pueden ser aproximados por interpolación polinomial. Los nodos son usados para crear las funciones interpoladoras, que dependen de los valores de x e y cuyo sistema de referencia está situado en el nodo 1 del elemento finito, conforme a la Figura 2.3a. Tales funciones interpoladores son dadas por

$$u(x, y) = c_1 + c_2x + c_3y + c_4xy,$$

$$v(x, y) = c_5 + c_6x + c_7y + c_8xy,$$

o, en relación a los desplazamiento nodales, por

$$u = \sum_{i=1}^4 N_i u_i,$$

$$v = \sum_{i=1}^4 N_i v_i,$$

siendo las funciones de la forma:

$$\begin{aligned} N_1(x, y) &= \left(1 - \frac{x}{l}\right) \left(1 - \frac{y}{h}\right), \\ N_2(x, y) &= \left(1 - \frac{y}{h}\right) \frac{x}{l}, \\ N_3(x, y) &= \frac{xy}{lh}, \\ N_4(x, y) &= \left(1 - \frac{x}{l}\right) \frac{y}{h}. \end{aligned}$$

Podemos simplificar la formulación si consideramos el sistema de coordenadas en el baricentro del elemento finito Ω_e , vea la Figura 2.3b. En este caso, las funciones de forma para los desplazamientos u y v son:

$$\begin{aligned} N_1(\xi, \eta) &= \frac{1}{4}(1 - \xi)(1 + \eta) \\ N_2(\xi, \eta) &= \frac{1}{4}(1 - \xi)(1 - \eta) \\ N_3(\xi, \eta) &= \frac{1}{4}(1 + \xi)(1 - \eta) \\ N_4(\xi, \eta) &= \frac{1}{4}(1 + \xi)(1 + \eta) \end{aligned}$$

siendo

$$\xi = \frac{x}{a} \quad \text{e} \quad \eta = \frac{y}{b}. \quad (2.13)$$

ahora, escribimos los desplazamientos en la forma matricial

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 & 0 \\ 0 & N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \\ u_4 \\ v_4 \end{bmatrix}$$

o

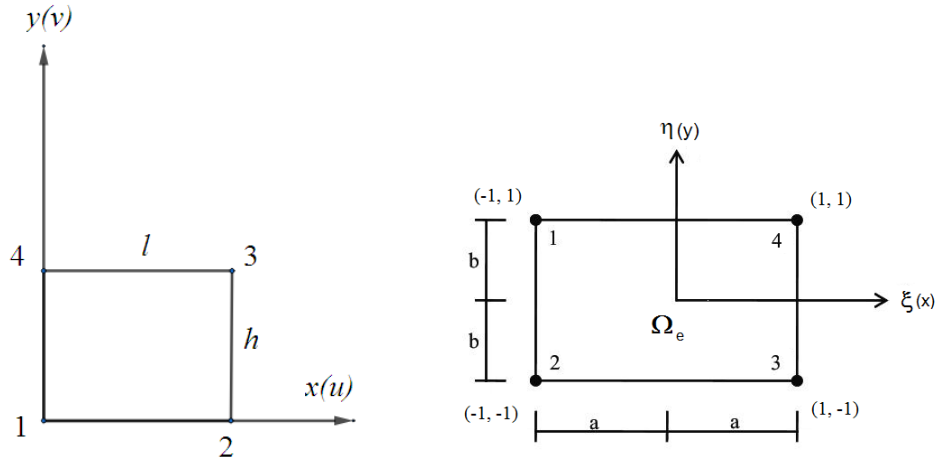
$$\mathbf{u} = \mathbf{Nw} \quad (2.14)$$

De (2.14) y (2.12), obtenemos

$$\boldsymbol{\varepsilon} = \mathbf{Lu} = \mathbf{LNw} = \mathbf{Bw} \quad (2.15)$$

donde

$$\mathbf{B} = \frac{1}{4} \begin{bmatrix} -\frac{1-\eta}{a} & 0 & \frac{1-\eta}{a} & 0 & \frac{1+\eta}{a} & 0 & -\frac{1+\eta}{a} & 0 \\ 0 & -\frac{1-\xi}{b} & 0 & -\frac{1+\xi}{b} & 0 & \frac{1+\xi}{b} & 0 & \frac{1-\xi}{b} \\ -\frac{1-\xi}{b} & -\frac{1-\eta}{a} & -\frac{1+\xi}{b} & \frac{1-\eta}{a} & \frac{1+\xi}{b} & \frac{1+\eta}{a} & \frac{1-\xi}{b} & -\frac{1+\eta}{a} \end{bmatrix}$$



(a) Origen del sistema en el nodo 1 (b) Sistema de coordenadas ξ e η ($\xi = \frac{x}{a}$ e $\eta = \frac{y}{b}$)

Figura 2.3: Representaciones del elemento finito rectangular

2.3.5. Matriz de rigidez y el vector de carga del elemento finito

De (2.8) podemos obtener la energía potencial para el elemento Ω_e . Además, considerando que el cuerpo está en equilibrio estático, la minimización de la energía potencial es obtenida de la anulación de la primera variación del funcional Π_e , es decir,

$$\delta\Pi_e = \int_{\Omega_e} \delta\boldsymbol{\varepsilon}^T C \boldsymbol{\varepsilon} d\Omega_e - \int_{\Gamma_e} \delta\mathbf{u}^{(e)T} \mathbf{b} d\Gamma_e = 0$$

o

$$\int_{\Omega_e} \delta\boldsymbol{\varepsilon}^T C \boldsymbol{\varepsilon} d\Omega_e = \int_{\Gamma_e} \delta\mathbf{u}^{(e)T} \mathbf{b} d\Gamma_e, \quad (2.16)$$

donde \mathbf{b} es la fuerza aplicada al elemento y Γ_e denota la frontera del elemento.

De (2.14) se puede escribir que la variación de los desplazamientos es dada por

$$\delta\mathbf{u}^{(e)} = \mathbf{N} \delta\mathbf{w}^{(e)} \quad (2.17)$$

y, de (2.15), se puede definir la primera variación de las deformaciones por

$$\delta\boldsymbol{\varepsilon} = \mathbf{B} \delta\mathbf{w}^{(e)}. \quad (2.18)$$

reemplazando (2.17) y (2.18) en (2.16), obtenemos

$$\delta\mathbf{w}^{(e)T} \left[\int_{\Omega_e} \mathbf{B}^T C \mathbf{B} d\Omega_e \right] \mathbf{w}^{(e)} = \delta\mathbf{w}^{(e)T} \int_{\Gamma_e} \mathbf{N}^T \mathbf{b} d\Gamma_e. \quad (2.19)$$

Considerando que los desplazamientos $\delta\mathbf{w}^{(e)}$ son arbitrarios y usando (2.11), la ecuación (2.19) se transforma

$$\left[\int_{\Omega_e} \mathbf{B}^T C \mathbf{B} d\Omega_e \right] \mathbf{w}^{(e)} = \int_{\Gamma_e} \mathbf{N}^T \mathbf{b} d\Gamma_e \quad (2.20)$$

o

$$\mathbf{E} \left[\int_{\Omega_e} \mathbf{B}^T \mathbf{D} \mathbf{B} d\Omega_e \right] \mathbf{w}^{(e)} = \int_{\Gamma_e} \mathbf{N}^T \mathbf{b} d\Gamma_e, \quad (2.21)$$

o todavía

$$\mathbf{E}^{(e)} \mathbf{K}^{(e)} \mathbf{w}^{(e)} = \mathbf{F}^{(e)}, \quad (2.22)$$

donde $\mathbf{K}^{(e)}$ es denominada matriz de rigidez del elemento, $\mathbf{w}^{(e)}$ es el vector de desplazamiento nodales del elemento y $\mathbf{F}^{(e)}$ es el vector de fuerzas nodales equivalentes del elemento. Considerando constante la espesura del elemento finito, h_e , podemos reescribir la matriz de rigidez del elemento como

$$\mathbf{K}^{(e)} = \int_{A_e} \mathbf{B}^T \mathbf{D} \mathbf{B} h_e dA_e, \quad (2.23)$$

donde A_e es el área de la mayor cara del elemento finito rectangular. De (2.13), obtenemos

$$dx dy = ab d\xi d\eta. \quad (2.24)$$

Así,

$$\mathbf{K}^{(e)} = ab h_e \int_{-1}^1 \int_{-1}^1 \mathbf{B}^T \mathbf{D} \mathbf{B} d\xi d\eta. \quad (2.25)$$

Note que la integral doble (2.25), la matriz \mathbf{B} depende de ξ e η . Una alternativa para calcular esa integral es el uso del método de integración de Gauss, que consiste en usar una aproximación del tipo

$$\int_{-1}^1 \int_{-1}^1 f(\xi, \eta) d\xi d\eta \approx \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} f(\xi_i, \eta_j) w_i w_j,$$

donde n_x y n_y representan, respectivamente, el número de puntos usados para aproximar la integral en ξ y en η , w_i y w_j son los pesos para cada punto de esa aproximación.

2.3.6. Montaje de la matriz de rigidez global

Para obtener el sistema de ecuaciones de equilibrio que represente la conectividad entre los elementos finitos, debemos adoptar alguna convención para la numeración de los grados de libertad asociados a cada elemento. En ese sentido, proponemos la siguiente convención para la conectividad entre los elementos: La numeración de los grados de la libertad de los elementos debe comenzar en la esquina superior izquierda del dominio y proseguir por columnas, de arriba hacia abajo, de modo que el último elemento sea aquel localizado en el esquina inferior derecha del dominio.

En la Figura 2.4, ilustra la convención para la numeración del dominio discretizado. El número de un elemento se muestra en el centro de éste. A la derecha de



Figura 2.4: Discretización del dominio de la viga en voladizo

cada nodo, está representado la enumeración de sus grados de libertad. Por ejemplo, el elemento 1 está conectado a sus grados de libertad 3, 4, 11, 12, 9, 10, 1 y 2 (exactamente en ese orden, comenzando en la esquina inferior izquierda del elemento y prosiguiendo en el sentido antihorario).

Para montar la matriz de rigidez global $\mathbf{K} \in \mathbb{R}^{2(n_n) \times 2(n_n)}$, donde n_n es el número total de nodos de la malla de elementos finitos, definimos la matriz $\mathbf{P}_e \in \mathbb{R}^{2(n_n) \times 8}$, que es formada por las columnas de la matriz de identidad $\mathbf{I} \in \mathbb{R}^{2(n_n) \times 2(n_n)}$ de tal forma que

$$\mathbf{K} = \sum_{e=1}^{nel} \mathbf{E}^{(e)} \hat{\mathbf{K}}_e$$

donde $\hat{\mathbf{K}}_e = \mathbf{P}_e \mathbf{K}^{(e)} \mathbf{P}_e^T$, $e = 1, \dots, nel$. Por ejemplo, el elemento 5 de la Figura 2.4 está conectada a sus grados de libertad 13, 14, 21, 22, 19, 20, 11 y 12, en ese orden. Luego, la matriz \mathbf{P}_5 , asociada a ese elemento, será dada por

$$\mathbf{P}_5 = [\mathbf{I}_{13} \quad \mathbf{I}_{14} \quad \mathbf{I}_{21} \quad \mathbf{I}_{22} \quad \mathbf{I}_{19} \quad \mathbf{I}_{20} \quad \mathbf{I}_{11} \quad \mathbf{I}_{12}]$$

donde \mathbf{I}_i representa la i -ésima columna de la matriz de identidad \mathbf{I} .

Por lo tanto, usando esa convención de numeración para los elementos, obtenemos la ecuación de equilibrio asociada para toda la estructura

$$\mathbf{K} \mathbf{w} = \mathbf{F}$$

donde $\mathbf{K} \in \mathbb{R}^{2(n_n) \times 2(n_n)}$ es la matriz de rigidez global, $\mathbf{w} \in \mathbb{R}^{2(n_n)}$ es el vector global de desplazamientos nodales de toda la estructura, y $\mathbf{F} \in \mathbb{R}^{2(n_n)}$ es el vector global de fuerzas nodales equivalentes. Este vector de fuerzas será obtenido por la superposición de los vectores de fuerzas nodales equivalentes $\mathbf{F}^{(e)}$ de manera similar a lo que se hizo para la matriz \mathbf{K} , utilizando estrictamente la misma convención para los índices definidos arriba.

2.4. Tablero de ajedrez

Un fenómeno típico que ocurre en los métodos de OT es la distribución heterogénea del material en la topología óptima, o sea la topología contiene regiones con elementos vacíos y con elementos totalmente rellenos, dispuestos de manera alternada, semejante a un tablero de ajedrez, véase Figura 2.5. Según Sigmund [7], la posible causa de esa formación del tablero de ajedrez se debe al uso de elementos finitos rectangulares. Para hacer frente a este fenómeno, utilizamos el siguiente filtro.

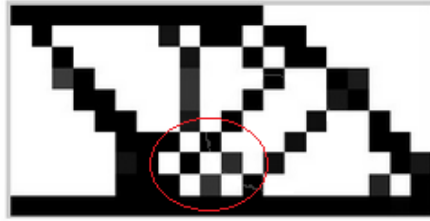


Figura 2.5: La región de la circunferencia de color rojo contiene el tablero de ajedrez.

2.4.1. Filtro de la media ponderada de las densidades

Este filtro se aplica a los elementos del dominio que se sustituyen por la media ponderada de los elementos contenidos en una vecindad, es decir, se sustituye x_e por

$$\hat{x}_e = \frac{1}{\sum_{i \in N_e} H_{ei}} \sum_{i \in N_e} H_{ei} x_i$$

donde N_e es el conjunto de elementos i para los cuales la distancia $\Delta(e, i)$ entre los centros de los elementos e e i es menor que el radio del filtro $r_{\text{mín}}$. Los pesos que se utilizan en esta media ponderada se definen como $H_{ei} = r_{\text{mín}} - \Delta(e, i)$.

2.5. Método de las Asíntotas Móviles

En [15], Svanberg desarrolló el método de las Asíntotas móviles o (MMA) que resuelve una secuencia de subproblemas explícitos que son generados de un problema de optimización no lineal. Este subproblema contiene funciones de aproximaciones de tipo convexa para las funciones del problema de programación no lineal. En esta sección, presentamos algunas propiedades de las funciones asociadas al método y su algoritmo.

Consideremos el siguiente problema general de optimización estructural con limitantes superior e inferior finitos para la variable de diseño \mathbf{x} .

$$\begin{aligned} \text{mín} \quad & C_0(\mathbf{x}) \\ \text{s.a} \quad & C_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & \hat{L}_j \leq x_j \leq \hat{U}_j, \quad j = 1, \dots, n \end{aligned} \quad (2.26)$$

donde $C_i : \mathbb{R} \rightarrow \mathbb{R}$, $i = 0, 1, \dots, m$ son funciones dos veces diferenciables.

2.5.1. Secuencia de subproblemas de aproximación convexa

La construcción de aquellas funciones de aproximación convexas y sus respectivos limitantes superior e inferior (asíntotas móviles), son hechas usando los gradientes de las funciones involucradas en el problema (2.26) y cuyos limitantes son actualizados de forma heurística. Se sigue la formulación de los subproblemas con las funciones de aproximación convexa del método MMA,

$$\begin{aligned} \text{mín} \quad & r_0^{(k)} + \sum_{j=1}^n \left(\frac{p_{0j}^{(k)}}{U_j^{(k)} - x_j} + \frac{q_{0j}^{(k)}}{x_j - L_j^{(k)}} \right) \\ \text{s.a} \quad & r_i^{(k)} + \sum_{j=1}^n \left(\frac{p_{ij}^{(k)}}{U_j^{(k)} - x_j} + \frac{q_{ij}^{(k)}}{x_j - L_j^{(k)}} \right) \leq 0, \quad i = 1, \dots, m \\ & \text{máx}\{\hat{L}_j, 0, 9L_j^{(k)} + 0, 1x_j^{(k)}\} \leq x_j \leq \text{mín}\{\hat{U}_j, 0, 9U_j^{(k)} + 0, 1x_j^{(k)}\}, \quad j = 1, \dots, n \end{aligned} \quad (2.27)$$

donde

$$p_{ij}^{(k)} = \text{máx}\{0, (U_j^{(k)} - x_j^{(k)})^2 \frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)})\},$$

$$q_{ij}^{(k)} = \text{mín}\{0, -(x_j^{(k)} - L_j^{(k)})^2 \frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)})\},$$

y

$$r_i^{(k)} = C_i(\mathbf{x}^{(k)}) - \sum_{j=1}^n \left(\frac{p_{ij}^{(k)}}{U_j^{(k)} - x_j^{(k)}} + \frac{q_{ij}^{(k)}}{x_j^{(k)} - L_j^{(k)}} \right).$$

Las propiedades de las funciones

$$\hat{C}_i^{(k)}(\mathbf{x}) = r_i^{(k)} + \sum_{j=1}^n \left(\frac{p_{ij}^{(k)}}{U_j^{(k)} - x_j} + \frac{q_{ij}^{(k)}}{x_j - L_j^{(k)}} \right), \quad i = 0, 1, \dots, m$$

del subproblema son expresada abajo

a) Las funciones \hat{C}_i son de primera orden, es decir

$$\begin{aligned}\hat{C}_i^{(k)}(\mathbf{x}^{(k)}) &= C_i(\mathbf{x}^{(k)}), \quad i = 0, 1, \dots, m \\ \frac{\partial \hat{C}_i^{(k)}}{\partial x_j}(\mathbf{x}^{(k)}) &= \frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}), \quad j = 1, \dots, n, i = 0, 1, \dots, m\end{aligned}$$

b) El calculo de las segundas derivadas de las funciones \hat{C}_i son:

$$\begin{aligned}\frac{\partial^2 \hat{C}_i^{(k)}}{\partial x_j^2}(\mathbf{x}) &= \frac{2p_{ij}^{(k)}}{(U_j^{(k)} - x_j)^3} + \frac{2q_{ij}^{(k)}}{(x_j - L_j^{(k)})^3} \quad \forall i, j \\ \frac{\partial^2 \hat{C}_i^{(k)}}{\partial x_j \partial x_l}(\mathbf{x}) &= 0 \quad \text{si } j \neq l\end{aligned}$$

c) Observe que, $p_{ij}^{(k)}$ e $q_{ij}^{(k)}$ son no negativos. Luego, \hat{C}_i es una función convexa. En particular, en el punto $\mathbf{x} = \mathbf{x}^{(k)}$, tenemos

$$\frac{\partial^2 \hat{C}_i^{(k)}}{\partial x_j^2}(\mathbf{x}^{(k)}) = \begin{cases} \frac{2}{U_j^{(k)} - x_j^{(k)}} \left(\frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}) \right), & \text{si } \partial C_i / \partial x_j(\mathbf{x}^{(k)}) > 0 \\ -\frac{2}{x_j^{(k)} - L_j^{(k)}} \left(\frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}) \right), & \text{si } \partial C_i / \partial x_j(\mathbf{x}^{(k)}) < 0. \end{cases} \quad (2.28)$$

Analicemos el comportamiento de las aproximaciones convexas con relación a los tres puntos $L_j^{(k)}$, $U_j^{(k)}$ e $x_j^{(k)}$.

d) Observe que (2.28), el valor de la segunda derivada no mixta será mayor, si $L_j^{(k)}$ e $U_j^{(k)}$ estuviesen más próximo de $x_j^{(k)}$, o sea, tenemos mayor curvatura de la función aproximadora $\hat{C}_i^{(k)}$. De este modo, las funciones aproximadoras serán más conservativas. Para ver esa propiedad, suponga que \hat{C}_i e \tilde{C}_i son dos funciones de aproximación diferentes que corresponden a los limitantes $\{\hat{L}_j^{(k)}, \hat{U}_j^{(k)}\}$ y $\{\tilde{L}_j^{(k)}, \tilde{U}_j^{(k)}\}$, respectivamente. Si consideramos $\hat{L}_j^{(k)} \leq \tilde{L}_j^{(k)} < x_j^{(k)} < \tilde{U}_j^{(k)} \leq \hat{U}_j^{(k)}$ para todo j . Entonces, todo los puntos \mathbf{x} tal que $\tilde{L}_j^{(k)} < x_j < \tilde{U}_j^{(k)}$ para todo j , verificaran que $\hat{C}_i(\mathbf{x}) \leq \tilde{C}_i(\mathbf{x})$.

e) En el caso que las asíntotas móviles $L_j^{(k)}$ y $U_j^{(k)}$ estuviesen suficientemente distantes de $\mathbf{x}^{(k)}$, entonces las funciones aproximadoras convexas $\hat{C}_i^{(k)}$ tienden a tener un comportamiento lineal en torno del punto $\mathbf{x}^{(k)}$. En particular, si $L_j^{(k)} \rightarrow -\infty$ y $U_j^{(k)} \rightarrow +\infty$, las funciones $\hat{C}_i^{(k)}$ se convierten en realidad lineales, es decir

$$\hat{C}_i^{(k)}(\mathbf{x}) = C_i(\mathbf{x}^{(k)}) + \nabla C_i(\mathbf{x}^{(k)})^T (\mathbf{x} - \mathbf{x}^{(k)}).$$

En MMA, el caso de nuestro interés es cuando $L_j^{(k)}$ e $U_j^{(k)}$ son finitos. De esa forma, cada función aproximadora $\hat{C}_i^{(k)}$ es estrictamente convexa en todas las variables x_j , excepto en los casos en que $\frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}) = 0$. Cuando eso ocurre, $\hat{C}_i^{(k)}$ es independiente de la variable x_j .

La actualización de los límites inferior $L_j^{(k)}$ y superior $U_j^{(k)}$ se hacen siguiendo las premisas de abajo:

- Si los iterando tienden a oscilar, entonces deben ser estabilizados. En este caso, las asíntotas deberán estar más cerca del punto actual $\mathbf{x}^{(k)}$.
- Si el proceso de resolución está monótono y lento, entonces las asíntotas deben ser “relajadas”, es decir, ellas deberán estar más distantes de el iterando actual $\mathbf{x}^{(k)}$.

Según Svanberg, en cada iteración los limitantes inferior $L_j^{(k)}$ y superior $U_j^{(k)}$ son actualizados como

- Para el caso de las dos primeras iteraciones $k = 1$ y $k = 2$, definamos

$$\begin{aligned} L_j^{(k)} &= x_j^{(k)} - 0,5(\hat{U}_j - \hat{L}_j), \\ U_j^{(k)} &= x_j^{(k)} + 0,5(\hat{U}_j - \hat{L}_j). \end{aligned}$$

- para $k \geq 3$, tomamos

$$\begin{aligned} L_j^{(k)} &= x_j^{(k)} - \gamma_j^{(k)}(x_j^{(k-1)} - L_j^{(k-1)}), \\ U_j^{(k)} &= x_j^{(k)} + \gamma_j^{(k)}(U_j^{(k-1)} - x_j^{(k-1)}). \end{aligned}$$

donde

$$\gamma_j^{(k)} = \begin{cases} 0,7, & \text{si } (x_j^{(k)} - x_j^{(k-1)})(x_j^{(k-1)} - x_j^{(k-2)}) < 0, \\ 1,2, & \text{si } (x_j^{(k)} - x_j^{(k-1)})(x_j^{(k-1)} - x_j^{(k-2)}) > 0, \\ 1, & \text{si } (x_j^{(k)} - x_j^{(k-1)})(x_j^{(k-1)} - x_j^{(k-2)}) = 0. \end{cases}$$

Algoritmo 1: Asíntotas Móviles

Mientras no es verificado algún criterio de parada, hacer
Determine $x^{(k+1)}$, solución de

$$\begin{aligned} \text{minimizar} \quad & r_0^{(k)} + \sum_{j=1}^n \left(\frac{p_{0j}^{(k)}}{U_j^{(k)} - x_j} + \frac{q_{0j}^{(k)}}{x_j - L_j^{(k)}} \right) \\ \text{s. a} \quad & r_i^{(k)} + \sum_{j=1}^n \left(\frac{p_{ij}^{(k)}}{U_j^{(k)} - x_j} + \frac{q_{ij}^{(k)}}{x_j - L_j^{(k)}} \right) \leq 0, \quad i = 1, \dots, m \\ & \alpha_j^{(k)} \leq x_j \leq \beta_j^{(k)}, \quad j = 1, \dots, n, \end{aligned} \tag{2.29}$$

$$\begin{aligned} \alpha_j^{(k)} &= \max\{\hat{L}_j, 0,9L_j^{(k)} + 0,1x_j^{(k)}\}, \\ \beta_j^{(k)} &= \min\{\hat{U}_j, 0,9U_j^{(k)} + 0,1x_j^{(k)}\}. \end{aligned}$$

2.5.2. Método dual para resolver los subproblemas

Para resolver los subproblemas descritos por el Algoritmo 1, vamos introducir algunos conceptos de dualidad en optimización no lineal. Sea un problema de optimización no lineal, denominado problema primal, podemos definir un problema dual asociado. Bajo ciertas hipótesis de convexidad y de cualificación de las restricciones, los problemas primal y dual tienen valores óptimos iguales para las respectivas funciones objetivo y, de esa manera, es posible resolver el primal indirectamente a través de la resolución del problema dual. El problema dual asociado al problema (2.29) es definido:

$$\begin{aligned} \text{máx} \quad & \mathcal{L}(\mathbf{x}, \mu) \\ \text{s. a} \quad & \mu \geq 0, \end{aligned}$$

donde

$$\mathcal{L}(\mathbf{x}, \mu) = \hat{C}_0^{(k)}(x) + \sum_{i=1}^m \mu_i \hat{C}_i^{(k)}(x)$$

donde $\mu \in \mathbb{R}^m$ es el vector de los multiplicadores de Lagrange o vector de variables duales asociados a las funciones aproximadoras $\hat{C}_i^{(k)}$, $i = 0, 1, \dots, m$. Obsérvese que el problema (2.29) es convexo, pues cada $\hat{C}_i^{(k)}$ es una función convexa, y separable, pues cada término del sumatorio de $\hat{C}_i^{(k)}$ depende de una variable. Luego, de algunas manipulaciones algebraicas, obtenemos:

$$\mathcal{L}(\mathbf{x}, \mu) = r_0^{(k)} + \mu^T r^{(k)} + \sum_{j=1}^n \mathcal{L}_j(x_j, \mu)$$

donde

$$\mathcal{L}_j(x_j, \mu) = \frac{p_{0j}^{(k)} + \mu^T p_j^{(k)}}{U_j^{(k)} - x_j} + \frac{q_{0j}^{(k)} + \mu^T q_j^{(k)}}{x_j - L_j^{(k)}}$$

y

$$\begin{aligned} r^{(k)} &= (r_1^{(k)} \cdots r_m^{(k)})^T, & p_j^{(k)} &= (p_{1j}^{(k)} \cdots p_{mj}^{(k)})^T, \\ q_j^{(k)} &= (q_{1j}^{(k)} \cdots q_{mj}^{(k)})^T, & \mu &= (\mu_1 \cdots \mu_m)^T. \end{aligned}$$

Luego, la función objetivo dual $W(\mu)$ es dada por

$$\begin{aligned} W(\mu) &= \min_{\mathbf{x}} \{ \mathcal{L}(\mathbf{x}, \mu); \alpha_j^{(k)} \leq \mathbf{x}_j \leq \beta_j^{(k)} \text{ para todo } \mathbf{j} \} \\ &= r_0^{(k)} + \mu^T r^{(k)} + \sum_{j=1}^n W_j(\mu), \end{aligned}$$

donde

$$W_j(\mu) = \min_{x_j} \{ \mathcal{L}_j(x_j, \mu); \alpha_j^{(k)} \leq x_j \leq \beta_j^{(k)} \}. \quad (2.30)$$

Para encontrar la solución explícita de la expresión de $W_j(\mu)$, determinamos la solución óptima del problema de optimización (2.30). Para esto, note que $\mu \geq 0$ implica que $p_{0j}^{(k)} + \mu^T p_j^{(k)} \geq 0$ y $q_{0j}^{(k)} + \mu^T q_j^{(k)} \geq 0$. Luego, $\mathcal{L}_j(x_j, \mu)$ es una función convexa de x_j . Cuando $p_{0j}^{(k)} + \mu^T p_j^{(k)} = 0$ y $q_{0j}^{(k)} + \mu^T q_j^{(k)} = 0$, $\mathcal{L}_j(x_j, \mu)$ no depende de x_j . En este caso, cualquier x_j entre $\alpha_j^{(k)}$ y $\beta_j^{(k)}$ minimiza $\mathcal{L}_j(x_j, \mu)$. De aquí en adelante, supondremos que al menos uno de los términos $p_{0j}^{(k)} + \mu^T p_j^{(k)} \geq 0$ o $q_{0j}^{(k)} + \mu^T q_j^{(k)} \geq 0$ sea estrictamente positivo. La derivada de $\mathcal{L}_j(x_j, \mu)$ con respecto a x_j es dada por:

$$\mathcal{L}'_j(x_j, \mu) = \frac{p_{0j}^{(k)} + \mu^T p_j^{(k)}}{(U_j^{(k)} - x_j)^2} - \frac{q_{0j}^{(k)} + \mu^T q_j^{(k)}}{(x_j - L_j^{(k)})^2},$$

y la segunda derivada de $\mathcal{L}_j(x_j, \mu)$ en relación a x_j es

$$\mathcal{L}''_j(x_j, \mu) = \frac{2(p_{0j}^{(k)} + \mu^T p_j^{(k)})}{(U_j^{(k)} - x_j)^3} + \frac{2(q_{0j}^{(k)} + \mu^T q_j^{(k)})}{(x_j - L_j^{(k)})^3}.$$

Como $\mathcal{L}''_j(x_j, \mu)$ es estrictamente positivo, $\mathcal{L}'_j(x_j, \mu)$ es estrictamente creciente en x_j . De esta forma, podemos obtener las siguientes conclusiones sobre la solución $x_j^*(\mu)$ en (2.30):

1. Si $\mathcal{L}'_j(\alpha_j^{(k)}, \mu) \geq 0$, entonces $x_j^*(\mu) = \alpha_j^{(k)}$.
2. Si $\mathcal{L}'_j(\beta_j^{(k)}, \mu) \leq 0$, entonces $x_j^*(\mu) = \beta_j^{(k)}$.
3. Si $\mathcal{L}'_j(\alpha_j^{(k)}, \mu) < 0$ y $\mathcal{L}'_j(\beta_j^{(k)}, \mu) > 0$, entonces

$$x_j^*(\mu) = \frac{(p_{0j}^{(k)} + \mu^T p_j^{(k)})^{1/2} L_j^{(k)} + (q_{0j}^{(k)} + \mu^T q_j^{(k)})^{1/2} U_j^{(k)}}{(p_{0j}^{(k)} + \mu^T p_j^{(k)})^{1/2} + (q_{0j}^{(k)} + \mu^T q_j^{(k)})^{1/2}} \quad (2.31)$$

es la única solución de la ecuación $\mathcal{L}'_j(x_j, \mu) = 0$. Como la expresión para la solución $x_j^*(\mu)$ del problema de minimización en (2.30) es dada de forma explícita, también existe una expresión explícita para la función objetivo dual $W(\mu)$, que es dada por:

$$W(\mu) = r_0^{(k)} + \mu^T r^{(k)} + \sum_{j=1}^n \frac{p_{0j}^{(k)} + \mu^T p_j^{(k)}}{U_j^{(k)} - x_j(\mu)} + \frac{q_{0j}^{(k)} + \mu^T q_j^{(k)}}{x_j(\mu) - L_j^{(k)}}$$

La derivada de $W(\mu)$ en relación a las variables duales μ_i son dadas por:

$$\frac{\partial W}{\partial \mu_i}(\mu) = r_i^{(k)} + \sum_{j=1}^n \left(\frac{p_{ij}^{(k)}}{U_j^{(k)} - x_j(\mu)} + \frac{q_{ij}^{(k)}}{x_j(\mu) - L_j^{(k)}} \right) \quad (2.32)$$

Por lo tanto, el problema dual correspondiente al subproblema (2.29) es dado por:

$$\begin{aligned} \text{máx} \quad & W(\mu) \\ \text{s.a} \quad & \mu_i \geq 0, \quad i = 1, \dots, m. \end{aligned} \tag{2.33}$$

Como $x_j(\mu)$ depende continuamente de μ , a menos que $p_{0j}^{(k)} + \mu^T p_j^{(k)} = 0$ y $q_{0j}^{(k)} + \mu^T q_j^{(k)} = 0$, tenemos, de la ecuación (2.32), que $W(\mu)$ es una función “suave”. Se puede demostrar fácilmente que $W(\mu)$ es una función cóncava, pues ella es un mínimo de un conjunto de funciones que son lineales en μ . Según Svanberg [15], es importante recordar que el problema (2.33) puede ser resuelto por cualquier método basado apenas en el vector gradiente, desde que se tome un cuidado especial con las variables no negativas μ . Una vez resuelto el problema dual (2.33), la solución óptima del subproblema primal es obtenido directamente substituyendo la solución óptima dual μ^* en la expresión (2.31).

2.6. Aspectos prácticos de MMA

En esta sección, presentamos el problema aumentado que fue propuesto por Svanberg [16]. El problema en cuestión es obtenido adicionando variables artificiales al problema original (2.26). De acuerdo con Svanberg [16], el problema aumentado posee soluciones factibles y un conjunto no vacío de puntos que verifican las condiciones Karush-Kuhn-Tucker (KKT).

Problema Aumentado

En el problema (2.26), adicionamos las variables artificiales y_1, y_2, \dots, y_m y z , obteniendo el siguiente problema aumentado

$$\begin{aligned} \text{Minimizar} \quad & C_0(\mathbf{x}) + a_0 z + \sum_{i=1}^m (c_i y_i + \frac{1}{2} d_i y_i^2) \\ \text{s. a} \quad & C_i(\mathbf{x}) - a_i z - y_i \leq 0, & i = 1, \dots, m \\ & \hat{L}_j \leq x_j \leq \hat{U}_j, & j = 1, \dots, n, \\ & y_i \geq 0 \\ & z \geq 0. \end{aligned} \tag{2.34}$$

donde C_0, C_1, \dots, C_m son dos veces diferenciable, de valores reales, $a_0 > 0, a_i \geq 0, c_i \geq 0, d_i \geq 0$, e $c_i + d_i > 0$ para $i = 1, \dots, m$. Además de eso, $a_i c_i > a_0$ para todo i tal que $a_i > 0$. De acuerdo con Svanberg [16], para hacer el problema (2.34), equivalente al problema original (2.26), primero fijamos $a_0 = 1$ y $a_i = 0$ para todo $i > 0$. Por lo tanto, z la componente de cualquier solución óptima de (2.34) es cero. Si fijamos, también, $d_i = 1$ e $c_i = a$ (escalar muy grande) para todo $i > 0$. En consecuencia, tendremos $y_i = 0$ en cualquier solución óptima del problema (2.34), y el vector correspondiente x será una solución óptima del problema (2.26).

Obsérvese que el problema (2.34) siempre tiene soluciones factibles, y al menos una solución óptima. Esto es válido mismo que el problema (2.26) no tenga soluciones factibles, en cuyo caso algunos $y_i > 0$ en la solución óptima de (2.34).

Si siguiendo las notaciones usadas en la sección 2.5.1, escribimos el subproblema convexo, con algunas modificaciones en las funciones de aproximación convexa mostradas por Svanberg [16]. El subproblema convexo es dado por

$$\begin{aligned}
\text{minimizar } & \hat{C}_0^{(k)}(\mathbf{x}) + a_0 z + \sum_{i=1}^m (c_i y_i + \frac{1}{2} d_i y_i^2) \\
\text{s a } & \hat{C}_i^{(k)}(\mathbf{x}) - a_i z - y_i \leq 0, & i = 1, \dots, m \\
& \alpha_j^{(k)} \leq x_j \leq \beta_j^{(k)}, & j = 1, \dots, n, \\
& y_i \geq 0 & i = 1, \dots, m \\
& z \geq 0.
\end{aligned} \tag{2.35}$$

donde

$$\begin{aligned}
\alpha_j^{(k)} &= \text{máx}\{\hat{L}_j, 0, 9L_j^{(k)} + 0, 1x_j^{(k)}\}, \\
\beta_j^{(k)} &= \text{mín}\{\hat{U}_j, 0, 9U_j^{(k)} + 0, 1x_j^{(k)}\}.
\end{aligned}$$

Las funciones de aproximación son dadas por

$$\hat{C}_i^{(k)}(\mathbf{x}) = \sum_{j=1}^n \left(\frac{p_{ij}^{(k)}}{U_j^{(k)} - x_j} + \frac{q_{ij}^{(k)}}{x_j - L_j^{(k)}} \right) + r_i^{(k)}, \quad i = 1, \dots, m,$$

donde

$$\begin{aligned}
p_{ij}^{(k)} &= (U_j^{(k)} - x_j^{(k)})^2 \left(\left(\frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}) \right)^+ + k_{ij}^{(k)} \right), \\
q_{ij}^{(k)} &= (x_j^{(k)} - L_j^{(k)})^2 \left(\left(\frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}) \right)^- + k_{ij}^{(k)} \right), \\
r_i^{(k)} &= C_i(\mathbf{x}^{(k)}) - \sum_{j=1}^n \left(\frac{p_{ij}^{(k)}}{U_j^{(k)} - x_j^{(k)}} + \frac{q_{ij}^{(k)}}{x_j^{(k)} - L_j^{(k)}} \right), \\
\left(\frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}) \right)^+ &= \text{máx} \left\{ 0, \frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}) \right\}, \\
\left(\frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}) \right)^- &= \text{máx} \left\{ 0, -\frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}) \right\}.
\end{aligned}$$

Los valores de las variables $k_{ij}^{(k)}$ son dados por

$$\begin{aligned}
k_{ij}^{(k)} &= 0, \text{ para } i = 1, \dots, m, \quad j = 1, \dots, n \\
k_{0j}^{(k)} &= 10^{-3} \left| \frac{\partial C_i}{\partial x_j}(\mathbf{x}^{(k)}) \right| + \frac{10^{-8}}{U_j^{(k)} - L_j^{(k)}}, \text{ para } j = 1, \dots, n.
\end{aligned}$$

2.6.1. Métodos de puntos interiores para resolver los subproblemas convexos

En esta sección, se describe el método de puntos interiores primal - dual (Svanberg [16]) para resolver el subproblema (2.35). Note que los subproblemas (2.35) son estrictamente convexos y la condición de cualificación de restricción Slater [5] es verificada.

Para simplificar la notación, el índice k será eliminado en el subproblema (2.35). Además, denotaremos $b_i = -r_i^{(k)}$ y descartaremos la constante $r_0^{(k)}$ de la función objetivo. De este modo, el subproblema (2.35) se torna

$$\begin{aligned} \text{minimizar} \quad & g_0(x) + a_0 z + \sum_{i=1}^m (c_i y_i + \frac{1}{2} d_i y_i^2) \\ \text{s a} \quad & g_i(x) - a_i z - y_i \leq b_i, & i = 1, \dots, m \\ & \alpha_j \leq x_j \leq \beta_j, & j = 1, \dots, n, \\ & y_i \geq 0 \\ & z \geq 0. \end{aligned} \tag{2.36}$$

donde

$$g_i(x) = \sum_{j=1}^n \left(\frac{p_{ij}}{U_j - x_j} + \frac{q_{ij}}{x_j - L_j} \right), \quad i = 0, 1, \dots, m,$$

con

$$L_j < \alpha_j < \beta_j < U_j \quad \text{para todo } j.$$

La función Lagrangiana correspondiente al problema (2.36) es dado por

$$\begin{aligned} L = g_0(x) + a_0 z + \sum_{i=1}^m (c_i y_i + \frac{1}{2} d_i y_i^2) + \sum_{i=1}^m \lambda_i (g_i(x) - a_i z - y_i - b_i) + \\ \sum_{j=1}^n (\xi_j (\alpha_j - x_j) + \eta_j (x_j - \beta_j)) - \sum_{i=1}^m \mu_i y_i - \zeta z, \end{aligned}$$

donde $\lambda = (\lambda_1, \dots, \lambda_m)^T$, $\xi = (\xi_1, \dots, \xi_n)^T$, $\eta = (\eta_1, \dots, \eta_n)^T$, $\mu = (\mu_1, \dots, \mu_m)^T$, e ζ son los multiplicadores de Lagrange no negativos para las restricciones de (2.36). Definiendo la función

$$\psi(x, \lambda) = g_0(x) + \sum_{i=1}^m \lambda_i g_i(x) = \sum_{j=1}^n \left(\frac{p_j(\lambda)}{U_j - x_j} + \frac{q_j(\lambda)}{x_j - L_j} \right),$$

donde

$$p_j(\lambda) = p_{0j} + \sum_i^m \lambda_i p_{ij} \quad \text{y} \quad q_j(\lambda) = q_{0j} + \sum_i^m \lambda_i q_{ij},$$

la función de Lagrange puede ser escrita como

$$\begin{aligned} L = \psi(x, \lambda) + (a_0 - \zeta) z + \sum_{j=1}^n (\xi_j (\alpha_j - x_j) + \eta_j (x_j - \beta_j)) + \\ \sum_{i=1}^m (c_i y_i + \frac{1}{2} d_i y_i^2 - \lambda_i a_i z - \lambda_i y_i - \lambda_i b_i - \mu_i y_i). \end{aligned}$$

En ese caso, las condiciones KKT para el subproblema (2.36) son dadas por

$$\begin{aligned} \frac{\partial \psi}{\partial x_j} - \xi_j + \eta_j &= 0, \quad j = 1, \dots, n & (\partial L / \partial x_j = 0) & (2.37a) \\ c_i + d_i y_i - \lambda_i - \mu_i &= 0, \quad i = 1, \dots, m & (\partial L / \partial y_i = 0) & (2.37b) \\ a_0 - \zeta - \lambda^T a &= 0, & (\partial L / \partial z = 0) & (2.37c) \\ g_i(x) - a_i z - y_i - b_i &\leq 0, \quad i = 1, \dots, m & (\text{factibilidad primal}) & (2.37d) \\ \lambda_i (g_i(x) - a_i z - y_i - b_i) &= 0, \quad i = 1, \dots, m & (\text{complementaridad}) & (2.37e) \\ \xi_j (\alpha_j - x_j) &= 0, \quad j = 1, \dots, n & (\text{complementaridad}) & (2.37f) \\ \eta_j (x_j - \beta_j) &= 0, \quad j = 1, \dots, n & (\text{complementaridad}) & (2.37g) \\ -\mu_i y_i &= 0, \quad i = 1, \dots, m & (\text{complementaridad}) & (2.37h) \\ -\zeta z &= 0, & (\text{complementaridad}) & (2.37i) \\ \alpha_j \leq x_j \leq \beta_j, & \quad j = 1, \dots, n & (\text{factibilidad primal}) & (2.37j) \\ -z \leq 0, \quad -y_i \leq 0, & \quad i = 1, \dots, m & (\text{factibilidad primal}) & (2.37k) \\ \xi_j \geq 0, \quad \eta_j \geq 0, & \quad j = 1, \dots, n & (\text{factibilidad dual}) & (2.37l) \\ \zeta \geq 0, \quad \mu_i \geq 0, & \quad i = 1, \dots, m & (\text{factibilidad dual}) & (2.37m) \\ \lambda_i \geq 0, & \quad i = 1, \dots, m & (\text{factibilidad dual}) & (2.37n) \end{aligned}$$

donde

$$\frac{\partial \psi}{\partial x_j} = \frac{p_j(\lambda)}{(U_j - x_j)^2} - \frac{q_j(\lambda)}{(x_j - L_j)^2} \quad \text{y} \quad \lambda^T a = \sum_{i=1}^m \lambda_i a_i.$$

2.6.2. Condiciones de optimalidad para el subproblema

Cuando usamos un método de puntos interiores para resolver el subproblema (2.36), los ceros del lado derecho de las condiciones de complementariedad (2.37e) - (2.37i) son reemplazadas por un parámetro pequeño $\varepsilon > 0$. Además, son introducidas las variables de holgura s_i para las restricciones (2.37d). De modo que las condiciones de optimalidad son dadas por

$$\begin{aligned} \frac{\partial \psi}{\partial x_j} - \xi_j + \eta_j &= 0, \quad j = 1, \dots, n & (2.38a) \\ c_i + d_i y_i - \lambda_i - \mu_i &= 0, \quad i = 1, \dots, m & (2.38b) \\ a_0 - \zeta - \lambda^T a &= 0, & (2.38c) \\ g_i(x) - a_i z - y_i + s_i - b_i &= 0, \quad i = 1, \dots, m & (2.38d) \\ \xi_j (x_j - \alpha_j) - \varepsilon &= 0, \quad j = 1, \dots, n & (2.38e) \\ \eta_j (\beta_j - x_j) - \varepsilon &= 0, \quad j = 1, \dots, n & (2.38f) \\ \mu_i y_i - \varepsilon &= 0, \quad i = 1, \dots, m & (2.38g) \\ \zeta z - \varepsilon &= 0, & (2.38h) \\ \lambda_i s_i - \varepsilon &= 0, \quad i = 1, \dots, m & (2.38i) \\ x_j - \alpha_j > 0, \quad \xi_j > 0 & \quad j = 1, \dots, n & (2.38j) \\ \beta_j - x_j > 0, \quad \eta_j > 0 & \quad j = 1, \dots, n & (2.38k) \\ y_i > 0, \quad \mu_i > 0 & \quad i = 1, \dots, m & (2.38l) \\ z > 0, \quad \zeta > 0 & & (2.38m) \\ s_i > 0, \quad \lambda_i > 0 & \quad i = 1, \dots, m. & (2.38n) \end{aligned}$$

Para cada $\varepsilon > 0$, existe una única solución $(x, y, z, \xi, \mu, \zeta, s)$ de estas condiciones. Esto ocurre porque las condiciones (2.38) son equivalentes a las condiciones KKT de el siguiente problema estrictamente convexo en las variables x, y, z e s .

$$\begin{aligned} \text{minimizar } & g_0(x) + a_0z + \sum_i (c_i y_i + \frac{1}{2} d_i y_i^2) + \\ & -\varepsilon \sum_j \log(x_j - \alpha_j) - \varepsilon \sum_j \log(\beta_j - x_j) + \\ & -\varepsilon \sum_i \log y_i - \varepsilon \sum_i \log s_i - \varepsilon \log z \\ \text{s a } & g_i(x) - a_i z - y_i + s_i \leq b_i, \quad i = 1, \dots, m \\ & (\alpha_j < x_j < \beta_j, y_j > 0, z > 0, s_i > 0), \end{aligned}$$

donde las desigualdades estrictas son automáticamente satisfechas debido a la función objetivo presenta términos de logaritmos.

2.6.3. Una dirección de Newton para las condiciones de optimalidad

Dado un punto inicial $w = (x, y, z, \lambda, \xi, \eta, \mu, \zeta, s)$ que satisfaga (2.38j) - (2.38n) al aplicar el método de Newton al sistema de ecuaciones no lineales (2.38a) - (2.38i), el siguiente sistema de ecuaciones lineales debe ser resuelto

$$\begin{bmatrix} \Psi & & G^T & \langle -e \rangle & \langle e \rangle & & & & & & & \\ & \langle d \rangle & \langle -e \rangle & & & & \langle -e \rangle & & & & & \\ & & -a^T & & & & & -1 & & & & \\ G & \langle -e \rangle & -a & & & & & & \langle e \rangle & & & \\ \langle \xi \rangle & & & \langle x - \alpha \rangle & & & & & & & & \\ \langle -\eta \rangle & & & & \langle \beta - x \rangle & & & & & & & \\ & \langle \mu \rangle & & & & & \langle y \rangle & & & & & \\ & & \zeta & & & & z & & & & & \\ & & & \langle s \rangle & & & & & \langle \lambda \rangle & & & \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta \lambda \\ \Delta \xi \\ \Delta \eta \\ \Delta \mu \\ \Delta \zeta \\ \Delta s \end{bmatrix} = \begin{bmatrix} -\delta_x \\ -\delta_y \\ -\delta_z \\ -\delta_\lambda \\ -\delta_\xi \\ -\delta_\eta \\ -\delta_\mu \\ -\delta_\zeta \\ -\delta_s \end{bmatrix} \quad (2.39)$$

donde $\delta_x, \dots, \delta_s$ son los términos del lado izquierdo en (2.38a) - (2.38i), $\Psi \in \mathbb{R}^{n \times n}$ es la matriz diagonal con

$$(\Psi)_{jj} = \frac{\partial^2 \psi}{\partial x_j^2} = \frac{2p_j(\lambda)}{(U_j - x_j)^3} + \frac{2q_j(\lambda)}{(x_j - L_j)^3},$$

$G \in \mathbb{R}^{m \times n}$ es la matriz con componentes

$$(G)_{ij} = \frac{\partial g_i}{\partial x_j} = \frac{p_{ij}}{(U_j - x_j)^2} - \frac{q_{ij}}{(x_j - L_j)^2},$$

$\langle d \rangle$ es una matriz diagonal con el vector $d = (d_1, \dots, m)^T$ en la diagonal, $\langle x - \alpha \rangle$ es una matriz diagonal con el vector $x - \alpha$ en la diagonal y e es el vector $(1, \dots, 1)^T$, con dimensión adecuada, de modo que $\langle e \rangle$ es la matriz identidad con dimensiones adecuadas.

En el sistema de Newton (2.39), las variables $\Delta\xi, \Delta\eta, \Delta\mu, \Delta\zeta$ y Δs pueden ser eliminadas por

$$\begin{aligned}\Delta\xi &= -\langle x - \alpha \rangle^{-1} \langle \xi \rangle \Delta x - \xi + \varepsilon \langle x - \alpha \rangle^{-1} e, \\ \Delta\eta &= -\langle \beta - x \rangle^{-1} \langle \eta \rangle \Delta x - \eta + \varepsilon \langle \beta - x \rangle^{-1} e, \\ \Delta\mu &= -\langle y \rangle^{-1} \langle \mu \rangle \Delta y - \mu + \varepsilon \langle y \rangle^{-1} e, \\ \Delta\zeta &= -(\zeta/z)\Delta z - \zeta + \varepsilon/z, \\ \Delta s &= -\langle \lambda \rangle^{-1} \langle s \rangle \Delta \lambda - s + \varepsilon \langle \lambda \rangle^{-1} e.\end{aligned}$$

Así, el siguiente sistema en $(\Delta x, \Delta y, \Delta z, \Delta \lambda)$ es obtenido:

$$\begin{bmatrix} D_x & & & G^T \\ & D_y & & \langle -e \rangle \\ & & \zeta/z & -a^T \\ G & \langle -e \rangle & -a & -D_\lambda \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -\tilde{\delta}_x \\ -\tilde{\delta}_y \\ -\tilde{\delta}_z \\ -\tilde{\delta}_\lambda \end{bmatrix}, \quad (2.40)$$

donde

$$\begin{aligned}D_x &= \Psi + \langle x - \alpha \rangle^{-1} \langle \xi \rangle + \langle \beta - x \rangle^{-1} \langle \eta \rangle \quad (\text{matriz diagonal}) \\ D_y &= \langle d \rangle + \langle y \rangle^{-1} \langle \mu \rangle, \quad (\text{matriz diagonal}) \\ D_\lambda &= \langle \lambda \rangle^{-1} \langle s \rangle, \quad (\text{matriz diagonal}) \\ \tilde{\delta}_x &= \frac{\partial \psi}{\partial x} - \varepsilon \langle x - \alpha \rangle^{-1} e + \varepsilon \langle \beta - x \rangle^{-1} e, \\ \tilde{\delta}_y &= c + \langle d \rangle y - \lambda - \varepsilon \langle y \rangle^{-1} e, \\ \tilde{\delta}_z &= a_0 - \lambda^T a - \varepsilon/z, \quad (\text{escalar}) \\ \tilde{\delta}_\lambda &= g(x) - az - y - b + \varepsilon \langle \lambda \rangle^{-1} e,\end{aligned}$$

siendo $\frac{\partial \psi}{\partial x}$ el vector con las n componentes

$$\frac{\partial \psi}{\partial x_j} = \frac{p_j(\lambda)}{(U_j - x_j)^2} - \frac{q_j(\lambda)}{(x_j - L_j)^2}, \quad j = 1, \dots, n$$

En seguida, Δy puede ser eliminado del sistema (2.40) usándose

$$\Delta y = D_y^{-1} \Delta \lambda - D_y^{-1} \tilde{\delta}_y.$$

Así, el siguiente sistema en $\Delta x, \Delta \lambda$ e $\Delta \lambda$ es obtenido:

$$\begin{bmatrix} D_x & & G^T \\ & \zeta/z & -a^T \\ G & -a & -D_{\lambda y} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -\tilde{\delta}_x \\ -\tilde{\delta}_z \\ -\tilde{\delta}_{\lambda y} \end{bmatrix}, \quad (2.41)$$

donde

$$\begin{aligned}D_{\lambda y} &= D_\lambda + D_y^{-1}, \quad (\text{matriz diagonal}) \\ \tilde{\delta}_{\lambda y} &= \tilde{\delta}_\lambda + D_y^{-1} \tilde{\delta}_y.\end{aligned}$$

En el sistema (2.41), podemos eliminar Δx o $\Delta \lambda$. Si Δx es eliminado por

$$\Delta x = -D_x^{-1}G^T \Delta \lambda - D_x^{-1}\tilde{\delta}_x,$$

el siguiente sistema en $\Delta \lambda$ y Δz es obtenido:

$$\begin{bmatrix} D_{\lambda y} + GD_x^{-1}G^T & a \\ a^T & -\zeta/z \end{bmatrix} \begin{bmatrix} \Delta \lambda \\ \Delta z \end{bmatrix} = \begin{bmatrix} \tilde{\delta}_{\lambda y} - GD_x^{-1}\tilde{\delta}_x \\ \tilde{\delta}_z \end{bmatrix} \quad (2.42)$$

Si, en vez de eso, $\Delta \lambda$ es eliminado del sistema (2.41) por

$$\Delta \lambda = D_{\lambda y}^{-1}G\Delta x - D_{\lambda y}^{-1}a\Delta z + D_{\lambda y}^{-1}\tilde{\delta}_{\lambda y},$$

el siguiente sistema en Δx y Δz es obtenido

$$\begin{bmatrix} D_x + G^T D_{\lambda y}^{-1}G & -G^T D_{\lambda y}^{-1}a \\ -a^T D_{\lambda y}^{-1}G & \zeta/z + a^T D_{\lambda y}^{-1}a \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = \begin{bmatrix} -\tilde{\delta}_x - G^T D_{\lambda y}^{-1}\tilde{\delta}_{\lambda y} \\ -\tilde{\delta}_z + a^T D_{\lambda y}^{-1}\tilde{\delta}_{\lambda y} \end{bmatrix} \quad (2.43)$$

La mejor manera de resolver el sistema de ecuaciones (2.41) depende de la resolución del sistema de ecuaciones (2.42) o del sistema de ecuación (2.43). El sistema (2.42) debe ser escogido si $n > m$, mientras el sistema (2.43) debe ser escogido si $m > n$.

2.6.4. Búsqueda lineal en la dirección de Newton

Luego de haber obtenido la dirección de Newton $\Delta w = (\Delta x, \Delta y, \Delta z, \Delta \lambda, \Delta \xi, \Delta \eta, \Delta \mu, \Delta \zeta, \Delta s)$ a partir de un punto $w = (x, y, z, \lambda, \xi, \eta, \mu, \zeta, s)$, debemos calcular el tamaño de paso t en esa dirección. Para que el paso t no viole las restricciones (2.38j) - (2.38n), comenzamos la búsqueda con $t \leq 1$ tal que

$$\begin{aligned} x_j + t\Delta x_j - \alpha_j &\geq 0, 01(x_j - \alpha_j), \\ \beta_j - (x_j + t\Delta x_j) &\geq 0, 01(\beta_j - x_j), \end{aligned}$$

para todo j , y que

$$(y, z, \lambda, \xi, \eta, \mu, \zeta, s) + t\Delta w \geq 0, 01(y, z, \lambda, \xi, \eta, \mu, \zeta, s).$$

De acuerdo con Svanberg [16], el nuevo punto también debe mejorar el vector residuo $\delta(w)$ definido por el lado izquierdo de las condiciones KKT (2.38a) - (2.38i) en relación al punto anterior. Por lo tanto, tómesese τ como el mayor número de la secuencia $t, t/2, t/4, t/8, \dots$ tal que

$$\|\delta(w + \tau\Delta w)\|_2 < \|\delta(w)\|_2,$$

Observe que siempre es posible obtener este tamaño de paso τ porque la dirección de Newton es una dirección de descenso para $\|\delta(w)\|_2$.

Dados el punto $w^{(1)} = (x^{(1)}, y^{(1)}, z^{(1)}, \lambda^{(1)}, \xi^{(1)}, \eta^{(1)}, \mu^{(1)}, \zeta^{(1)}, s^{(1)})$, el parámetro $\varepsilon^{(1)}$ y el parámetro $0 < \epsilon < 1$ muy pequeño. Como ejemplo, uno puede tomar

$$\begin{aligned} \varepsilon^{(1)} &= 1, \quad x_j^{(1)} = \frac{1}{2}(\alpha_j + \beta_j), \quad y_i^{(1)} = 1, \quad z^{(1)} = 1, \quad \zeta^{(1)} = 1, \quad \lambda_i^{(1)} = 1, \quad s_i^{(1)} = 1, \\ \xi_j^{(1)} &= \max\{1, 1/(x_j^{(1)} - \alpha_j)\}, \quad \eta_j^{(1)} = \max\{1, 1/(\beta_j - x_j^{(1)})\}, \quad \mu_i^{(1)} = \max\{1, c_i/2\}. \end{aligned}$$

El algoritmo de abajo detalla el método de puntos interiores para resolver el subproblema (2.36).

Algoritmo 2: Puntos interiores primal - dual

Mientras $\varepsilon^\ell > \epsilon$, hacer

Paso 1 : Dado ε^ℓ y w^ℓ que verifique (2.38j) - (2.38n),

Calcule Δw^ℓ conforme descrito en la Sección 2.6.3.

Paso 2: Calcule el tamaño de paso τ^ℓ conforme descrito en la Sección 2.6.4.

Paso 3: $w^{\ell+1} \leftarrow w^\ell + \tau^\ell \Delta w^\ell$.

Paso 4: Si $\|\delta(w^{\ell+1})\|_\infty < 0, 9\varepsilon^\ell$, hacer

$$\varepsilon^{(\ell+1)} \leftarrow 0, 1\varepsilon^\ell$$

Caso contrario

$$\varepsilon^{(\ell+1)} \leftarrow \varepsilon^\ell.$$

$$\ell \leftarrow \ell + 1.$$

Según Svanberg [17], el problema aumentado (2.34) tiene al menos una solución óptima. Para demostrar esa afirmación, consideraremos el problema equivalente al problema (2.34) con una variable $x = (x_1, \dots, x_n)$

$$\begin{aligned} \text{Minimizar} \quad & C_0(x) + a_0 \max_{i \in A_1} \left\{ \frac{C_i^+(x)}{a_i} \right\} + \sum_{i \in A_0} \left(c_i C_i^+(x) + \frac{1}{2} d_i (C_i^+(x))^2 \right) \\ \text{s a} \quad & x \in X \end{aligned} \quad (2.44)$$

donde

$$\begin{aligned} X &= \{x \in R^n \mid \hat{L}_j \leq x_j \leq \hat{U}_j, j = 1, \dots, m\}, \\ A_1 &= \{i \in \{1, \dots, m\} \mid a_i > 0\}, \\ A_0 &= \{i \in \{1, \dots, m\} \mid a_i = 0\}, \\ C_i^+(x) &= \max\{0, C_i(x)\}. \end{aligned}$$

Esta formulación (2.44) no será usada para resolver el problema (2.34), mas será útil para demostrar la proposición.

Proposición 2.1 *Si $x \in X$ es mantenido fijo en el problema (2.34), los correspondientes valores óptimos de las variables y e z son únicas. Estos únicos valores son los siguientes:*

Se $A_1 = \emptyset$, entonces $z = 0$ e $y_i = C_i^+(x)$ para $i \in \{1, \dots, m\}$

Se $A_1 \neq \emptyset$, entonces

$$z = \max_{i \in A_1} \left\{ \frac{C_i^+(x)}{a_i} \right\}, \quad y_i = 0 \text{ para } i \in A_1 \text{ e } y_i = C_i^+(x) \text{ para } i \in A_0.$$

Si $A_1 = \emptyset$, tomando en cuenta $a_0 > 0$ y $c_i + d_i > 0, \forall i$, es fácil ver que si $x \in X$, entonces $(x, C^+(x), 0)$ es un punto viable para (2.34). Además, dado (x, y, z) un punto viable para (2.34) entonces $y_i \geq C_i^+(x), i = 1, \dots, m$, luego:

$$C_0(x) + \sum_{i=1}^m \left(c_i C_i^+(x) + \frac{d_i}{2} (C_i^+(x))^2 \right) \leq C_0(x) + a_0 z \sum_{i=1}^m \left(c_i y_i + \frac{d_i}{2} y_i^2 \right) \quad (2.45)$$

De forma semejante podemos demostrar para el caso $A_1 \neq \emptyset$ considerando $a_i c_i > a_0$ para todo $i \in A_1$.

De este modo, podemos eliminar la variables y_i y z del problema (2.34) y obtener el problema (2.44). Así, tenemos la siguiente proposición.

Proposición 2.2 *Si el par $(\hat{x}, \hat{y}, \hat{z})$ es una solución óptima global del problema (2.34) si y solo si \hat{x} es solución óptima global del problema (2.44) mientras que \hat{y} , \hat{z} son como en la Proposición 2.1*

la demostración se sigue de la Proposición 2.1

Proposición 2.3 *Existe al menos una solución óptima global del problema (2.34).*

En el problema (2.44), la función objetivo es continua sobre un conjunto compacto X . De esta forma, existe al menos una solución óptima del problema (2.44). Luego de la Proposición 2.2, tenemos que existe al menos una solución óptima del problema (2.34).

Proposición 2.4 *Si $(\hat{x}, \hat{y}, \hat{z})$ es una solución, local o global, del problema (2.34), entonces existen los multiplicadores de Lagrange que junto con $(\hat{x}, \hat{y}, \hat{z})$ satisfacen las condiciones KKT.*

Demostración. Consideremos la condición de cualificación propuesta por Arrow, Hurwicz y Uzawa [1], sea x^* un minimizador local del problema:

$$\begin{aligned} \text{mín} \quad & C_0(x) \\ \text{s. a.} \quad & C_1(x) \leq 0, \dots, C_m(x) \leq 0, \end{aligned} \quad (2.46)$$

donde C_0 y C_i son continuamente diferenciable. Si existe un vector $d \in \mathbb{R}^n$ tal que

$$\nabla C_i(x^*)^T d < 0. \quad \forall i \in A(x^*)$$

donde $A(x^*) = \{i \mid C_i(x) = 0\}$ es el conjunto de índices de las restricciones activas del problema (2.46) en x^* , entonces existen λ_i multiplicadores de Lagrange que junto con x^* verifican las condiciones KKT para problema (2.46), esas condiciones KKT consiste en resolver el siguiente sistema de ecuaciones

$$\begin{aligned} \frac{\partial C_0}{\partial x_j}(x^*) + \sum_{i=1}^m \lambda_i \frac{\partial C_i}{\partial x_j}(x^*) &= 0, \quad j = 1, \dots, n \\ C_i(x^*) &\leq 0, \quad i = 1, \dots, m \\ \lambda_i &\geq 0, \quad i = 1, \dots, m \\ \lambda_i C_i(x^*) &= 0, \quad i = 1, \dots, m \end{aligned}$$

La condición de cualificación de Arrow, Hurwicz y Uzawa [1] es un caso particular de la conocida condición de cualificación de Mangasarian-Fromovitz [12] para problema con restricciones de desigualdad, que es verificado en el problema

(2.34). En efecto, sea $(\hat{x}, \hat{y}, \hat{z})$ una solución óptima del problema (2.34) y el vector $d = (d_x, d_y, d_z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}$:

$$(d_x)_j = \begin{cases} 1, & \text{Si } \hat{x}_j = \hat{L}_j \\ -1, & \text{Si } \hat{x}_j = \hat{U}_j, \quad j = 1, \dots, n, \\ 0, & \text{Caso contrario.} \end{cases} \quad (d_y)_i = 1 + \sum_{j=1}^n \left| \frac{\partial C_i}{\partial x_j}(\hat{x}) \right|, \quad i = 1, \dots, m.$$

y $d_z = 1$. El resultado del producto interno de d y el vector gradiente, calculado en $(\hat{x}, \hat{y}, \hat{z})$, de cualquier restricción activa en el problema (2.34) es estrictamente negativo. De este modo, existe un vector multiplicador de Lagrange que junto con $(\hat{x}, \hat{y}, \hat{z})$ satisface las condiciones KKT del problema (2.34). ■

2.7. Método de gradiente espectral proyectado no monótono

En [6], los autores combinaron el método de gradiente proyectado con dos ingredientes de optimización. El primero, extendieron las estrategias de globalización típicas a estos métodos para el esquema de búsqueda lineal no monótona desarrollada por Grippo, Lampariello y Lucidi [8] para el método de Newton lo que represento una gran mejoría en relación a los métodos de gradiente proyectados tradicionales. El segundo ingrediente, propusieron la asociación del paso espectral, introducido por Barzilai y Borwein [3]. Esta elección de tamaño de paso requiere poco esfuerzo computacional y aumenta la velocidad de convergencia de los métodos de gradiente.

Paso espectral

Consideremos el siguiente problema:

$$\min_{x \in \mathbb{R}^n} C : \mathbb{R}^n \rightarrow \mathbb{R}$$

y sea el modelo cuadrático:

$$m_k(x) = C(x_k) + \nabla C(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^T B_k(x - x_k),$$

donde B_k es una aproximación de $\nabla^2 C(x_k)$. Si B_k verifica la ecuación secante:

$$B_k s = y, \tag{2.47}$$

donde $s = x_k - x_{k-1}$ y $y = \nabla C(x_k) - \nabla C(x_{k-1})$, entonces el modelo cuadrático verificara $m_k(x_k) = C(x_k)$, $\nabla m_k(x_k) = \nabla C(x_k)$ y $\nabla m_k(x_{k-1}) = \nabla C(x_{k-1})$. Para reducir costo computacional imponemos que $B_k = \omega I$. Probablemente la matriz $B_k = \omega I$ no verificará la ecuación (2.47). Entonces podemos formular el problema de calcular la matriz ωI que verifique la ecuación (2.47) en el sentido de mínimos cuadrados. De este modo, tenemos el siguiente problema:

$$\begin{aligned} \min_{\omega} \|\omega I s - y\|_2^2 &= (\omega I s - y)^T (\omega I s - y) = q(\omega) \\ q(\omega) &= \omega^2 s^T s - 2\omega s^T y + y^T y \\ q'(\omega) &= 2\omega s^T s - 2s^T y = 0 \Rightarrow \omega^* = \frac{s^T y}{s^T s} \end{aligned}$$

como $q''(\omega) = 2s^T s > 0$, pues $\|s\|_2^2 > 0$, entonces ω^* es minimizador. De este modo, como $B_k = \omega I$, tenemos la versión más popular del paso espectral de Barzilai y Borwein.

$$\alpha_k^{(1)} = (\omega^*)^{-1} = \frac{s^T s}{s^T y}.$$

Por otro lado, denotamos $H_k = B_k^{-1}$ y si H_k verifica la ecuación

$$s = H_k y, \quad (2.48)$$

entonces al imponer una condición $H_k = \theta I$ y formular el problema de encontrar la matriz de la forma θI que verifica la ecuación (2.48) en el sentido de mínimos cuadrados, tenemos:

$$\begin{aligned} \min_{\theta} \|s - \theta I y\|_2^2 &= (s - \theta I y)^T (s - \theta I y) = q(\theta) \\ q(\theta) &= \theta^2 y^T y - 2\theta s^T y + s^T s \\ q'(s) &= 2\theta y^T y - 2s^T y = 0 \Rightarrow \theta^* = \frac{s^T y}{y^T y} \end{aligned}$$

como $q''(\theta) = 2y^T y > 0$, pues $\|y\|_2^2 > 0$, entonces θ^* es minimizador. De este modo, $H_k = B_k^{-1}$ y $H_k = \theta I$, tenemos otra versión del paso espectral:

$$\alpha_k^{(2)} = \theta^* = \frac{s^T y}{y^T y}.$$

En este trabajo, utilizamos el paso espectral $\alpha_k^{(1)}$ y los salvaguardas $0 < \alpha_{\min} < \alpha_{\max} < \infty$. Así, en cada iteración tenemos el paso espectral:

$$\alpha_{k+1} = \min(\alpha_{\max}, \max(\alpha_{\min}, \frac{s_k^T s_k}{s_k^T y_k}))$$

Búsqueda Lineal

La búsqueda lineal monótona de Armijo, que exige un decrecimiento suficiente en cada iteración, puede no poseer un buen desempeño. Para amenizar esto, Grippo, Lampariello, y Lucidi [8] introdujeron la búsqueda lineal no monótona, que exige un decrecimiento en cada M iteraciones, lo que permite que la función objetivo aumente en algunas iteraciones. Estas iteraciones del método SPG son de la forma $x_{k+1} = x_k + \lambda_k d_k$ donde $d_k = P_{\Omega}(x_k - \alpha_k \nabla C(x_k)) - x_k$ y $\lambda_k > 0$ tal que x_{k+1} verifica la condición de Armijo no monótono dada por

$$C(x_{k+1}) \leq C_{\max} + \lambda_k \gamma_{arm} \nabla C(x_k)^T d_k$$

donde $C_{\max} = \max\{C(x_{k-j}) \mid 0 \leq j \leq \min\{k, M-1\}\}$. Para nuestro análisis de la convergencia del método y descripción del Algoritmo, definamos el gradiente proyectado escalado como:

$$g_t(x) = [P(x - t g(x)) - x]$$

para todo $x \in \Omega, t > 0$ donde $g(x) = \nabla C(x)$ y $P(\cdot)$ define la proyección ortogonal sobre Ω .

En el Algoritmo 3, describimos como calcular en cada iteración el paso espectral, considerando las salvaguardas, α_{\min} y α_{\max} , que evitan valores cerca de cero o

valores muy grandes, respectivamente. Supongamos que sean dados $x_0 \in \Omega$, $\epsilon > 0$, $\alpha_0 \in [\alpha_{\min}, \alpha_{\max}]$, $0 < \alpha_{\min} < \alpha_{\max} < +\infty$, $\gamma \in (0, 1)$, $M \geq 1$, $0 < \sigma_1 < \sigma_2 < 1$.

Algoritmo 3: Gradiente Espectral Proyectado no monótono

Paso 1: Detectar si el punto actual es estacionario.

Si $\|P(x_k - g(x_k)) - x_k\| = 0$, parar, declarando que x_k es estacionario.

Paso 2: Búsqueda lineal no monótona

Paso 2.1 Calcule $d_k = P(x_k - \alpha_k g(x_k)) - x_k$. Hacer $\lambda \leftarrow 1$.

Paso 2.2 Hacer $x_+ = x_k + \lambda d_k$.

Paso 2.3 Si

$$C(x_+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} C(x_{k-j}) + \gamma \lambda \langle d_k, g(x_k) \rangle, \quad (2.49)$$

entonces, definimos

$$\lambda_k = \lambda, x_{k+1} = x_+, s_k = x_{k+1} - x_k, y_k = g(x_{k+1}) - g(x_k),$$

ir al **Paso 3**.

Si (2.49) no es verificada, defina

$$\lambda_{new} \in [\sigma_1, \sigma_2 \lambda],$$

Hacer $\lambda \leftarrow \lambda_{new}$, ir al **Paso 2.2**

Paso 3. Calcule $b_k = \langle s_k, y_k \rangle$.

Si $b_k \leq 0$, hacer $\alpha_{k+1} \leftarrow \alpha_{\max}$,

caso contrario,

calcule $a_k = \langle s_k, s_k \rangle$ y $\alpha_{k+1} = \min\{\alpha_{\max}, \max\{\alpha_{\min}, a_k/b_k\}\}$.

Para obtener λ_{new} , calculamos el mínimo de la función interpoladora cuadrática de una dimensión $q(w)$ tal que $q(0) = C(x_k)$, $q(\lambda) = C(x_k + \lambda d_k)$ e $q'(0) = \nabla C(x_k)^T d_k$. Observe que el proceso de salvaguarda emplea el intervalo $[\sigma_1, \sigma_2 \lambda]$. Esto significa que, cuando λ_{new} no pertenece a ese intervalo, se juzga que la interpolación no es confiable y el proceso más conservador de bisección es utilizado.

Lema 2.1 Para todo $x \in \Omega$, $t \in (0, \max]$,

i) $\langle g(x), g_t(x) \rangle \leq -\frac{1}{t} \|g_t(x)\|_2^2 \leq -\frac{1}{\alpha_{\max}} \|g_t(x)\|_2^2$.

ii) el vector $g_t(\bar{x})$ es cero si y solo si \bar{x} es un punto estacionario.

Demostración. ver [12].

2.8. Definiciones de términos básicos

Discretizar: Dividir o particionar un todo en partes con menor complejidad, con el fin de facilitar cálculos. En el área de estadística y matemáticas el término se utiliza para expresar la individualización (discretización) de una unidad continua en unidades individuales (discretas).

Equilibrio Estático: El concepto de equilibrio estático, o equilibrio mecánico estático, es utilizado en física para describir un estado estacionario en el cuál la posición relativa de los componentes de un sistema no cambia con el tiempo o se mueven todos a una velocidad constante. Dicho en otras palabras, en el estado de equilibrio estático el sistema está en reposo o su centro de masas se mueve a velocidad constante.

Deformación: La deformación es el cambio en el tamaño o forma de un cuerpo debido a esfuerzos internos producidos por una o más fuerzas aplicadas sobre el mismo o la ocurrencia de dilatación térmica.

Módulo de Young: El módulo de Young o módulo de elasticidad longitudinal es un parámetro que caracteriza el comportamiento de un material elástico, según la dirección en la que se aplica una fuerza.

Coefficiente de Poisson: El coeficiente de Poisson es una constante elástica que proporciona una medida del estrechamiento de sección de un prisma de material elástico lineal e isótropo cuando se estira longitudinalmente y se adelgaza en las direcciones perpendiculares a la de estiramiento.

Layout: La noción de layout suele utilizarse para nombrar al esquema de distribución de los elementos dentro de un diseño.

III. HIPÓTESIS Y VARIABLES

3.1. Hipótesis general e hipótesis específicas

Hipótesis general

El método de gradiente espectral proyectado no monótono resuelve problemas de programación no lineal sobre un conjunto convexo.

Hipótesis específicas

1. El método de elementos finitos rectangulares es adecuado para dominio de forma rectangular.
2. La implementación del método SPG es eficiente para resolver el problema de la minimización de la flexibilidad de una estructura debido a que tiene una restricción lineal y restricciones de cajas.
3. El método SPG tiene convergencia global.

3.2. Operacionalización de variables

| Variable | Dimensiones | Indicadores |
|---|--|---------------------------------|
| Métodos de optimización no lineal para problemas de estructuras | Método de las asíntotas móviles | Función de aproximación convexa |
| | | Problema aumentado |
| | Método de gradiente espectral proyectado no monótono | Paso espectral |
| | | Búsqueda lineal |

IV. METODOLOGÍA DE LA INVESTIGACIÓN

4.1. Tipo de investigación

Nuestra investigación es básica. Si obtenemos la solución del problema de minimización de la flexibilidad de una estructura con el método SPG, posiblemente este método podría resolver otros problemas de optimización de estructuras.

4.2. Diseño de la investigación

En este trabajo de investigación presenta el siguiente diseño:

- Discretizar el dominio de la estructura, considerando las condiciones de contorno, en elementos finitos rectangulares.
- Formular el problema de minimización de la flexibilidad de una estructura.
- Implementar el método SPG para resolver el problema de minimización.
- Comparar los resultados obtenidos del método SPG con el método MMA.

4.3. Población y muestra

Como la investigación es básica no considera población ni muestra.

4.4. Técnicas e instrumentos de recolección de datos

Para nuestra investigación, se revisará artículos científicos, tesis y libros que abordaron el problema en cuestión.

4.5. Plan de análisis estadísticos de datos

Por la naturaleza del trabajo no se realiza ningún análisis estadístico.

V. RESULTADOS

En este capítulo, presentamos la demostración de la convergencia global del método SPG que fue propuesto por Birgin, Martinez y Raydan [6]. Además, mostramos los resultados de las implementaciones de los métodos SPG y MMA para problemas de estructuras.

5.1. Convergencia del método SPG

Demostraremos que el Algoritmo 3 está bien definido y tiene la propiedad de convergencia global, todo punto de acumulación \bar{x} es un punto estacionario, o sea

$$\langle g(\bar{x}), x - \bar{x} \rangle \geq 0 \quad \text{para todo } x \in \Omega.$$

Teorema 5.1 *El algoritmo SPG esta bien definido, y cualquier punto de acumulación de la secuencia $\{x_k\}$ que es generada es un punto estacionario.*

Demostración.

Supongamos que x_k no es un punto estacionario, entonces del Lema 2.1

$$\langle g(x_k), d_k \rangle = \langle g(x_k), g_{\alpha_k}(x_k) \rangle \leq -\frac{1}{\alpha_{\text{máx}}} \|g_{\alpha_k}(x_k)\|_2^2 < 0,$$

y la dirección de búsqueda es una dirección de descenso. Así, encontraremos un paso en un número finito de iteraciones, y el Algoritmo 3 está bien definido.

Sea $\bar{x} \in \Omega$ un punto de acumulación de la secuencia $\{x_k\}$, y consideremos una subsecuencia, con la misma notación de la secuencia anterior, que convergen a \bar{x} . Analicemos dos casos:

Caso 1: Asumimos que $\inf \lambda_k = 0$. Supongamos que \bar{x} no es un punto estacionario. Por la continuidad y compacidad, existe $\delta > 0$ tal que

$$\left\langle g(\bar{x}), \frac{P(\bar{x} - \alpha g(\bar{x})) - \bar{x}}{\|P(\bar{x} - \alpha g(\bar{x})) - \bar{x}\|} \right\rangle < -\delta \quad \text{para todo } \alpha \in [\alpha_{\text{mín}}, \alpha_{\text{máx}}].$$

Esto implica que

$$\left\langle g(x_k), \frac{P(x_k - \alpha g(x_k)) - x_k}{\|P(x_k - \alpha g(x_k)) - x_k\|} \right\rangle < -\delta/2 \quad \text{para todo } \alpha \in [\alpha_{\text{mín}}, \alpha_{\text{máx}}]. \quad (5.50)$$

y k suficientemente grande en la secuencia que converge a \bar{x} . Desde que $\inf \lambda_k = 0$, existe una secuencia $\{x_k\}_K$ tal que

$$\lim_{k \in K} \lambda_k = 0.$$

en ese caso, la manera en que λ_k se elige en (2.49), existe un índice \bar{k} suficientemente grande tal que para todo $k \geq \bar{k}$, $k \in K$, existe ρ_k , $0 < \sigma_1 \leq \rho_k \leq \sigma_2$, para el cual $\lambda_k/\rho_k > 0$ la condición (2.49) no es verificada, o sea:

$$C(x_k + \frac{\lambda_k}{\rho_k} d_k) > \max_{0 \leq j \leq M-1} C(x_{k-j}) + \gamma \frac{\lambda_k}{\rho_k} \langle g(x_k), d_k \rangle \geq C(x_k) + \gamma \frac{\lambda_k}{\rho_k} \langle g(x_k), d_k \rangle.$$

Luego,

$$\frac{C(x_k + \frac{\lambda_k}{\rho_k} d_k) - C(x_k)}{\lambda_k/\rho_k} > \gamma \langle g(x_k), d_k \rangle.$$

por el teorema del valor medio, la relación de arriba podemos escribirlo como:

$$\langle g(x_k + t_k d_k) \rangle > \gamma \langle g(x_k), d_k \rangle \quad \text{para todo } k \in K, k \geq \bar{k}, \quad (5.51)$$

donde t_k es un escalar en el intervalo $[0, \lambda_k/\rho_k]$ que va para cero cuando $k \in K$ va para el infinito. Por otro lado, la secuencia $d_k/\|d_k\|$ es limitada, así posee una secuencia convergente a d . reemplazando esa subsecuencia en la ecuación (5.51) y tomando el límite, tenemos que $(1 - \gamma) \langle g(\bar{x}), d \rangle \geq 0$. (Observe que la secuencia $\{\|d_k\|\}_K$ es limitada, y entonces $t_k \|d_k\| \rightarrow 0$). Como $(1 - \gamma) > 0$ y $\langle g(x_k), d_k \rangle < 0, \forall k$, entonces $\langle g(\bar{x}), d \rangle = 0$. Por la continuidad y definición de d_k , tenemos que para k suficientemente grande sobre la subsecuencia,

$$\langle g(x_k), \frac{P(x_k - \alpha_k g(x_k)) - x_k}{\|P(x_k - \alpha_k g(x_k)) - x_k\|} \rangle > -\delta/2,$$

esto contradice a desigualdad (5.50).

Caso 2: Asuma que $\inf \lambda_k \geq \rho > 0$. Supongamos que el punto \bar{x} no sea estacionario. Entonces, del Lema 2.1, tenemos que $\|g_\lambda(\bar{x})\| > 0$ para todo $\lambda \in (0, \alpha_{\max}]$. Por la continuidad y compacidad, existe un $\delta > 0$ tal que $\|g_\lambda(\bar{x})\| \geq \delta > 0$ para todo $\lambda \in [\rho, \alpha_{\max}]$. Sea $l(k)$ un entero tal que $k - \min\{k, M-1\} \leq l(k) \leq k$ y

$$C(x_{l(k)}) = \max_{0 \leq j \leq \min\{k, M-1\}} C(x_{k-j}).$$

es una secuencia monótona no creciente. En efecto,

$$\begin{aligned} C(x_{l(k+1)}) &= \max_{0 \leq j \leq \min\{k+1, M-1\}} C(x_{k+1-j}) \\ &\leq \max_{0 \leq j \leq \min\{k+1, M-1\}} C(x_{k+1-j}) = \max\{C(x_{l(k)}), C(x_{k+1})\} = C(x_{l(k)}). \end{aligned}$$

De (2.49), se sigue que para $k > M-1$,

$$\begin{aligned} C(x_{l(k)}) &\leq C(x_{l(k)-1} + \lambda_{l(k)-1} d_{l(k)-1}) \\ &\leq \max_{0 \leq j \leq \min\{l(k)-1, M-1\}} C(x_{l(k)-1-j}) + \gamma \lambda_{l(k)-1} \langle g(x_{l(k)-1}), g_{\alpha_{l(k)-1}}(x_{l(k)-1}) \rangle \\ &= C(x_{l(l(k)-1)}) + \gamma \lambda_{l(k)-1} \langle g(x_{l(k)-1}), g_{\alpha_{l(k)-1}}(x_{l(k)-1}) \rangle. \end{aligned}$$

Por la continuidad, para $k \geq \bar{k}$ suficientemente grande, $\|g_{\alpha_k}(x_k)\| \geq \delta/2$. Entonces, usando el lema 2.1, tenemos

$$C(x_{l(k)}) \leq C(x_{l(l(k)-1)}) - \frac{\gamma \rho}{\alpha_{\max}} \|g_{\alpha_{l(k)-1}}(x_{l(k)-1})\|_2^2 \leq C(x_{l(l(k)-1)}) - \frac{\gamma \delta^2 \rho}{4 \alpha_{\max}}.$$

Cuando $k \rightarrow \infty$, claramente $C(x_{l(k)}) \rightarrow -\infty$, lo cual es una contradicción. En realidad, C es una función continua y así $C(x_k)$ converge a $C(\bar{x})$. ■

5.2. Resultados numéricos

En esta sesión mostramos los resultados de las implementaciones del método de las asíntotas móviles y el método de gradiente espectral proyectado no monótono que fueron aplicados a los problemas clásicos de optimización topológica como la viga en voladizo, la viga *Messerschmitt-Bölkow-Blohm* (MBB) y la estructura de puente. Cada uno de estos problemas de OT fueron modelados con el método SIMP. En todos los problemas, los dominios fueron discretizados en elementos finitos rectangulares de 4 nodos, con dos grados de libertad.

La metodología aquí propuesta y todas las rutinas necesarias para la aplicación del método de los elementos finitos fueron implementados en Matlab. Todos los resultados fueron obtenidos en un ordenador Intel CORE 2 2.40 GHz, con 4 GB de memoria RAM.

De la Figura 5.6, *la viga en voladizo* tiene un apoyo (engaste), impidiendo los desplazamientos nodales tanto en la horizontal como en la vertical, así como una fuerza $F = 1N$ aplicada en el punto inferior de la arista lateral derecha. La longitud y ancho del dominio son, respectivamente, 300 cm y 150 cm. El dominio es discretizado en $300 \times 150 = 45000$ elementos finitos.

La viga MBB es apoyada en los cantos inferiores, de tal forma que apenas los desplazamientos horizontales son permitidos en el canto derecho, y los desplazamientos horizontales y verticales del canto izquierdo son impedidos. Una carga externa $F = 1N$ es aplicada para abajo en el centro de borde superior. El largo y ancho del dominio son, respectivamente, 840 cm y 140 cm. Debido a la simetría, consideramos apenas la mitad derecha de la estructura, que es discretizada en $420 \times 140 = 58800$ elementos finitos.

La estructura del puente es apoyada en los cantos inferiores, de tal forma que apenas los desplazamientos horizontales son permitidos en el canto derecho, y los desplazamientos horizontales y verticales del canto izquierdo son impedidos. La estructura tiene una fuerza de magnitud $2F$ aplicada para abajo en el centro del borde inferior y dos fuerzas de magnitud F aplicadas en el borde inferior, con distribución simétrica. En nuestros experimentos, usamos $F = 1N$. El largo y ancho del dominio son, respectivamente, 480 cm y 240 cm. Debido a la simetría, consideramos apenas la mitad de la estructura, que es discretizada en $240 \times 240 = 57600$ elementos finitos.

Los radios del filtro para la viga en voladizo, viga MBB y estructura de puente son 12,6 , 10,5 y 9,9, veces la longitud de la arista de los elementos, respectivamente. Además, definimos un punto de inicio viable como $x_e = 0,4, \forall e$. Tener en cuenta que el filtro de media ponderada \tilde{x} se aplica sólo en la función objetivo y en la restricción del volumen.

En todos los problemas, supondremos que la estructura tiene 1 cm de espesura, que el coeficiente de Poisson es igual a 0,3 y que el módulo de Young del material

es $E = 1 \text{ N/cm}^2$, representado en la topología óptima con el color negro, el módulo de Young artificial $E_0 = 10^{-3}$ es representado en la topología óptima con el color blanco, las estructuras deben contener, como máximo, el 40% del volumen total del dominio, el volumen fraccionario del material que la compone es $\eta = 0,4$ y el número máximo de iteraciones es 100.

Los valores de los parámetros del algoritmo de SPG son $\lambda_{\min} = 10^{-30}$, $\lambda_{\max} = 10^{30}$, $\sigma_1 = 0,1$, $\sigma_2 = 0,9$, $\gamma_{arm} = 10^{-4}$ y $M = 10$. Las Tablas 5.2 y 5.3 muestran los resultados de las implementaciones de los métodos MMA y SPG, y las estructuras obtenidas se presentan en las Figuras 5.7 y 5.8, respectivamente.

El criterio de parada de los algoritmos es basado en el decrecimiento relativo de la función objetivo $\frac{|\mathbf{C}(x^{(k+1)}) - \mathbf{C}(x^{(k)})|}{|\mathbf{C}(x^{(k+1)})|} \leq 5 \times 10^{-5}$ por tres veces seguidas.

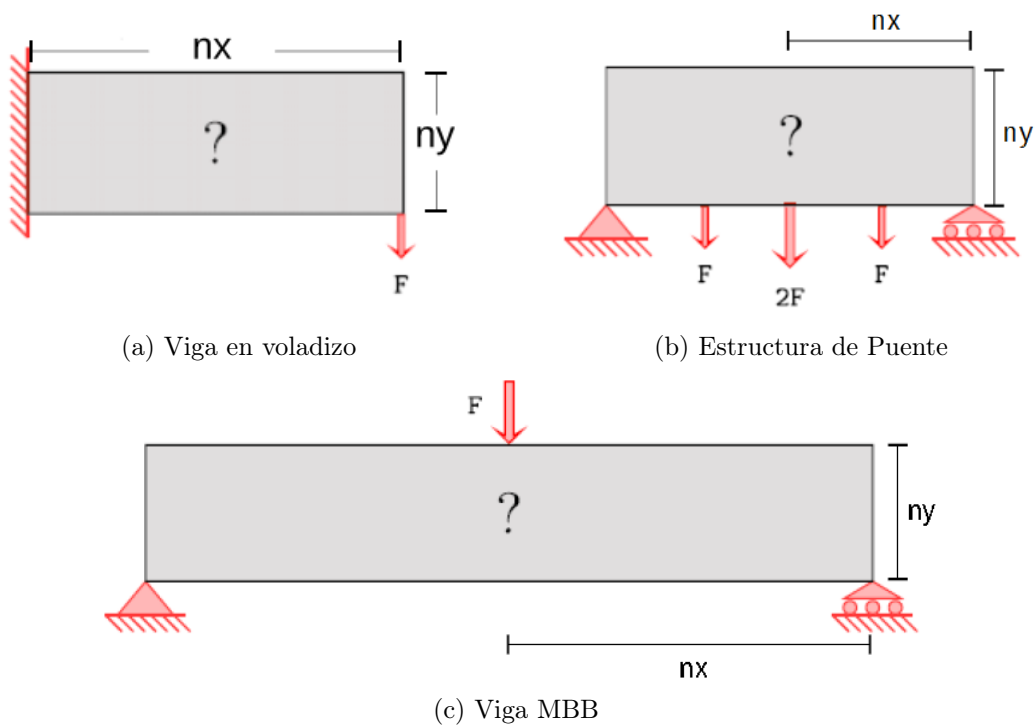


Figura 5.6: Problemas de optimización topológica

Tabla 5.2: Resultados de los Problemas de OT usando el método MMA.

| Problemas | Radio de Filtro | Número de Iteraciones | Función Objetivo | Tiempo (seg) |
|----------------------|-----------------|-----------------------|------------------|--------------|
| Viga en Voladizo | 12.6 | 100 | 112.15211 | 531.6382 |
| Viga MBB | 10.5 | 100 | 276.82527 | 736.4496 |
| Estructura de Puente | 9.9 | 100 | 304.44583 | 687.4628 |

Tabla 5.3: Resultados dos Problemas de OT usando el método SPG

| Problemas | Radio de Filtro | Número de Iteraciones | Función Objetivo | Tiempo (seg) |
|----------------------|-----------------|-----------------------|------------------|--------------|
| Viga en Voladizo | 12.6 | 76 | 111.65831 | 183.2663 |
| Viga MBB | 10.5 | 88 | 277.17903 | 266.9556 |
| Estructura de Puente | 9.9 | 31 | 302.49632 | 93.2882 |

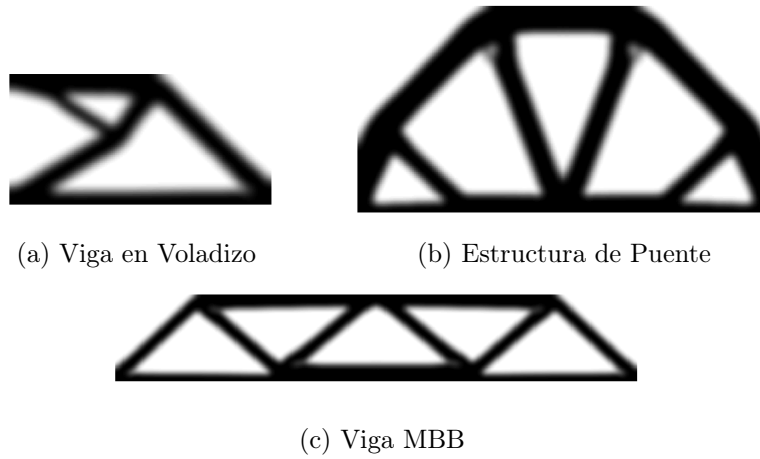


Figura 5.7: Estructuras Óptimas (MMA)

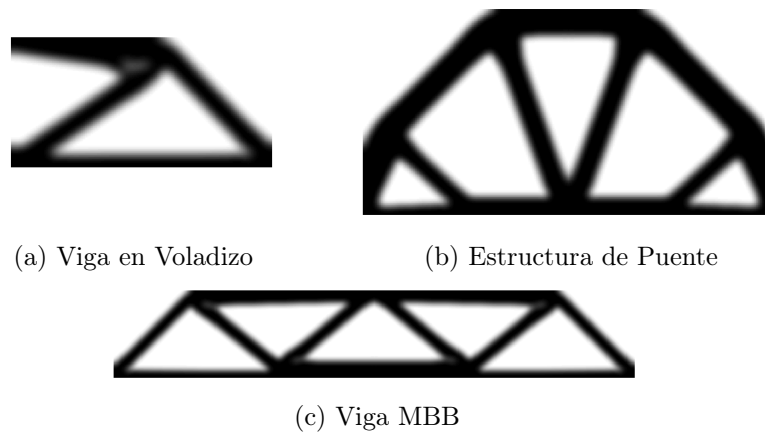


Figura 5.8: Estructuras Óptimas (SPG)

VI. CONCLUSIONES

En este trabajo, implementamos el método SIMP en conjunto con el método SPG para resolver el problema de minimización de la flexibilidad de una estructura con restricciones de volumen y variables de gran escala. En el análisis de los resultados, se verificó que el método SPG tuvo un desempeño superior que el método MMA, obteniendo estructuras similares en menor tiempo.

VII. RECOMENDACIONES

Como estudios posteriores a este trabajo, sugerimos:

- Usar otro lenguaje de programación, como C_{++} o Fortran, con el objetivo de reducir el tipo total gastado en la obtención de las soluciones óptimas.
- Obtener topologías óptimas de otros problemas de optimización topológica, como por ejemplo Mecanismos Flexibles, para obtener conclusiones sobre el desempeño de los métodos de optimización abordados en este trabajo.

VIII. REFERENCIAS BIBLIOGRÁFICAS

- [1] Arrow K. J., and Uzama H. *Constraint qualifications in maximization problems. Naval Research logistic Quartely*,8:175-191 1961.
- [2] Assan A. E., *Métodos dos elementos finitos: Primeiros passos. Editora Unicamp, são paulo, 2 edition, 2003.*
- [3] Barzilai J. and Borwein J. M., *Two point step gradient method. IMA J. Number. Anal.*, 8:141-148, 1988.
- [4] Bendøe M. P., *Optimal shape design as a material distribution problem. Struct. Optim.*, 1:193-202, 1989
- [5] Bertsekas, D. P. *Nonlinear Programming, 2 3d. athena Sientific, Belmont MA, 1999.*
- [6] Birgin E. G., Martinez J. M. and Raydan M., *Nonmonotone spectral projected gradient methods on convex sets, SIAM J. optim.*,v. 10, p.1196-1211,2000.
- [7] Diaz A. R., Sigmund O., *Checkerboard pattern in layout optimization, Struct. Optim.*, 10:40-45,1995.
- [8] Grippo L., Lampariello F. and Lucidi S., *A monotone line search technique for newton's method. SIAM J. on Number. Anal.*, 23:707-716, 1986.
- [9] Goodier T. *Teoria da Elasticidade. Guanabara Dois, Rio de Janeiro,3 edition, 1980.*
- [10] Haftka R. T., and Gurdal Z., *Elements of structural optimization, 3rd edn. Kluwer Academic Publishers, Boston, 1992*
- [11] Haftka, R. T., and Grandhi, R. V., *Structural shape optimization - a survey. Computer Methods in Applied Mechanics and Engineering* 57 (1986), 91-106.
- [12] Izmailov A. and Solodov M., *Otimização - volume 1: Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade. IMPA, Rio de Janeiro, 2 edition, 2007.*
- [13] Izmailov A. and Solodov M., *Otimização - volume 2: Métodos Computacionais. IMPA, Rio de Janeiro, 2 edition, 2007.*
- [14] Rietz A., *Sufficiency of a finite exponent in simp (power law) Methods. Struct. Multidisc. Optim.*, v. 21, p. 159-163, 2001.

- [15] Svanberg K., *The method of moving asymptotes-a new method for structural optimization. Int. J. Num. Meth. Eng.*, 24:359-373, 1987.
- [16] Svanberg K. *MMA and GCMMA - two method nonlinear optimization. Technical report 2007.*
- [17] Svanberg K. *A class of globally convergent optimization methods based on conservative convex separable approximations, SIAM J. Opt.*, vol. 12, no. 2, pp. 555-573, 2002.
- [18] Vanderplaats, G. N., *Numerical optimization techniques for engineering design: with applications. McGraw - Hill, New York, 1984*
- [19] Martinez J. M. and Santos S. A., *Métodos Computacionais de Otimização. IMPA, Rio de Janeiro, 1995.*

ANEXO

Matriz de consistencia

| PROBLEMA | OBJETIVO | HIPÓTESIS | METODOLOGÍA |
|---|---|--|--|
| <p>Problema General ¿Es posible adaptar el método de gradiente espectral proyectado no monótono para resolver problemas de optimización estructural?</p> | <p>Objetivo General Resolver el problema de minimización de la flexibilidad de una estructura usando el método de gradiente espectral proyectado no monótono.</p> | <p>Hipótesis General El método de gradiente espectral proyectado no monótono resuelve, problemas de programación no lineal sobre conjunto convexo.</p> <p>Hipótesis específicas 1) El método de elementos finitos rectangulares es adecuado para el dominio de forma rectangular.</p> <p>2) La implementación del método SPG es eficiente para resolver el problema de la minimización de la flexibilidad de una estructura debido a que tiene una restricción lineal y restricciones de cajas.</p> <p>3) El método SPG tiene convergencia global.</p> | <p>Tipo de investigación Nuestra investigación es básica. Si obtenemos la solución del problema de la minimización de la flexibilidad de una estructura con el método SPG, posiblemente este método podría resolver otros problemas de optimización de estructuras.</p> <p>Diseño 1) Discretizar el dominio de la estructura, considerando las condiciones de contorno, en elementos finitos rectangulares. 2) Formular el problema de la minimización de la flexibilidad de una estructura. 3) Implementar el método SPG para resolver el problema de minimización. 4) Compara los resultados obtenidos del método SPG con el método MMA.</p> |
| <p>Problema específico ¿Es posible resolver el problema de la minimización de la flexibilidad de un estructura con el método SPG? ¿Es posible que el método SPG obtenga la solución del problema de minimización en menor tiempo que el método de MMA?</p> | <p>Objetivo específico 1) Estudiar la convergencia del método SPG. 2) Implementar el algoritmo del método SPG en conjunto con el método de los elementos finitos para resolver el problema de minimización. 3) Analizar los resultados numéricos del método propuesto.</p> | <p>2) La implementación del método SPG es eficiente para resolver el problema de la minimización de la flexibilidad de una estructura debido a que tiene una restricción lineal y restricciones de cajas.</p> <p>3) El método SPG tiene convergencia global.</p> | <p>1) Discretizar el dominio de la estructura, considerando las condiciones de contorno, en elementos finitos rectangulares. 2) Formular el problema de la minimización de la flexibilidad de una estructura. 3) Implementar el método SPG para resolver el problema de minimización. 4) Compara los resultados obtenidos del método SPG con el método MMA.</p> |