

Métodos secantes de cambio mínimo para la solución de sistemas de ecuaciones con restricciones

por

María de los Ángeles Martínez Arraigada

Presentado ante la Facultad de Matemática, Astronomía, Física y Computación como parte de los requerimientos para la obtención del grado de

Doctora en Matemática

de la

Universidad Nacional de Córdoba



Marzo 2019

©FAMAF – UNC 2019

Director: Dr. Damián Fernández Ferreyra



Métodos secantes de cambio mínimo para la solución de sistemas de ecuaciones con restricciones por María de los Ángeles Martínez Arraigada se distribuye bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

*...of course, a mathematician has to be in tune with
some of the basics of a mathematical way of life,
such as pleasure in spending hours in quiet contemplation,
and in dedication to writing projects.
But we all know that this somewhat solitary
side of mathematical life also brings
with it a kind of social life that few people outside of
our professional world can even imagine.
The frequent travel that's not just tied to a few laboratories,
the network of friends and research collaborators
in different cities and even different countries,
the extended family of former students, and the interactions
with current students—what fun, and what an opportunity to
explore music, art, nature, and our many other interests.
All these features keep me going too..”*

R. Tyrrell Rockafellar

Agradecimientos

Lo que parece ser el final, suele ser en realidad un nuevo comienzo...

Quiero agradecer en primer lugar a mi director Damián Fernández, quien me acompañó en este recorrido con gran dedicación y paciencia. Ninguno de mis logros hubiera sido posible sin su apoyo. A Cristina Turner, una mujer maravillosa, llena de vitalidad y entusiasmo.

Agradezco también a la Facultad de Matemática, Astronomía, Física y Computación, por ser mi segundo hogar durante estos años. A la Universidad Nacional de Córdoba, pública, gratuita y laica. A CONICET por brindarme la oportunidad de realizar este doctorado.

Gracias a mi familia que me apoya de manera incondicional, especialmente a Rosario y María.

Gracias a mis compañeros, que con pequeñas charlas de pasillo o café de por medio, hicieron que mi estadía en la facultad sea más amena. A todos los que pasaron por la oficina 418, gracias por compartir el día a día.

A Federico y Fiorela, mis primeros amigos de la facultad, gracias por tantos años de una amistad intacta, a pensar de la distancia. A Raquel, una personita que ocupa un lugar especial en mi corazón, gracias por tantas charlas y consejos. A Melina, por seguir transitando juntas este último trayecto. A Alexander, porque a pesar del tiempo y la distancia, siempre hay un punto de encuentro.

Gracias a Melina, los caminos de la vida nos encontraron y espero nos sigan encontrando juntas, gracias por tu amistad incondicional. A Eugenia y Lucia, personas maravillosas de las cuales he aprendido y aprendo día a día.

A Ezequiel, por su compañía, sus palabras no siempre acertadas y sus malos consejos. Gracias por no siempre decir lo que uno quiere escuchar. A Gabriel, por sus consejos como hermano académico y amigo. A Fernando, por las interminables charlas de pasillo, comentarios políticos y resúmenes de charlas TEDx, se que algún día escucharé alguna.

A Valentin, a quien quiero con el alma y admiro, porque si hay personas que nos marcan en esta vida, sin duda el es una de ellas. A Marianne, por hacerme sentir como una hija. Un pedacito de mí siempre seguirá allá... Vielen Dank für alles!

A mi amigo Javier, una persona incondicional, mi colombiano favorito. Gracias por estar siempre conmigo, por darme hermosos consejos y dejarme volar, te quiero mucho.

A mi mejor amigo y hermano, Luciano, aunque el tiempo pase, nuestra amistad sigue intacta.

Por último, a mis mujeres de danza, gracias por curarme el alma, no me imagino una vida sin bailar con ustedes, sin sentirme libre...

Resumen

En este trabajo comenzamos con un recorrido por algunos de los métodos newtonianos más tradicionales de la literatura, para resolver problemas de optimización y sistemas de ecuaciones no lineales. Estos problemas surgen en distintas áreas de la ciencia, la ingeniería y la industria. En las últimas décadas, el requerimiento en cada uno de estos ámbitos ha desencadenado un mayor estudio y desarrollo sobre diferentes herramientas para tratar tales problemas.

En el primer capítulo comenzamos dando algunos conceptos básicos que son clave a la hora del análisis de los métodos que presentamos a lo largo de este trabajo. Derivamos el método clásico de Newton para la solución de sistemas de ecuaciones, presentamos el algoritmo básico y un resultado de convergencia del mismo bajo ciertas hipótesis. Adicionamos algunos comentarios relacionados a los métodos de Newton inexactos, los cuales son una herramienta útil a la hora de encarar problemas de gran escala. Además, derivamos este método para problemas de optimización sin restricciones y restricciones mixtas, haciendo una breve introducción a las condiciones de optimalidad y regularidad de las funciones involucradas. Terminamos este capítulo introduciendo el método LP-Newton, propuesto por Facchinei, Fischer y Herrich que permite resolver sistemas de ecuaciones no lineales con restricciones, y posibles soluciones localmente no aisladas, lo que lo hace altamente competitivo con otros ya existentes en la literatura que presentan algunas falencias a la hora de resolver, por ejemplo, sistemas KKT provenientes de problemas de optimización, desigualdades variacionales, etc.

En el segundo capítulo presentamos un método basado en el método LP-Newton, donde desarrollamos una estrategia quasi-Newton para aproximar la información de primer orden de la función por una matriz que utiliza valores funcionales ya calculados. Este método combina las propiedades de las actualizaciones quasi-Newton, que evitan el cálculo de derivadas que pueden ser costosas o peor aún, que no estén analíticamente disponibles, y del método LP-Newton, capaz de tratar con restricciones de los sistemas y el hecho que las soluciones sean no aisladas. Presentamos un análisis de convergencia local del método y mostramos que posee velocidad de convergencia q -lineal. Algunos ejemplos numéricos para mostrar el desempeño del algoritmo son dados al finalizar el capítulo.

En el tercer capítulo analizamos un nuevo método que intenta mejorar la velocidad de convergencia del propuesto en el capítulo anterior, alcanzando convergencia q -superlineal. En este caso, algunas hipótesis extras son necesarias en relación a las matrices quasi-

Newton pero solo son requeridas para el análisis teórico. En la última sección presentamos algunos experimentos numéricos donde estos supuestos no son verificados y el algoritmo igualmente realiza una buena performance.

En el último capítulo presentamos las conclusiones generales del trabajo y algunas discusiones son planteadas.

Palabras claves: sistema de ecuaciones no lineales con restricciones, soluciones no aisladas, métodos quasi-Newton, convergencia local superlineal.

2010 Mathematics subject Classification: 90C30, 65K05.

Índice general

1. Introducción	1
1.1. Método de Newton	2
1.1.1. Introducción al Análisis Local	3
1.1.2. Método de Newton para Sistemas de Ecuaciones	4
1.1.3. Método de Newton para Problemas de Optimización	7
1.1.3.1. Optimización sin Restricciones	9
1.1.3.2. Optimización con Restricciones de Igualdad	11
1.1.3.3. Optimización con Restricciones Mixtas	15
1.2. Métodos Secantes	20
1.2.1. Métodos Secantes para Sistemas de Ecuaciones	21
1.2.2. Métodos Secantes para Problemas de Optimización	23
1.3. Un Método tipo Newton para la Resolución de Sistemas de Ecuaciones	25
2. Un Método LP quasi-Newton Modificado	33
2.1. Introducción	33
2.2. Análisis de Convergencia Local	35
2.3. Ejemplos Numéricos	45
2.3.1. Comentarios	54
3. Un Método LP-Newton Secante Modificado	57
3.1. Introducción	57
3.2. Análisis de Convergencia	58
3.3. Ejemplos Numéricos	66
3.3.1. Comentarios	71
4. Conclusiones Generales	73
Bibliografía	75

Acrónimos

BFGS	Broyden-Fletcher-Goldfarb-Shanno
CQ	Constraint Qualification Condición de regularidad
DFP	Davidon-Fletcher-Powell
KKT	Karush-Kuhn-Tucker
LICQ	Linear Independence Constraint Qualification Condición de regularidad de independencia lineal
LPN	Linear Programming Newton method Método LP-Newton
MFCQ	Mangasarian-Fromovitz Constraint Qualification Condición de regularidad de Mangasarian-Fromovitz
qNLP	quasi-Newton Linear Programming method Método LP quasi-Newton
SMFCQ	Strict Mangasarian-Fromovitz Constraint Qualification Condición de regularidad de Mangasarian-Fromovitz estricta
SMLPN	Secant Modified Linear Programming Newton method Método LP-Newton Secante Modificado
SONC	Second-Order Necessary optimality Condition Condiciones de optimalidad necesarias de segundo orden
SOSC	Second-Order Sufficient optimality Condition Condiciones de optimalidad suficientes de segundo orden
SQP	Sequential Quadratic Programming Programación cuadrática secuencial

Notación

Espacios

\mathbb{R}	Espacio real unidimensional
\mathbb{R}^n	Espacio real n -dimensional
\mathbb{R}_+^n	Ortante no negativo en el espacio real n -dimensional
$\mathbb{R}^{n \times m}$	Espacio de matrices reales $n \times m$

Vectores

z	Vector columna en \mathbb{R}^n
$y^T x$	Producto interno euclídeo de vectores $x, y \in \mathbb{R}^n$
$\ z\ $	Norma vectorial en \mathbb{R}^n o \mathbb{R}^m y su norma asociada inducida en $\mathbb{R}^{m \times n}$
$\{z^k\}$	Sucesión de elementos $\{z^1, z^2, \dots, z^k, \dots\}$
$\{z^k\} \rightarrow z$	La sucesión $\{z^k\}$ converge a z

Matrices

M^T	Traspuesta de la matriz M
$\text{rang}(M)$	Número de columnas linealmente independientes de la matriz M
M^{-1}	Inversa de una matriz cuadrada M no singular
$\ M\ _*$	Norma de M en $\mathbb{R}^{m \times n}$ asociada a un producto interno
$\langle M, M \rangle$	Producto interno asociado a la norma $\ M\ _*$

Conjuntos

$ S $	Cardinal de un conjunto S
$\mathcal{B}_\varepsilon(z)$	Bola cerrada con centro en $z \in \mathbb{R}^n$ y radio ε dada por $\{w \in \mathbb{R}^n \mid \ w - z\ \leq \varepsilon\}$
$\mathcal{B}_\varepsilon^*(M)$	Bola cerrada con centro en $M \in \mathbb{R}^{n \times m}$ y radio ε dada por $\{N \in \mathbb{R}^{n \times m} \mid \ N - M\ _* \leq \varepsilon\}$
$\mathcal{C}(z)$	Cono crítico de un problema de optimización en z
$\mathcal{I}(z)$	Conjunto de índices de las restricciones de desigualdad activas en z
\mathcal{Z}	Conjunto solución de un problema

Mapeos

$\text{dist}(z, S)$	Distancia euclídea de un punto $z \in \mathbb{R}^n$ a un conjunto $S \subset \mathbb{R}^n$, definida como $\inf_{y \in S} \ y - z\ $
$F'(z)$	Matriz Jacobiana de $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ en el punto z
$\nabla f(z)$	Vector gradiente de $f : \mathbb{R}^n \rightarrow \mathbb{R}$, definido como $f'(z)^T$
$\nabla^2 f(z)$	Matriz Hessiana de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ en el punto z
$\mathcal{L}(z, \lambda, \mu)$	Función Lagrangiana de un problema de optimización en (z, λ, μ)

Capítulo 1

Introducción

Los métodos tipo Newton se encuentran entre las herramientas más utilizadas para resolver distintos tipos de problemas que surgen en la práctica, tales como sistemas de ecuaciones no lineales, problemas de minimización y de cuadrados mínimos no lineales, entre otros. Bajo ciertas hipótesis ellos son matemáticamente equivalentes, pero pueden ser tratados de manera separada para poder explotar sus diferentes estructuras. Tales problemas tienen origen en diversas áreas de investigación, tales como economía, física, química, ingeniería, etc. Abarcando desde simples problemas no lineales de una variable, hasta aquellos de más de cientos o miles de variables, que son costosos de resolver en términos computacionales. A continuación mencionamos algunos ejemplos que pueden servir de motivación al lector.

- ✓ El siguiente es un problema de control extraído de [66, Ejemplo 11.1]. El desafío es analizar la estabilidad de una aeronave en respuesta a los comandos del piloto. Para ello se planteó un modelo simplificado basado en el balance de fuerzas. Las ecuaciones de equilibrio para dicha nave están dadas por un sistema de 5 ecuaciones con 8 incógnitas, de la forma $F(z) = Az + \phi(z)$, donde $F : \mathbb{R}^8 \rightarrow \mathbb{R}^5$, $A \in \mathbb{R}^{5 \times 8}$ y $\phi : \mathbb{R}^8 \rightarrow \mathbb{R}^5$ es la parte no lineal de la ecuación. El objetivo es encontrar un punto $z \in \mathbb{R}^8$ tal que $F(z) = 0$. Las primeras variables z_1 , z_2 y z_3 representan las tasas de balanceo, inclinación y guiñada, respectivamente, mientras que z_4 es el ángulo de ataque incremental y z_5 el ángulo de deslizamiento lateral. Las últimas 3 variables z_6 , z_7 y z_8 son los controles y representan las desviaciones del elevador, alerón y timón, respectivamente. Entonces, para cierta elección de las variables de control, obtenemos un sistema de 5 ecuaciones con 5 incógnitas. Si deseamos estudiar el comportamiento de la nave al cambiar los controles, debemos resolver un sistema de ecuaciones no lineales con incógnitas z_1, \dots, z_5 para cada ajuste de controles.
- ✓ Otro problema interesante en el área de biología molecular, es determinar la geometría de una molécula. La idea es calcular las posiciones de los átomos en el espacio que minimizan el potencial de energía asociado. Si n es el número de átomos y tomamos $z_i \in \mathbb{R}^3$ como la posición espacial del i -ésimo átomo, el vector

$z = (z_1, \dots, z_n) \in \mathbb{R}^{3n}$ está asociado a la energía potencial $F(z)$. Dicha energía es la suma de diferentes términos, entre ellos los referidos a longitud de enlace, energía, ángulos de balance, etc. Esta función es fuertemente no lineal e involucra más de 10^3 variables. Para más detalles del problema y la función involucrada, ver [7, Ejemplo 1.2.1].

Estos son solo dos de los numerosos ejemplos con los que podemos encontrarnos. En algunas ocasiones los problemas involucran pocas variables, en cuyo caso no son difíciles de resolver, sin embargo, podría ocurrir que las evaluaciones funcionales sean costosas. Por otro lado, puede que la función involucrada no venga dada por una expresión algebraica, sino que sus valores provengan de ciertos experimentos, simulaciones o cajas negras, en cuyo caso sus derivadas analíticas no están disponibles, aunque la función sea lo suficientemente suave. De aquí la motivación de desarrollar algoritmos efectivos en ausencia de expresiones analíticas, tanto de la función como de sus derivadas.

En este capítulo presentamos una breve descripción de algunos de los métodos newtonianos más conocidos en la literatura para la solución de sistemas de ecuaciones y problemas de optimización. Comenzamos definiendo algunos conceptos básicos para facilitar la interpretación de los posteriores resultados. En la Sección 1.1 presentamos el método clásico de Newton y continuamos con los llamados métodos secantes en la Sección 1.2. Finalmente, en la Sección 1.3 exhibimos un método para la solución de sistemas de ecuaciones con restricciones, el cual cubre una brecha importante entre los métodos ya existentes y la necesidad de resolver ciertos problemas de manera eficiente y aprovechando su estructura.

1.1. Método de Newton

En 1669 Issac Newton (1643–1727) en su paper titulado *De Analysi per Equationes Numero Terminorum Infinitas* describe, entre otras cosas, un procedimiento iterativo para aproximar, con la precisión deseada, las raíces reales de una ecuación polinomial de tercer grado. En 1690 Joseph Raphson (1648-1715) en su libro titulado *Analysis Aequationum Universalis* propone un procedimiento similar para resolver ecuaciones más generales. Si bien los métodos eran equivalentes (aunque de espíritu diferente), Raphson desconocía el trabajo de Newton. Por esta razón, el método es comúnmente llamado de Newton-Raphson.

Finalmente, fue Thomas Simpson (1710-1761) quien en 1740 estableció el método en la forma en que hoy lo conocemos, para una ecuación no necesariamente polinomial. También se le atribuye la versión multidimensional del método de Newton.

Para comentarios históricos relacionados a este método, invitamos al lector a ver [24].

A continuación presentamos algunos conceptos básicos del análisis de convergencia que son indispensables a la hora de estudiar el comportamiento de algoritmos tipo Newton.

1.1.1. Introducción al Análisis Local

Los métodos numéricos que presentaremos en esta tesis son métodos iterativos, es decir, intentan resolver el problema dado mediante aproximaciones sucesivas a la solución, comenzando desde algún punto inicial. En otras palabras, dado un punto $z^0 \in \mathbb{R}^n$, en cada iteración k el algoritmo genera un punto z^k de manera que la sucesión de iterados $\{z^k\}$ tienda a alguna solución z^* . Además estos métodos son locales, es decir, el punto de inicio debe estar suficientemente cerca de la solución.

Otra cuestión importante sobre estos algoritmos es la tasa (velocidad) a la que los iterados z^k convergen a una solución. En general, el estudio de la tasa de convergencia es una herramienta clave para elegir un algoritmo por sobre otro, a la hora de resolver algún problema en particular. Si bien hay varios enfoques para cuantificar la tasa de convergencia, en este trabajo nos concentramos en el llamado análisis local.

Con análisis local nos referimos al comportamiento local del método en un entorno de la solución z^* , ignorando lo que sucede fuera de dicho entorno. A continuación damos algunas definiciones referidas a la tasa de convergencia de sucesiones.

Definición 1.1.1. Dada $\|\cdot\|$ una norma vectorial, sean $z^* \in \mathbb{R}^n$ y $\{z^k\} \subseteq \mathbb{R}^n$ una sucesión. Decimos que z^k tiende a z^* y lo denotamos por $z^k \rightarrow z^*$, si

$$\lim_{k \rightarrow \infty} \|z^k - z^*\| = 0.$$

Definimos algunas de las tasas de convergencia más utilizadas en el análisis local de los algoritmos.

✓ Decimos que la sucesión converge q -linealmente a z^* , si existe $\sigma \in (0, 1)$ tal que

$$\limsup_{k \rightarrow \infty} \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|} \leq \sigma.$$

✓ Decimos que la convergencia es q -superlineal a z^* si

$$\limsup_{k \rightarrow \infty} \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|} = 0.$$

✓ Decimos que la sucesión converge q -cuadráticamente a z^* si

$$\limsup_{k \rightarrow \infty} \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|^2} < \infty.$$

El coeficiente q se debe al término *quotient* (cociente). No es difícil ver que cualquier sucesión que converge q -cuadráticamente también converge q -superlinealmente, y si converge q -superlinealmente también lo hace q -linealmente. Las recíprocas no son ciertas.

Podemos definir tasas de convergencia mayores (cúbica, cuártica, etc), pero en la práctica, las que hemos mencionado son las más comunes. Además de la q -convergencia, podemos mencionar la r -convergencia.

- ✓ Decimos que una sucesión converge r -linealmente a z^* si

$$\|z^k - z^*\|^{\frac{1}{k}} \leq \infty$$

para todo k .

- ✓ La sucesión converge r -superlinealmente a z^* si

$$\|z^k - z^*\|^{\frac{1}{k}} \rightarrow 0.$$

El coeficiente r se debe al término *root* (raíz). Para más información sobre tasas de convergencia ver [67, 7, 2].

Antes de comenzar haremos algunos comentarios sobre la notación que aparecerá a lo largo de este trabajo. Usamos $\|\cdot\|$ para denotar una norma vectorial en \mathbb{R}^n o \mathbb{R}^m según corresponda y para su norma inducida en $\mathbb{R}^{m \times n}$ (siempre quedará claro por contexto). Definimos la bola cerrada centrada en un punto $z \in \mathbb{R}^n$ de radio $r > 0$ como $\mathcal{B}_r(z) = \{w \mid \|z - w\| \leq r\}$ y la distancia entre punto y conjunto como $\text{dist}(w, \mathcal{Z}) = \inf_{z \in \mathcal{Z}} \|z - w\|$. En lo que sigue, usaremos $\|\cdot\|_*$ para una norma en $\mathbb{R}^{m \times n}$ asociada a un producto interno y $\mathcal{B}_r^*(M)$ para la bola en dicha norma, es decir $\mathcal{B}_r^*(M) = \{N \mid \|M - N\|_* \leq r\}$.

1.1.2. Método de Newton para Sistemas de Ecuaciones

El método de Newton clásico surge a partir del problema de encontrar las raíces de una función F dada, es decir, encontrar un punto $z \in \mathbb{R}^n$ tal que

$$F(z) = 0, \tag{1.1}$$

donde $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es continuamente diferenciable. En esta subsección presentamos el método de Newton para la solución del problema (1.1) y discutimos sus requerimientos, implementación y características de convergencia.

Este método puede ser derivado al asumir que contamos con una buena estimación w de alguna solución z^* de (1.1) y que en un entorno de w , la función F puede ser aproximada por los dos primeros términos de su expansión en serie de Taylor, es decir

$$F(w + d) \approx F(w) + F'(w)d.$$

Por lo tanto, si $F'(w)$ es no singular en w (donde $[F'(w)]_{ij} = \frac{\partial F_i}{\partial x_j}(w)$ denota la matriz Jacobiana de F en el punto w), tomando $F(w + d) \approx 0$, podemos resolver el sistema

$$F(w) + F'(w)d = 0.$$

Dicho de otro modo, resolver el sistema anterior es encontrar $d \in \mathbb{R}^n$ solución de

$$F'(w)d = -F(w) \tag{1.2}$$

y obtener así una nueva aproximación de la solución dada por $z = w + d$. A continuación mostramos las iteraciones del método para $n = 1$.

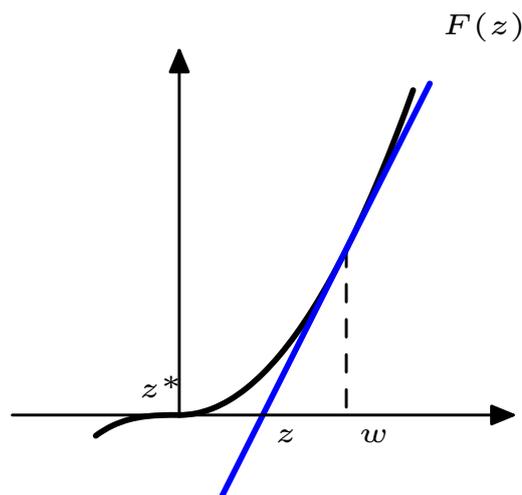


Figura 1.1: Método de Newton.

Formalmente describimos el algoritmo de Newton para resolver el problema (1.1).

Algoritmo 1

- Paso 0. Elegir $z^0 \in \mathbb{R}^n$ y fijar $k = 0$.
- Paso 1. Definir $z^{k+1} = z^k + d^k$ con d^k solución del problema $F'(z^k)d = -F(z^k)$.
- Paso 2. Fijar $k = k + 1$ y volver al Paso 1.

A continuación presentamos un resultado que asegura la convergencia local q -cuadrática (o q -superlineal bajo hipótesis relajadas) del método de Newton, ver [42, Teorema 3.2.1]

La demostración de este teorema y posteriores en este capítulo no son mostrados, por ello algunas referencias son dadas al lector.

Teorema 1.1.2. Sea $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ diferenciable en un entorno de un punto $z^* \in \mathbb{R}^n$, con derivada continua en este punto. Sea z^* una solución de (1.1) con $F'(z^*)$ no singular.

Entonces para cualquier punto inicial $z^0 \in \mathbb{R}^n$, suficientemente próximo a z^* , el Algoritmo 1 genera una sucesión de iterados $\{z^k\}$ que converge a z^* . La tasa de convergencia de este método es q -superlineal, pero si además pedimos que la derivada de la función sea Lipschitz continua en un entorno de z^* , la tasa es q -cuadrática.

Bajo las hipótesis de este teorema, se tiene que z^* es una solución aislada de (1.1), es decir, en un entorno de z^* este punto es única solución del problema.

En general, las cuestiones de existencia y unicidad están más allá de la capacidad de los algoritmos que resuelven estos problemas no lineales y no serán considerados en esta tesis. Por lo cual siempre supondremos la existencia de al menos una solución de los problemas de interés.

Este método funciona muy bien cuando el problema está cerca de ser lineal, dado que para una función F afín (no singular), encuentra la solución en una sola iteración. Si $F'(z^*)$ es singular, el teorema anterior no asegura convergencia y en caso de llegar a una solución, la tasa alcanzada no necesariamente será q -cuadrática (en estos casos la tasa suele ser q -lineal). Por otro lado, es importante tener en cuenta que la prueba de convergencia del método requiere que el punto inicial z^0 esté suficientemente cerca de una solución. En los siguientes ejemplos mostramos que la sucesión generada por el algoritmo puede no converger a una solución si alguna de las hipótesis del Teorema no se satisface.

Ejemplo 1.1.3. Consideremos el problema de encontrar un cero de la función $F(z) = z^2$, que tiene una única solución en $z^* = 0$, donde $F'(z^*)$ es singular.

Comenzando desde cualquier punto inicial z^0 , el Algoritmo 1 genera la siguiente sucesión de iterados

$$z^k = \frac{1}{2^k} z^0,$$

la cual converge a la solución, pero la tasa de convergencia es q -lineal, puesto que

$$\lim_{k \rightarrow \infty} \frac{\|z^{k+1} - z^*\|}{\|z^k - z^*\|^r} = \frac{1}{2},$$

para $r = 1$, pero para $r > 1$ dicho límite diverge.

En el siguiente ejemplo mostramos lo importante que es comenzar cerca de la solución. Este problema fue extraído de [37, Ejemplo 2.8].

Ejemplo 1.1.4. Consideremos la función $F: \mathbb{R} \rightarrow \mathbb{R}$ dada por

$$F(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}.$$

Como antes, el problema consiste en encontrar un cero de dicha función. Es fácil ver que $z^* = 0$ es la única solución de nuestro problema.

Si comenzamos con $z^0 = 1$, el algoritmo converge a $z^* = -2.35e-13$, un punto próximo a la solución.

En cambio, si comenzamos con $z^0 = 1.1$, el último iterado antes de llegar a un overflow numérico es $z^* = -2.30e+4$ y $F(z^*) = 1$, lo que significa que la sucesión diverge.

En ciertas ocasiones, especialmente cuando hablamos de implementaciones computacionales, si el problema a tratar es de tamaño considerable, en lugar de intentar encontrar una solución exacta en el Paso 1 del Algoritmo 1, se itera hasta que alguna tolerancia sea alcanzada, es decir, dado $\varepsilon > 0$, buscamos d solución de $\|F(z) + F'(z)d\| \leq \varepsilon$. Este tipo de esquemas puede conducir a métodos más prácticos, con menor costo computacional por iteración y con tasa de convergencia aún suficientemente buena. Entre estos esquemas, podemos encontrar los llamados Métodos de Newton perturbados, y una clase particular de éstos, los métodos de Newton truncados, que fueron estudiados sistemáticamente en [20]. Algunas implementaciones específicas de los métodos truncados pueden

también ser encontradas en [66, Capítulo 11]. La idea en estos casos es resolver el sistema (1.2) de manera inexacta, es decir, dado un punto z^k , se define $d \in \mathbb{R}^n$ como solución de

$$\|F(z^k) + F'(z^k)d\| \leq \sigma_k \|F(z^k)\|, \quad (1.3)$$

donde $\sigma_k \in [0, 1)$. La sucesión $\{\sigma_k\}$, conocida como sucesión forzante¹, puede estar definida de antemano o ser calculada en cada iteración.

En este caso, bajo las hipótesis del Teorema 1.1.2, la sucesión $\{z^k\}$ generada bajo la regla (1.3) converge q -superlinealmente a una solución z^* si $\sigma_k \rightarrow 0$.

En la Tabla 1.1 presentamos un resumen con las principales ventajas y desventajas del método de Newton, las cuales han sido un incentivo a la hora de desarrollar nuevos algoritmos que conserven sus bondades y mejoren los puntos más débiles del método.

Ventajas

- ✓ Convergencia q -cuadrática desde un punto inicial z^0 suficientemente próximo a z^* donde $F'(z^*)$ es no singular.
- ✓ Solución exacta en una iteración para funciones afines.

Desventajas

- ✓ No es globalmente convergente.
 - ✓ Requiere el cálculo de $F'(z^k)$ en cada iteración.
 - ✓ En cada iteración se resuelve un sistema de ecuaciones lineales que puede ser singular o mal condicionado.
-
-

Tabla 1.1: Ventajas y desventajas del método de Newton.

En las siguientes subsecciones presentamos el método de Newton para resolver problemas de optimización.

1.1.3. Método de Newton para Problemas de Optimización

Maximizar las ganancias en una empresa o minimizar los riesgos, diseñar un puente minimizando su ancho o maximizando su longitud, definir una rutina de vuelo empleando el menor tiempo posible o cantidad de combustible, son problemas que surgen con frecuencia en el campo de las ciencias y la ingeniería. El continuo desarrollo de herramientas capaces de resolver este tipo de problemas se desprende de la necesidad de tomar mejores decisiones en los diferentes ámbitos de trabajo. En esta subsección hacemos un breve recorrido por diferentes modelos que plasman matemáticamente este tipo de problemas.

Comenzamos estudiando problemas de optimización sin restricciones, con restricciones de igualdad y finalmente con restricciones mixtas, es decir, de igualdad y desigualdad. Damos además una breve introducción a las condiciones de optimalidad, las cuales se tornan visiblemente útiles al momento de considerar los algoritmos de optimización, ya que

¹En inglés, forcing sequence.

frecuentemente proveen las bases para el desarrollo y análisis de los mismos. En particular, los algoritmos reconocen soluciones al chequear si se satisfacen las condiciones de optimalidad y terminan cuando tales condiciones valen de manera satisfactoria.

Aclaremos que los conceptos de la teoría de optimización que discutiremos a lo largo de estas secciones, son los necesarios para el posterior análisis de los métodos numéricos. Para mayor información sobre el tema, invitamos al lector a revisar la siguiente bibliografía [3, 8, 7].

El problema de optimización consiste en determinar un punto que minimice una dada función objetivo $f : \mathcal{D} \rightarrow \mathbb{R}$ sobre un conjunto factible $\mathcal{D} \subseteq \mathbb{R}^n$. Matemáticamente lo denotamos de la siguiente forma

$$\begin{aligned} & \underset{z}{\text{minimizar}} && f(z) \\ & \text{s. a} && z \in \mathcal{D}. \end{aligned} \tag{1.4}$$

El problema (1.4) se dice factible si \mathcal{D} es no vacío y un punto $z \in \mathcal{D}$ se dice factible para el problema (1.4). A continuación damos la definición de minimizador local de una función f .

Definición 1.1.5. *Un punto z^* es un minimizador local de f si $f(z^*) \leq f(z)$ para todo z en un entorno de z^* .*

Si además $f(z^) < f(z)$ para todo z en ese entorno (con $z \neq z^*$), entonces se dice que z^* es un minimizador local estricto.*

Casos de particular importancia son aquellos donde, dado \mathcal{D} poliedral, es decir $\mathcal{D} = \{z \in \mathbb{R}^n \mid Az \leq b, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m\}$, f es una función lineal o cuadrática. En este caso (1.4) es un problema de programación lineal o de programación cuadrática, respectivamente.

Las condiciones de optimalidad que veremos a continuación están asociadas al concepto de estacionariedad.

Definición 1.1.6. *Un punto $z^* \in \mathbb{R}^n$ es un punto estacionario del problema general (1.4) (o punto crítico de f) si $-\nabla f(z^*) \in \mathcal{N}_{\mathcal{D}}(z^*)$, donde el cono normal a \mathcal{D} en $z^* \in \mathcal{D}$ es*

$$\mathcal{N}_{\mathcal{D}}(z^*) = \left\{ v \in \mathbb{R}^n \left| \begin{array}{l} \forall \{z^k\} \subset \mathcal{D} \setminus \{z^*\} \text{ con } z^k \rightarrow z^* \\ \text{vale que } \limsup_{k \rightarrow +\infty} \frac{v^T(z^k - z^*)}{\|z^k - z^*\|} \leq 0 \end{array} \right. \right\}.$$

A continuación presentamos un gráfico que muestra el cono normal a un conjunto en un punto dado.

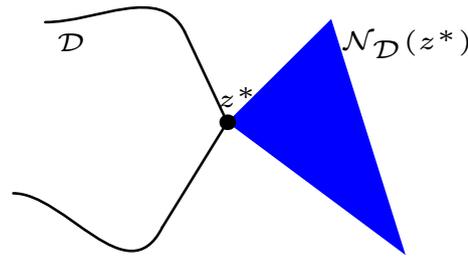


Figura 1.2: Cono normal al conjunto \mathcal{D} en el punto z^* .

1.1.3.1. Optimización sin Restricciones

Consideremos el problema (1.4) con $\mathcal{D} = \mathbb{R}^n$, es decir,

$$\begin{aligned} & \underset{z}{\text{minimizar}} && f(z) \\ & \text{s. a} && z \in \mathbb{R}^n, \end{aligned} \tag{1.5}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función suave en \mathbb{R}^n . Este problema es llamado problema de optimización sin restricciones.

Antes de presentar el algoritmo que nos permita encontrar un minimizador del problema, damos las condiciones de optimalidad necesarias que debe cumplir un punto z^* que sea minimizador local de (1.5). El resultado siguiente fue tomado de [43, Teoremas 1.7 y 1.8].

Teorema 1.1.7. *Sea z^* un minimizador local del problema (1.5), donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función diferenciable en z^* . Entonces,*

$$\nabla f(z^*) = 0. \tag{1.6}$$

Además, si f es dos veces diferenciable en dicho punto,

$$\nabla^2 f(z^*) \text{ es semidefinida positiva,} \tag{1.7}$$

donde $\nabla^2 f(z^*)$ es la matriz Hessiana de f en z^* .

La ecuación (1.6) es llamada condición necesaria de primer orden y fue formulada por Fermat en 1637 en su trabajo titulado *Methodus as Disquirendam Maximam et Minimam*. Tal condición indica que el punto z^* debe ser un punto estacionario de (1.5) ya que es fácil verificar que $\mathcal{N}_{\mathcal{D}}(z^*) = \{0\}$.

La condición (1.7) es la condición necesaria de segundo orden (SONC).

Si f es convexa, las condiciones necesarias de primer orden son suficientes para asegurar la optimalidad de un punto. En casos generales, donde f no es convexa, otras condiciones son requeridas. A continuación presentamos las condiciones de optimalidad suficientes de segundo orden (SOSC) que debe cumplir un punto z^* para ser minimizador local. El resultado fue tomado de [43, Teorema 1.9].

Teorema 1.1.8. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función dos veces diferenciable en un punto z^* . Supongamos que dicho punto satisface

$$\nabla f(z^*) = 0 \quad \text{y} \quad \nabla^2 f(z^*) \quad \text{es definida positiva.}$$

Entonces, z^* es un minimizador local estricto de f .

Además, existen constantes positivas γ y ε tales que

$$f(z) \geq f(z^*) + \frac{\gamma}{2} \|z - z^*\|^2$$

para todo z con $\|z - z^*\| < \varepsilon$.

Ahora estamos en condiciones de presentar el método de Newton para la resolución del problema (1.5). Por el Teorema 1.1.7, un minimizador local z^* de (1.5) debe cumplir $\nabla f(z^*) = 0$. Este problema es una ecuación del tipo (1.1), con lo cual podemos usar el método descrito anteriormente para su resolución.

Procediendo como antes, tenemos que para una estimación w de alguna solución z^* , la función puede ser aproximada por el siguiente modelo cuadrático

$$f(w + d) \approx f(w) + \nabla f(w)^T d + \frac{1}{2} d^T \nabla^2 f(w) d. \quad (1.8)$$

Una condición necesaria para que d sea un minimizador de la aproximación cuadrática en (1.8), es que su gradiente sea igual a cero, es decir,

$$\nabla f(w) + \nabla^2 f(w) d = 0.$$

Entonces, tenemos que

$$\nabla^2 f(w) d = -\nabla f(w). \quad (1.9)$$

Por lo tanto, el método de Newton para un problema de optimización sin restricciones es el siguiente.

Algoritmo 2

- Paso 0. Elegir $z^0 \in \mathbb{R}^n$ y fijar $k = 0$.
- Paso 1. Definir $z^{k+1} = z^k + d^k$ con d^k solución del problema $\nabla^2 f(z^k) d = -\nabla f(z^k)$.
- Paso 2. Fijar $k = k + 1$ y volver al Paso 1.

El siguiente resultado resume las propiedades de convergencia local del Algoritmo 2, ver [42, Corolario 3.2.2].

Teorema 1.1.9. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función dos veces diferenciable en un entorno de z^* , con derivada segunda continua en dicho punto. Sea z^* un punto estacionario del problema (1.5) y supongamos que $\nabla^2 f(z^*)$ es definida positiva.

Entonces, para cualquier punto inicial z^0 , suficientemente próximo a z^* , el Algoritmo 2 genera una sucesión de puntos $\{z^k\}$ que converge a z^* . Además, la tasa de convergencia es q -superlineal y si $\nabla^2 f$ es Lipschitz continua en un entorno de z^* , la convergencia es q -cuadrática.

Cabe destacar que bajo las hipótesis del resultado anterior, se tiene que z^* es un minimizador local de (1.5).

Notemos que si la condición suficiente de segundo orden se reemplaza por $\nabla^2 f(z^*)$ no singular, el Teorema 1.1.9 continúa valiendo, pero en este caso, el método no distingue minimizadores locales de otros puntos estacionarios del problema, como maximizadores locales o puntos de ensilladura. En este sentido las condiciones de optimalidad juegan un rol esencial en el desarrollo de métodos numéricos para la solución de problemas de optimización.

El Algoritmo 2 no es más que aplicar el método de Newton para sistemas de ecuaciones a ∇f . Desde esta perspectiva, estos métodos comparten las ventajas y desventajas mencionadas en la Tabla 1.1.

Para un análisis donde el método converge a un punto singular ver [18, 41, 19].

1.1.3.2. Optimización con Restricciones de Igualdad

Cuando el conjunto $\mathcal{D} \subsetneq \mathbb{R}^n$, se dice que (1.4) es un problema de optimización con restricciones, las cuales están dadas en general por igualdades y desigualdades. Si tenemos en cuenta esta estructura, obtenemos una colección de condiciones de optimalidad que involucra variables auxiliares llamadas multiplicadores de Lagrange. Estas nuevas variables facilitan la caracterización de soluciones óptimas.

La teoría de multiplicadores de Lagrange puede ser abordada de varias formas, pero aquí la introduciremos desde un punto de vista de penalización. A grandes rasgos, cada problema con restricciones será reformulado como un problema penalizado sin restricciones. Este enfoque es muy simple y sienta las bases para importantes algoritmos.

Consideremos el siguiente problema de optimización con restricciones de igualdad,

$$\begin{aligned} & \underset{z}{\text{minimizar}} && f(z) \\ & \text{s. a} && h(z) = 0, \end{aligned} \tag{1.10}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ son mapeos suaves. Aquí, $\mathcal{D} = \{z \in \mathbb{R}^n \mid h(z) = 0\}$.

A continuación damos algunas condiciones de regularidad (CQ)² sobre la función h en z^* , las cuales nos aseguran la existencia de multiplicadores de Lagrange para el problema en cuestión.

Hay un amplio abanico de condiciones de regularidad, pero aquí solo haremos mención de las más tradicionales para problemas como (1.10).

²En inglés, Constraint Qualification.

Definición 1.1.10. Sea $z \in \mathbb{R}^n$.

- ✓ *Condición de linealidad de la función restricción. Esta condición vale si h es afín, es decir,*

$$h(z) = Az + b, \quad (1.11)$$

con $A \in \mathbb{R}^{p \times n}$ y $b \in \mathbb{R}^p$.

- ✓ *Condición de independencia lineal. Decimos que la condición de regularidad de independencia lineal (LICQ)³ vale en z si $\nabla h_1(z), \dots, \nabla h_p(z)$ son linealmente independientes. Es decir,*

$$\text{rang}(\nabla h(z)) = p. \quad (1.12)$$

Bajo alguna de estas condiciones se puede ver que dado $z \in \mathbb{R}^n$ podemos definir el cono normal a D en tal punto, como

$$\mathcal{N}_{\mathcal{D}}(z) = \left\{ \sum_{i=1}^p \lambda_i \nabla h_i(z) \mid \lambda \in \mathbb{R}^p \right\}.$$

Cuando el cono normal puede escribirse en términos de los gradientes de las restricciones, se conoce como condición de regularidad de Guignard. Existen otras condiciones de regularidad que no se mencionan en este trabajo, pero todas garantizan esta última propiedad.

El siguiente enunciado, extraída de [43, Teoremas 1.11 y 1.12] establece las condiciones necesarias para el problema (1.10).

Teorema 1.1.11. *Sea f diferenciable en z^* y h diferenciable en un entorno de z^* , con derivada continua en ese punto. Sea z^* un minimizador local de f sujeto a $h(z^*) = 0$. Supongamos que vale (1.11) o (1.12) en z^* . Entonces existe un vector $\lambda^* = (\lambda_1^*, \dots, \lambda_p^*)$, llamado vector multiplicador de Lagrange, tal que*

$$\nabla f(z^*) + \sum_{i=1}^p \lambda_i^* \nabla h_i(z^*) = 0. \quad (1.13)$$

Además, si f y h son dos veces diferenciables en z^ , con h diferenciable en un entorno de z^* y su derivada continua en dicho punto y vale (1.12) en z^* , entonces para el único multiplicador de Lagrange λ^* se cumple*

$$y^T \left(\nabla^2 f(z^*) + \sum_{i=1}^p \lambda_i^* \nabla^2 h_i(z^*) \right) y \geq 0, \quad (1.14)$$

para todo $y \in \mathcal{C}(z^*) = \{y \mid \nabla h_i(z^*)^T y = 0, \quad i = 1, \dots, p\}$.

La ecuación (1.13) es llamada condición necesaria de primer orden y (1.14) condición necesaria de segundo orden (SONC).

³En inglés, Linear Independence Constraint Qualification.

Vale destacar que un multiplicador de Lagrange asociado a un punto estacionario de (1.10) es único si y solo si la condición (1.12) vale en tal punto. Si no vale ninguna CQ podrían no existir multiplicadores de Lagrange para un mínimo local, como nos muestra el siguiente ejemplo.

Ejemplo 1.1.12. Consideremos el siguiente problema extraído de [43, Ejemplo 3.1.1].

$$\begin{aligned} & \underset{z}{\text{minimizar}} && z_1 + z_2 \\ & \text{s. a} && (z_1 - 1)^2 + z_2^2 - 1 = 0, \\ & && (z_1 - 2)^2 + z_2^2 - 4 = 0. \end{aligned}$$

No es difícil ver que el único punto factible y por lo tanto mínimo local es $z^* = (0, 0)^T$. En este punto, $\nabla f(z^*) = (1, 1)^T$ no puede ser expresado como combinación lineal de los gradientes de las restricciones, ya que $\nabla h_1(z^*) = (-2, 0)^T$ y $\nabla h_2(z^*) = (-4, 0)^T$. Entonces no existen λ_1^* y λ_2^* tal que

$$\nabla f(z^*) + \lambda_1^* \nabla h_1(z^*) + \lambda_2^* \nabla h_2(z^*) = 0.$$

Algunas veces es conveniente escribir las condiciones necesarias directamente en términos de la función Lagrangiana, la cual es definida a continuación.

Definición 1.1.13. La función Lagrangiana asociada al problema (1.10) es $\mathcal{L} : \mathbb{R}^{n+p} \rightarrow \mathbb{R}$ dada por

$$\begin{aligned} \mathcal{L}(z, \lambda) &= f(z) + \lambda^T h(z) \\ &= f(z) + \sum_{i=1}^p \lambda_i h_i(z). \end{aligned}$$

Entonces, si el punto z^* es un minimizador local satisfaciendo las condiciones de la Proposición 1.1.11, tales condiciones junto con la restricción $h(z^*) = 0$ pueden ser escritas de la siguiente forma

$$\begin{aligned} \nabla_z \mathcal{L}(z^*, \lambda^*) &= 0, \\ \nabla_\lambda \mathcal{L}(z^*, \lambda^*) &= 0 \end{aligned}$$

y

$$y^T \nabla_{zz}^2 \mathcal{L}(z^*, \lambda^*) y \geq 0 \quad \text{para todo } y \in \mathcal{C}(z^*).$$

Las condiciones de primer orden representan un sistema de $n + p$ ecuaciones con $n + p$ incógnitas. Cada minimizador local z^* junto con su multiplicador asociado serán soluciones del sistema. Sin embargo, una solución de dicho sistema podría no corresponder a un minimizador local y si a un maximizador local o punto silla. Para reducir el abanico de posibilidades y garantizar que un punto es efectivamente minimizador local, haremos uso de las condiciones suficientes de segundo orden (SOSC), las cuales son dadas a continuación.

Teorema 1.1.14. Sean f y h dos veces diferenciables en el punto z^* . Sean $z^* \in \mathbb{R}^n$ y $\lambda^* \in \mathbb{R}^p$ tal que

$$\nabla_z \mathcal{L}(z^*, \lambda^*) = 0, \quad \nabla_\lambda \mathcal{L}(z^*, \lambda^*) = 0$$

y

$$y^T \nabla_{zz}^2 \mathcal{L}(z^*, \lambda^*) y > 0 \quad \text{para todo } y \in \mathcal{C}(z^*) \setminus \{0\}.$$

Entonces z^* es un minimizador local estricto de f sujeto a $h(z) = 0$.

Además, existen constantes positivas γ y ε tales que

$$f(z) \geq f(z^*) + \frac{\gamma}{2} \|z - z^*\|^2$$

para todo z factible con $\|z - z^*\| < \varepsilon$.

El resultado anterior fue tomado de [43, Teorema 1.13], donde se usa la notación $\nabla h(z) = h'(z)^T$.

Una vez planteadas las condiciones de optimalidad para el problema (1.10), procedemos a derivar el método de Newton. Por la Proposición 1.1.11 tenemos que un minimizador z^* de (1.10) debe cumplir $\nabla \mathcal{L}(z^*, \lambda^*) = 0$, con λ^* su multiplicador asociado, donde

$$\nabla \mathcal{L}(z^*, \lambda^*) = \begin{bmatrix} \nabla f(z^*) + \nabla h(z^*) \lambda^* \\ h(z^*) \end{bmatrix}.$$

Procediendo como en el caso anterior, debemos resolver la ecuación (1.9), entonces dado (w, λ_w) una aproximación a la solución, encontrar $d^\lambda = (d, \lambda_d)$ tal que

$$\nabla^2 \mathcal{L}(w, \lambda_w) d^\lambda = -\nabla \mathcal{L}(w, \lambda_w), \quad (1.15)$$

donde

$$\nabla^2 \mathcal{L}(w, \lambda_w) = \begin{bmatrix} \nabla_{zz}^2 \mathcal{L}(w, \lambda_w) & \nabla h(w) \\ \nabla h^T(w) & 0 \end{bmatrix}.$$

Puede verse que el sistema lineal (1.15) corresponde a la condición necesaria de primer orden del problema

$$\begin{aligned} \text{minimizar}_d \quad & \nabla f(w)^T d + \frac{1}{2} d^T \nabla_{zz}^2 \mathcal{L}(w, \lambda_w) d \\ \text{s. a} \quad & h(w) + \nabla h(w)^T d = 0. \end{aligned}$$

El método de Newton para el sistema de Lagrange está descrito por el Algoritmo 3, que es conocido como método de Newton-Lagrange.

Algoritmo 3

Paso 0. Elegir $(z^0, \lambda_0) \in \mathbb{R}^{n \times p}$ y fijar $k = 0$.

Paso 1. Definir $(z^{k+1}, \lambda_{k+1}) = (z^k, \lambda_k) + (d^k, \lambda_{d^k})$ donde (d^k, λ_{d^k}) es solución de $\nabla^2 \mathcal{L}(z^k, \lambda_k) d = -\nabla \mathcal{L}(z^k, \lambda_k)$.

Paso 2. Fijar $k = k + 1$ y volver al Paso 1.

Recordemos que el método de Newton asegura convergencia (con tasa superlineal o cuadrática) cuando la matriz Jacobiana de la función involucrada es no singular (Teorema 1.1.2). Entonces es importante garantizar el cumplimiento de esta condición. El siguiente resultado fue tomado de [43, Lema 4.2].

Proposición 1.1.15. *Sean f y h dos veces diferenciables en $z^* \in \mathbb{R}^n$. Sea z^* un punto estacionario del problema (1.10), satisfaciendo LICQ.*

Sea λ^ el único multiplicador de Lagrange asociado a z^* y supongamos que la SOSC dada en el Teorema 1.1.14 vale. Entonces la matriz Hessiana de la función Lagrangiana \mathcal{L} en el punto (λ^*, μ^*) es no singular.*

El siguiente resultado asegura la convergencia del método de Newton-Lagrange.

Teorema 1.1.16. *Sean f y h funciones dos veces diferenciables en un entorno del punto z^* , con derivadas segundas continuas en dicho punto. Supongamos que z^* es un punto estacionario del problema (1.10) satisfaciendo LICQ. Sea λ^* el único multiplicador de Lagrange asociado. Finalmente, supongamos que SOSC del Teorema 1.1.14 se satisface.*

Entonces para todo punto inicial (z^0, λ_0) , suficientemente próximo al punto (z^, λ^*) , el Algoritmo 3 genera una sucesión que converge a (z^*, λ^*) . La tasa de convergencia es q -superlineal, y si las derivadas segundas de f y h son Lipschitz continuas en un entorno de z^* , la convergencia es q -cuadrática.*

Extendemos este análisis a los problemas de optimización con restricciones mixtas.

1.1.3.3. Optimización con Restricciones Mixtas

Finalmente, consideramos el problema de optimización con restricciones de igualdad y desigualdad,

$$\begin{aligned} & \underset{z}{\text{minimizar}} && f(z) \\ & \text{s. a} && h(z) = 0, \\ & && g(z) \leq 0, \end{aligned} \tag{1.16}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ y $g : \mathbb{R}^n \rightarrow \mathbb{R}^r$ son funciones suaves. Para este caso, $\mathcal{D} = \{z \in \mathbb{R}^n \mid h(z) = 0, g(z) \leq 0\}$.

A continuación presentamos algunas definiciones que serán necesarias al momento de enunciar las condiciones de optimalidad para el problema de restricciones mixtas.

Definición 1.1.17. *Dado z un punto factible de (1.16), decimos que la restricción de desigualdad que corresponde al índice $j \in \{1, \dots, r\}$ está activa en el punto z si $g_j(z) = 0$. El conjunto de índices de las restricciones activas en el punto z está dado por*

$$\mathcal{I}(z) = \{j \mid g_j(z) = 0\}.$$

Si $j \notin \mathcal{I}(z)$, decimos que la j -ésima restricción está inactiva en z .

Ahora presentamos algunas condiciones de regularidad que usaremos en el desarrollo de esta sección.

Definición 1.1.18. Sea $z^* \in \mathbb{R}^n$.

- ✓ Condición de linealidad de las funciones restricciones. Esta condición vale si h y g son afines, es decir,

$$h(z) = Az + b \quad y \quad g(z) = Bz + c,$$

con $A \in \mathbb{R}^{p \times n}$, $B \in \mathbb{R}^{r \times n}$, $b \in \mathbb{R}^p$ y $c \in \mathbb{R}^r$. Esta condición puede ser debilitada al pedir que la linealidad se cumpla solo en algún entorno de z^* .

- ✓ Condición de Slater. Sea h afín y g_j convexa en \mathbb{R}^n para $j = 1, \dots, r$. Entonces esta restricción vale si existe un punto $\hat{z} \in \mathbb{R}^n$ tal que $h(\hat{z}) = 0$ y $g(\hat{z}) < 0^4$.
- ✓ Condición de Mangasarian-Fromovitz (MFCQ). Sean h y g diferenciables en z^* . Decimos que esta condición vale en z^* si

$$\text{rang}(\nabla h(z^*)) = p$$

y existe $y \in \mathbb{R}^n$ tal que

$$\nabla h(z^*)^T y = 0 \quad y \quad \nabla g_{\mathcal{I}(z^*)}(z^*)^T y < 0.$$

- ✓ Condición de independencia lineal (LICQ). Un punto z^* satisface dicha condición si los gradientes de las restricciones de igualdad $\nabla h_1(z^*), \dots, \nabla h_p(z^*)$ y los gradientes de las restricciones de desigualdad activas, $\nabla g_j(z^*)$ con $j \in \mathcal{I}(z^*)$, son linealmente independientes. En otras palabras,

$$\text{rang} \left(\left[\nabla h(z^*), \nabla g_{\mathcal{I}(z^*)}(z^*) \right] \right) = p + |\mathcal{I}(z^*)|.$$

Bajo alguna de estas condiciones en z se puede ver que

$$\mathcal{N}_{\mathcal{D}}(z) = \left\{ \sum_{i=1}^p \lambda_i \nabla h_i(z) + \sum_{j \in \mathcal{I}(z)} \mu_j \nabla g_j(z) \mid \lambda \in \mathbb{R}^p, \mu_j \geq 0 \ j \in \mathcal{I}(z) \right\}.$$

Como en el caso de restricciones de igualdad, ésta es la condición de regularidad de Guignard.

Claramente si vale LICQ vale MFCQ. Como podemos ver, si solo contamos con restricciones de igualdad, es decir, el problema de optimización es como (1.10), las condiciones MFCQ y LICQ son equivalentes.

A continuación definimos la función Lagrangiana asociada al problema (1.16).

Definición 1.1.19. La función Lagrangiana $\mathcal{L} : \mathbb{R}^{n+p+r} \rightarrow \mathbb{R}$ asociada al problema (1.16) está dada por

$$\mathcal{L}(z, \lambda, \mu) = f(z) + \sum_{i=1}^p \lambda_i h_i(z) + \sum_{j=1}^r \mu_j g_j(z).$$

⁴Donde $g(z) < 0$ significa $g_i(z) < 0$ para todo $i = 1, \dots, m$.

Ya podemos presentar las condiciones de optimalidad necesarias de primer orden. El siguiente resultado fue extraído de [43, Teorema 1.14].

Teorema 1.1.20. *Sean f y g diferenciables en z^* , donde h es diferenciable en un entorno de z^* con derivada continua en ese punto. Supongamos además, que alguna de las condiciones de regularidad listadas anteriormente se cumple. Entonces si z^* es una solución del problema (1.16), existen $\lambda^* \in \mathbb{R}^p$ y $\mu^* \in \mathbb{R}_+^r$ tal que*

$$\nabla_z \mathcal{L}(z^*, \lambda^*, \mu^*) = 0 \quad \text{y} \quad \mu^{*T} g(z^*) = 0. \quad (1.17)$$

La segunda igualdad en (1.17) es conocida como condición de complementariedad.⁵ Cuando $\mu^* \geq 0$ y $g(z^*) \leq 0$ dicha condición implica

$$\mu_j^* = 0 \quad \text{para todo} \quad j \notin \mathcal{I}(z^*).$$

El sistema dado por

$$\nabla_z \mathcal{L}(z, \lambda, \mu) = 0, \quad h(z) = 0, \quad \mu \geq 0, \quad g(z) \leq 0 \quad \text{y} \quad \mu^T g(z) = 0, \quad (1.18)$$

en las variables $(z, \lambda, \mu) \in \mathbb{R}^{n \times p \times r}$, caracteriza los puntos estacionarios del problema (1.16) y los multiplicadores de Lagrange asociados y es conocido como el sistema de Karush-Kuhn-Tucker (KKT).

Definición 1.1.21. *Diremos que un punto estacionario z^* de (1.16) cumple con la condición de regularidad estricta de Mangasarian-Fromovitz (SMFCQ) si el multiplicador de Lagrange asociado es único.*

La siguiente definición introduce el concepto de cono crítico.

Definición 1.1.22. *Sea $z \in \mathbb{R}^n$ un punto factible del problema (1.16), definimos el cono crítico del problema en dicho punto como*

$$\mathcal{C}(z) = \left\{ y \in \mathbb{R}^n \mid \nabla h(z^*)^T y = 0, \quad \nabla g_{\mathcal{I}(z^*)}(z^*)^T y \leq 0, \quad \nabla f(z^*)^T y \leq 0 \right\}.$$

El siguiente teorema extraído de [43, Teorema 1.19] presenta las condiciones de optimalidad necesarias de primer orden.

Teorema 1.1.23. *Sean los mapeos f , g y h dos veces diferenciables en z^* , con h y g diferenciables en un entorno del mismo y sus derivadas continuas en z^* . Supongamos que MFCQ se satisfacen en z^* .*

Si z^ es una solución del problema (1.16) entonces, para cada $y \in \mathcal{C}(z^*)$ existen (λ^*, μ^*) multiplicadores de Lagrange tal que*

$$y^T \nabla_{zz} \mathcal{L}(z^*, \lambda^*, \mu^*) y \geq 0.$$

Si z^ satisface SMFCQ, tenemos que*

$$y^T \nabla_{zz} \mathcal{L}(z^*, \lambda^*, \mu^*) y \geq 0,$$

para el único vector multiplicador de Lagrange (λ^, μ^*) .*

⁵En inglés, complementarity condition.

En [43, Teorema 1.20] podemos encontrar el siguiente resultado relacionado a las condiciones suficientes de segundo orden.

Teorema 1.1.24. *Sean f , g y h dos veces diferenciables en z^* , factible para el problema (1.16). Si para todo $y \in \mathcal{C}(z^*)$, con $y \neq 0$, existen (λ^*, μ^*) multiplicadores de Lagrange tal que*

$$y^T \nabla_{zz} \mathcal{L}(z^*, \lambda^*, \mu^*) y > 0. \quad (1.19)$$

entonces z^* es una solución local estricta del problema de optimización mixto. Más aún, existen constantes positivas γ y ε tales que

$$f(z) \geq f(z^*) + \frac{\gamma}{2} \|z - z^*\|^2$$

para todo z factible con $\|z - z^*\| < \varepsilon$.

Es evidente que LICQ implica SMFCQ y ésta implica MFCQ, por lo tanto, los resultados anteriores son válidos para esta condición.

Consideremos nuevamente el sistema KKT definido en (1.18). La idea detrás de las siguientes formulaciones, es reescribir éste sistema como un sistema de ecuaciones no lineales. Con este fin, introducimos el concepto de función de complementariedad.

Definición 1.1.25. *Una función $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ se llama función de complementariedad, si el conjunto de soluciones de la ecuación*

$$\psi(a, b) = 0$$

coincide con el conjunto de soluciones del sistema

$$a \geq 0, \quad b \geq 0, \quad ab = 0.$$

Entre las funciones de complementariedad más comunes podemos encontrar las siguientes.

- ✓ La función del residuo natural

$$\psi(a, b) = \min\{a, b\} \quad \text{con } a, b \in \mathbb{R}.$$

- ✓ La función de Fischer-Burmeister dada por

$$\psi(a, b) = \sqrt{a^2 + b^2} - a - b \quad \text{con } a, b \in \mathbb{R}.$$

Para más información sobre funciones de complementariedad invitamos al lector a ver [76, 13].

Haciendo uso de estas funciones, podemos reformular el sistema KKT del problema (1.16) definido en (1.18) como sigue,

$$\nabla_z \mathcal{L}(z, \lambda, \mu) = 0, \quad h(z) = 0, \quad \psi(\mu_j, -g_j(z)) = 0 \quad \text{para } j = 1, \dots, p,$$

para alguna elección de ψ . Con lo cual, nuestro nuevo problema puede ser escrito como un sistema de ecuaciones

$$\Phi(z, \lambda, \mu) = 0,$$

donde

$$\Phi : \mathbb{R}^{n \times p \times r} \rightarrow \mathbb{R}^{n \times p \times r}$$

dada por,

$$\Phi(z, \lambda, \mu) = \begin{bmatrix} \nabla_z \mathcal{L}(z, \lambda, \mu) \\ h(z) \\ \psi(\mu_1, -g_1(z)) \\ \vdots \\ \psi(\mu_p, -g_p(z)) \end{bmatrix}. \quad (1.20)$$

Si bien podríamos pensar en aplicar el método de Newton para ecuaciones en el caso de (1.20), no hemos dicho nada acerca de la diferenciabilidad de la función ψ . En el caso de los dos ejemplos anteriores, las funciones no son diferenciables en $(0, 0)$, entonces si la condición de complementariedad estricta (1.17) (segunda ecuación) no se cumple, el sistema no será diferenciable en la solución (z^*, λ^*, μ^*) independientemente de la diferenciabilidad de las funciones involucradas. Si bien hay funciones de complementariedad suaves, estos casos conllevan otros problemas computacionales, por lo que el uso de funciones de complementariedad no diferenciables es más atractivo computacionalmente.

Para resolver la reformulación del sistema KKT otros métodos son empleados, tales como los conocidos métodos de Newton generalizados, que en vez de utilizar la derivada clásica hacen uso del llamado subdiferencial de Clarke o Bouligand, propuestas que incluyen métodos quasi-Newton no suaves [11, 52, 54], entre otros [12, 22, 57, 69, 75, 51, 72, 73].

Veamos entonces una alternativa para resolver el problema (1.16). El método de Newton para resolver dicho problema puede ser escrito como una sucesión de problemas de optimización, en este caso es conocido en la literatura como método SQP, Programación Cuadrática Secuencial⁶ (ver [4, 66]) y puede describirse de la siguiente manera: dado w una buena estimación de un punto estacionario z^* del problema (1.16), una iteración del método halla (z, λ, μ) , con $z = w + d$ donde (λ, μ) es un multiplicador de Lagrange asociado a un punto estacionario d del problema

$$\begin{aligned} \underset{d}{\text{minimizar}} \quad & \nabla f(w)^T d + \frac{1}{2} d^T \nabla_{zz}^2 \mathcal{L}(w, \lambda, \mu) d \\ \text{s. a} \quad & h(w) + \nabla h(w)^T d = 0, \\ & g(w) + \nabla g(w)^T d \leq 0 \end{aligned} \quad (1.21)$$

Entonces el método SQP puede ser descrito como se indica a continuación.

⁶En inglés, Sequential Quadratic Programming.

Algoritmo 4

- Paso 0. Elegir $(z^0, \lambda_0, \mu_0) \in \mathbb{R}^{n \times p}$ y fijar $k = 0$.
- Paso 1. Definir $z^{k+1} = z^k + d^k$ donde d^k es punto estacionario de (1.21) y multiplicadores de Lagrange λ_{k+1} y μ_{k+1} asociados a d^k .
- Paso 2. Fijar $k = k + 1$ y volver al Paso 1.

El siguiente resultado asume solo SMFCQ y SOSC en el punto estacionario del problema z^* con multiplicador de Lagrange (λ^*, μ^*) asociado. Las condiciones LICQ y de complementariedad estricta no son necesarias para obtener resultados de convergencia.

Teorema 1.1.26. *Sean f , h y g dos veces diferenciables en un entorno del punto z^* , con derivadas segundas continuas en este punto. Sea z^* una solución local del problema (1.16), satisfaciendo SMFCQ y SOSC para el multiplicador de Lagrange asociado (λ^*, μ^*) .*

Entonces existe una constante $\delta > 0$ tal que para cualquier punto inicial (z^0, λ_0, μ_0) , suficientemente próximo a (z^, λ^*, μ^*) , existe una sucesión de puntos $\{(z^k, \lambda_k, \mu_k)\}$ generada por el Algoritmo 4, que cumple*

$$\|(z^{k+1} - z^k, \lambda_{k+1} - \lambda_k, \mu_{k+1} - \mu_k)\| \leq \delta,$$

tal que la sucesión converge a (z^, λ^*, μ^*) y la tasa de convergencia es q -superlineal. Más aún, si las derivadas segundas de f , h y g son localmente Lipschitz continuas con respecto a z^* , la tasa q -cuadrática es alcanzada.*

A pesar de que los métodos presentados son sencillos de comprender e implementar, hay ciertas desventajas que debemos tener en cuenta, algunas de las cuales ya hemos mencionado. Para hacer frente a dichas desventajas surgen alternativas, tales como los conocidos métodos quasi-Newton. Estos métodos se basan en las ideas del método de Newton, pero en estos casos no es necesario calcular de manera explícita información de primer o segundo orden. Por otro lado, evitan resolver sistemas lineales, lo cual reduce ampliamente el costo computacional. Sin embargo, es inevitable que mejoras en un sentido impliquen algunas pérdidas en otros. Por ejemplo, la tasa de convergencia es, en general, más lenta (a lo sumo q -superlineal) y el número de iteraciones que se realiza es usualmente mayor, aunque menos costosas.

Entre las alternativas mencionadas, una de las más usadas son los llamados métodos quasi-Newton. En la siguiente sección damos una breve descripción de los mismos.

1.2. Métodos Secantes

William Davidon (1927-2013) en 1959 y Charles Broyden (1933-2011) en 1965, introdujeron nuevos métodos para el problema de minimización y resolución de sistemas de

ecuaciones respectivamente. Estos trabajos, junto con importantes modificaciones hechas en el trabajo de Davidon, Fletcher y Powell (1963), han generado una amplia investigación a finales de la década de 1960 y principios de 1970. Dicho trabajo ha conducido a una nueva clase de algoritmos que han sido llamados como métodos quasi-Newton, de actualización secante y modificaciones de Newton, entre otros.

Después de la publicación del libro de Dennis y Schnabel [22], muchos autores comenzaron a diferenciar los llamados métodos quasi-Newton de los métodos secantes. Con el término quasi-Newton nos referimos a métodos que siguen las líneas del método clásico de Newton, aproximando de alguna manera, la derivada de la función. Los métodos secantes pertenecen a una familia de los métodos quasi-Newton, donde además se pide que tales aproximaciones cumplan con la llamada ecuación secante. Más adelante estos comentarios serán esclarecidos. En esta sección nos enfocaremos en el análisis de estos últimos.

Las actualizaciones secantes permiten generar aproximaciones a la matriz Jacobiana, en el caso de problemas de ecuaciones no lineales, y a la matriz Hessiana, en el caso de problemas de optimización. Dado que las derivadas no son necesarias, al menos analíticamente, estos métodos son muchas veces más eficientes que el método de Newton en cualquiera de sus versiones.

En lo que sigue, discutimos algunos de estos métodos para problemas de tamaño chico y medio. Para problemas de gran escala (es decir, cuando el número de variables n es grande) otras estrategias han sido desarrolladas pero no son discutidas en profundidad en este trabajo. Para una lectura del tema invitamos al lector a ver [20, 23, 36, 55, 58, 60, 61, 62, 65].

Comenzamos nuestra discusión de los métodos secantes para problemas de ecuaciones no lineales, ya que las aproximaciones secantes a la matriz Jacobiana son más simples que aquéllas correspondientes a la matriz Hessiana.

1.2.1. Métodos Secantes para Sistemas de Ecuaciones

Recordemos que el problema en cuestión consiste en encontrar un punto $z \in \mathbb{R}^n$ tal que $F(z) = 0$, donde F es un mapeo suave. La idea es aproximar, en cada iteración, la matriz Jacobiana de la función usando solo valores funcionales que han sido previamente calculados.

Como ya hemos mencionado, podemos aproximar la función por su expansión en serie de Taylor, es decir $F(w + d) \approx F(w) + F'(w)d$. Entonces si M_+ es una buena aproximación a $F'(w)$, es razonable pensar en pedir que dicha matriz satisfaga la ecuación

$$F(w + d) = F(w) + M_+d$$

o de otra forma,

$$M_+d = y. \tag{1.22}$$

donde $d = z - w$ e $y = F(z) - F(w)$. La ecuación (1.22) es conocida como la ecuación secante. Entonces, si w es una buena aproximación a una solución z^* , el iterado siguiente

se define como

$$z = w - M_+^{-1}y. \quad (1.23)$$

En el método clásico de Newton $M_+ = F'(w)$. En muchos métodos, calcular un nuevo iterado en (1.23) no involucra el cálculo directo de la inversa de la matriz, de hecho, en algunos casos, la actualización de la inversa de una matriz se obtiene a partir de la inversa de la anterior, como veremos más adelante .

En algunas ocasiones, la matriz Jacobiana es rala⁷ y dicho patrón es conocido. En la práctica es conveniente que tal estructura sea aprovechada, especialmente si hablamos de problemas de gran escala. Ver [14, 15, 16, 53].

Si bien buscamos que la ecuación (1.22) se satisfaga, esta condición no determina unívocamente a M_+ . De esta forma, construir una buena aproximación secante consiste en tomar una buena elección dentro de todas las posibles, de manera que las propiedades de la matriz Jacobiana se conserven o faciliten su uso en un algoritmo quasi-Newton.

Si bien hay muchas formas de obtener estas aproximaciones, una de las más usadas en la literatura es elegir M_+ tratando de que el cambio en el modelo afín sea mínimo, respetando siempre la condición (1.22). Estos métodos son conocidos como métodos secantes de cambio mínimo. Diferentes interpretaciones de “mínimo” llevan a diferentes métodos quasi-Newton. A continuación damos algunos ejemplos de dichas actualizaciones, donde dada M , aproximación previa de la matriz Jacobiana, se define la siguiente como M_+ , todas ellas cumpliendo con la ecuación (1.22).

- ✓ La siguiente actualización fue propuesta por C. Broyden en 1965 (ver [9]) y es conocida como actualización de Broyden buena, siendo una de las más populares dentro de los métodos quasi-Newton.

$$M_+ = M + \frac{(y - Md)d^T}{d^T d}. \quad (1.24)$$

- ✓ Una segunda actualización, también sugerida por Broyden es la siguiente.

$$M_+^{-1} = M^{-1} + \frac{(d - M^{-1}y)y^T}{y^T y}. \quad (1.25)$$

Esta formulación tiene la ventaja de calcular directamente M_+^{-1} . En la práctica, los métodos que usan este tipo de actualizaciones han sido considerados menos exitosos que aquéllos que usan (1.24), de hecho, (1.25) es conocida como la mala actualización de Broyden.

- ✓ El tercer y último ejemplo es una generalización que abarca las dos actualizaciones anteriores.

$$M_+ = M + \frac{(y - Md)v^T}{v^T d}. \quad (1.26)$$

Dicha actualización está bien definida para todo $v \in \mathbb{R}^n$ tal que $v^T d \neq 0$ y cumple con la ecuación secante. Tomando $v = d$ y $v = M^T d$ recuperamos las actualizaciones de Broyden buena y mala respectivamente.

⁷En inglés, sparse.

Ahora presentamos un esquema del algoritmo para los métodos secantes.

Algoritmo 5

- Paso 0. Elegir $z^0 \in \mathbb{R}^n$, $M_0 \in \mathbb{R}^{n \times n}$ y fijar $k = 0$.
- Paso 1. Definir d^k solución de $M_k d = -F(z^k)$ y fijar $z^{k+1} = z^k + d^k$.
- Paso 2. Calcular M_{k+1} usando alguna actualización del tipo (1.24), (1.25), (1.26).
- Paso 3. Fijar $k = k + 1$ y volver al Paso 1.

Antes de pasar a la siguiente subsección, mencionamos un teorema de convergencia para este tipo de métodos, usando la actualización de Broyden (buena). Para más detalles ver [22, Teorema 8.2.2].

Teorema 1.2.1. *Sea $\varepsilon' > 0$. Supongamos $z^* \in \mathbb{R}^n$ es una solución de (1.1) y $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ continuamente diferenciable con derivada Lipschitz continua en todo punto $z \in \mathcal{B}_{\varepsilon'}(z^*)$ y $F'(z^*)$ no singular.*

Entonces, existen $\varepsilon > 0$ y $\delta > 0$ tal que si $z^0 \in \mathcal{B}_\varepsilon(z^)$ y $M_0 \in \mathcal{B}_\delta(F'(z^*))$, el Algoritmo 5 genera una sucesión de iterados $\{z^k\}$ que converge a z^* . La tasa de convergencia de este método es q -superlineal si la regla de actualización para las matrices es (1.24).*

Si bien solo hemos mencionado unos pocos, existen una variedad de métodos quasi-Newton. Entre los métodos de actualización secante de cambio mínimo podemos mencionar los de actualización directa vía factorización, donde puede escribirse la matriz Jacobiana de F como producto de matrices, es decir, $F'(z) = M(z)^{-1}N(z)$, para todo $z \in \mathbb{R}^n$. Esta descomposición puede corresponderse con la factorización LU y podemos encontrar un desarrollo detallado en [45]. Cuestiones relacionadas a factorizaciones ortogonales pueden encontrarse en [59].

Otros métodos de gran importancia son los llamados métodos estructurados. En estos casos, $F'(z) = M(z) + N(z)$, para todo $z \in \mathbb{R}^n$, donde $M(z)$ es fácilmente calculable a diferencia de $N(z)$. Algunos ejemplos de métodos estructurados pueden ser encontrados en [1, 39, 48].

Para una discusión más completa del tema invitamos al lector a ver [60].

1.2.2. Métodos Secantes para Problemas de Optimización

En esta subsección consideramos métodos secantes para problemas de optimización sin restricciones. Si bien podríamos pensar en aplicar las técnicas mencionadas anteriormente al problema $\nabla f(z) = 0$, debemos tener en cuenta que estaríamos desperdiciando ciertas propiedades de la matriz Hessiana, tales como la simetría y frecuentemente el hecho de ser definida positiva.

Las actualizaciones que presentamos aquí intentan mantener las propiedades de estas matrices. Si pensamos la Hessiana de una función como derivada del sistema $\nabla f(z) = 0$, entonces la ecuación (1.22) debe ser satisfecha, es decir

$$H_+(z - w) = \nabla f(z) - \nabla f(w), \quad (1.27)$$

donde H_+ es la aproximación secante de $\nabla^2 f(z)$. En lo que sigue damos algunos ejemplos de actualizaciones que cumplen la ecuación secante y mantienen la simetría.

- ✓ La conocida actualización secante simétrica o también llamada actualización de Powell-Broyden simétrica (PSB), fue presentada por Powell en 1970, ver [70].

$$H_+ = H + \frac{(y - Hd)d^T + d(y - Hd)^T}{d^T d} - \frac{d^T (y - Hd) d d^T}{(d^T d)^2}. \quad (1.28)$$

- ✓ Otra fórmula para H_+ fue propuesta por Greenstadt en 1970.

$$H_+^{-1} = H^{-1} + \frac{(d - H^{-1}y)y^T + y(d - H^{-1}y)^T}{y^T y} - \frac{y^T (d - H^{-1}y) y y^T}{(y^T y)^2}. \quad (1.29)$$

- ✓ La actualización simétrica de rango 1, SR1.

$$H_+ = H + \frac{(y - Hd)(y - Hd)^T}{(y - Hd)^T d}. \quad (1.30)$$

Lamentablemente en estos casos, H_+ no necesariamente es definida positiva aún cuando H lo es. En lo que sigue, presentamos otras actualizaciones que sí preservan esta última propiedad.

- ✓ La siguiente fórmula fue presentada por Broyden [10], Fletcher [33], Goldfarb [35] y Shanno [74], de manera independiente, en 1970, de allí que se la conoce con el nombre de actualización BFGS (o actualización secante definida positiva).

$$H_+ = H + \frac{yy^T}{y^T d} - \frac{H d d^T H}{d^T H d}. \quad (1.31)$$

- ✓ La primer actualización secante conocida es la llamada actualización DFP, debido a Davidon (1959) y posteriormente popularizada en los trabajos de Fletcher y Powell (1963). Ver [17].

$$H_+ = H + \frac{(y - Hd)y^T + y(y - Hd)^T}{y^T d} - \frac{(y - Hd)^T d y y^T}{(y^T d)^2}. \quad (1.32)$$

Ahora estamos en condiciones de presentar el Algoritmo quasi-Newton para este tipo de actualizaciones de la matriz Hessiana.

Algoritmo 6

- Paso 0. Elegir $z^0 \in \mathbb{R}^n$, $H_0 \in \mathbb{R}^{n \times n}$ y fijar $k = 0$.
- Paso 1. Definir d^k solución de $H_k d = -\nabla f(z^k)$ y fijar $z^{k+1} = z^k + d^k$.
- Paso 2. Calcular H_{k+1} usando alguna actualización del tipo secante.
- Paso 3. Fijar $k = k + 1$ y volver al Paso 1.

El siguiente teorema se debe a Broyden, Dennis y Moré (1973) y asegura la convergencia local de este tipo de métodos a un punto estacionario localmente aislado. Un resultado similar se obtiene si pedimos que la matriz Hessiana $\nabla^2 f(z^*)$ sea definida positiva. Para más detalles de los resultados y demostraciones ver [22, Teoremas 9.2.1 y 9.3.1].

Teorema 1.2.2. *Sea $\varepsilon' > 0$. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dos veces continuamente diferenciable con derivada localmente Lipschitz continua en todo punto $z \in \mathcal{B}_{\varepsilon'}(z^*)$, donde z^* es tal que $\nabla f(z^*) = 0$ y $\nabla^2 f(z^*)$ es no singular.*

Entonces, existen $\varepsilon > 0$ y $\delta > 0$ tal que si $z^0 \in \mathcal{B}_\varepsilon(z^)$ y $H_0 \in \mathcal{B}_\delta(\nabla^2 f(z^*))$ simétrica, el Algoritmo 6, actualizando las matrices según (1.28), (1.29) o (1.30), genera una sucesión de iterados $\{z^k\}$ que converge q -superlinealmente a z^* .*

Si además, se cumple que $\nabla^2 f(z^)$ es definida positiva, entonces se obtiene el mismo resultado para las actualizaciones (1.31) o (1.32).*

Un hecho importante de destacar es que cuando las matrices quasi-Newton conservan la propiedad de ser simétricas y definidas positivas, una de las estrategias más naturales para resolver el sistema lineal en el Paso 1 del Algoritmo 6, es el uso de la factorización de Cholesky (ver [7, 66]), la cual nos permite pensar nuestra matriz como un producto de la forma GG^T , donde G es triangular inferior con elementos positivos en la diagonal.

En la siguiente sección presentamos un método tipo Newton para resolver un problema como (1.1), donde la función involucrada es tal que $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ con $n \neq m$, y además pedimos que z se encuentre en algún conjunto $\Omega \subseteq \mathbb{R}^n$.

1.3. Un Método tipo Newton para la Resolución de Sistemas de Ecuaciones

En esta sección presentamos un método para resolver un sistema de ecuaciones con restricciones y posibles soluciones no aisladas, junto con una extensión del método aplicado a un caso particular de funciones no diferenciables. Este método, conocido como LP-Newton⁸ fue presentado por Facchinei, Fischer y Herrich en [25] y significa un importante avance dentro del campo de optimización.

⁸En inglés, Linear Programming Newton method.

Nuevamente el objetivo es encontrar un punto $z \in \mathbb{R}^n$ tal que

$$F(z) = 0, \quad z \in \Omega \quad (1.33)$$

donde $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es una función continua y $\Omega \subseteq \mathbb{R}^n$ es un conjunto cerrado y no vacío. Es importante destacar la primera diferencia con respecto a los métodos presentados anteriormente, en este caso el dominio de la función involucrada es un subconjunto de \mathbb{R}^n y su imagen uno de \mathbb{R}^m , con n y m posiblemente diferentes.

El desarrollo de métodos de este tipo intenta extender el método de Newton clásico en al menos tres direcciones.

- ✓ Relajación en las hipótesis de diferenciabilidad. Hasta la actualidad, el manejo de las hipótesis relajadas de diferenciabilidad ha demandado grandes esfuerzos, debido a su importancia práctica. Los sistemas de ecuaciones no diferenciables surgen naturalmente tanto en el modelado de fenómenos naturales como en el estudio de problemas que son considerados suaves, como vimos anteriormente, sistemas de ecuaciones no diferenciables basados en reformulaciones de sistemas KKT.
- ✓ Relajación en las hipótesis de no singularidad. Los avances en este campo son más recientes y surgen de los intentos de desarrollar métodos de rápida convergencia para la solución de problemas usuales como sistemas KKT que surgen de los problemas de optimización, desigualdades variacionales y problemas de equilibrio de Nash generalizados.
- ✓ Restricciones en el conjunto factible. Restringir el conjunto en el cual buscamos soluciones puede significar algunos beneficios tales como evitar regiones donde la función es no diferenciable o donde su matriz Jacobiana sea singular. Por otro lado, nos permite ahorrar esfuerzo si sabemos de antemano en qué regiones buscar nuestras soluciones (por ejemplo, en el caso de los sistemas KKT, los multiplicadores deben ser no negativos).

La idea general del método es generar una sucesión de puntos $\{z^k\} \subseteq \Omega$ convergente a alguna solución z^* de (1.33), resolviendo en cada interacción k un subproblema de optimización. La idea detrás de este procedimiento, es acotar el error en la aproximación lineal a la función F y la dirección definida por un iterado y el siguiente, en cada interacción.

Dado un punto $w \in \Omega$, consideremos

$$\begin{aligned} & \underset{d, \gamma}{\text{minimizar}} \quad \gamma \\ & \text{s. a} \quad \|F(w) + M(w)d\| \leq \gamma \|F(w)\|^2, \\ & \quad \|d\| \leq \gamma \|F(w)\|, \\ & \quad w + d \in \Omega. \end{aligned} \quad (1.34)$$

Donde $M(w)$ se corresponde con la derivada o derivada generalizada de la función F en el punto w .

Es fácil ver que el problema anterior tiene solución para cualquier $w \in \mathbb{R}^n$. Por otro lado, tomando $\|\cdot\|$ como la norma infinito y Ω un conjunto poliedral, (1.34) no es más que un problema de optimización lineal. De allí su nombre. Estos tipos de problemas son fáciles de resolver aún cuando tratamos con grandes dimensiones, lo cual es una importante ventaja de este método respecto a otros.

Ahora estamos en condiciones de describir el método para la solución de (1.33).

Algoritmo 7

Paso 0. Elegir un punto $z^0 \in \Omega$ y fijar $k = 0$.

Paso 1. Si $F(z^k) = 0$, parar.

Paso 2. Definir $z^{k+1} = z^k + d^k$ donde el par (d^k, γ_k) es una solución de

$$\begin{aligned} & \underset{d, \gamma}{\text{minimizar}} \quad \gamma \\ & \text{s. a.} \quad \|F(z^k) + M(z^k)d\| \leq \gamma \|F(z^k)\|^2, \\ & \quad \quad \|d\| \leq \gamma \|F(z^k)\|, \\ & \quad \quad z^k + d \in \Omega. \end{aligned}$$

Paso 3. Fijar $k = k + 1$ e ir al Paso 1.

En lo que resta de la sección nos detendremos en algunos detalles del análisis de convergencia del algoritmo, ya que éste da el puntapié inicial a los trabajos que serán desarrollados en los posteriores capítulos.

A continuación presentamos algunas hipótesis que serán necesarias a la hora de demostrar la convergencia del algoritmo. Definimos el conjunto solución de (1.33) como

$$\mathcal{Z} = \{z \in \Omega \mid F(z) = 0\}, \quad (1.35)$$

el cual tomaremos no vacío. Sea $z^* \in \mathcal{Z}$, $\varepsilon > 0$ y consideremos la bola centrada en z^* de radio ε .

Hipótesis 1. Existe una constante $L > 0$ tal que

$$\|F(w)\| \leq L \text{dist}(w, \mathcal{Z}),$$

para todo $w \in \mathcal{B}_\varepsilon(z^*) \cap \Omega$.

Esta hipótesis es muy débil y se cumple cuando, por ejemplo, F es localmente Lipschitz continua.

La próxima hipótesis corresponde a una cota de error local de la función F alrededor del punto z^* en Ω , la cual nos da una estimación de la distancia de un cierto punto al conjunto solución. Esta condición es clave en la prueba de convergencia en el caso que z^* no sea un punto aislado.

Hipótesis 2. Existe una constante $\ell > 0$ tal que

$$\text{dist}(w, \mathcal{Z}) \leq \ell \|F(w)\|,$$

para todo $w \in \mathcal{B}_\varepsilon(z^*) \cap \Omega$.

Dado que el problema (1.34) está bien definido, para cada w denotamos $\gamma(w)$ al valor óptimo, al cual supondremos uniformemente acotado para todo $w \in \mathcal{B}_\varepsilon(z^*) \cap \Omega$. Esta condición se refleja en la siguiente hipótesis.

Hipótesis 3. Existe una constante $\Gamma \geq 1$ tal que

$$\gamma(w) \leq \Gamma,$$

para todo $w \in \mathcal{B}_\varepsilon(z^*) \cap \Omega$.

Es importante entender en qué casos generales esta condición vale. En el trabajo de Facchinei, Fischer y Herrich se muestra que si F es diferenciable, con derivada localmente Lipschitz continua y si provee una cota de error local en Ω , entonces la hipótesis se satisface.

Finalmente presentamos la última hipótesis, que como la anterior, es más bien técnica, lo que puede dificultar su comprensión.

Hipótesis 4. Existe $\hat{\alpha} > 0$ tal que si

$$u \in \mathcal{L}(w, \alpha) = \{u \in \Omega \mid \|u - w\| \leq \alpha, \|F(w) + M(w)(u - w)\| \leq \alpha^2\}$$

entonces,

$$\|F(w)\| \leq \hat{\alpha} \alpha^2$$

para todo $w \in \mathcal{B}_\varepsilon(z^*) \cap \Omega$, con $w \notin \mathcal{Z}$ y $\alpha \in [0, \varepsilon]$.

Esta hipótesis requiere que el mapeo $u \rightarrow F(w) + M(w)(u - w)$ sea una buena aproximación, en algún sentido, del mapeo $u \rightarrow F(u)$, para $u \in \Omega$ con u suficientemente cerca de w . Bajo la condición de que F sea diferenciable y F' localmente Lipschitz, esta hipótesis vale.

A continuación presentamos el resultado que asegura la convergencia q -cuadrática del Algoritmo 7. Dicho resultado fue extraído de [25, Teorema 1].

Teorema 1.3.1. El Algoritmo 7 está bien definido para cualquier punto inicial $z^0 \in \Omega$. Si las Hipótesis 1 – 4 son satisfechas, entonces existe $\varepsilon > 0$ tal que cualquier sucesión $\{z^k\}$, generada por el algoritmo con punto inicial $z^0 \in \mathcal{B}_\varepsilon(z^*) \cap \Omega$, converge q -cuadráticamente a algún punto \bar{z} .

El lector podrá encontrar en [25] la demostración del teorema anterior y de algunos resultados auxiliares. En una segunda parte del trabajo, los autores presentan algunos resultados de convergencia para funciones F con una estructura especial, que serán explotados para sistemas KKT. Entre ellas se encuentran las funciones afines a trozos.

Definición 1.3.2. Sean $F^1, \dots, F^p : \mathbb{R}^n \rightarrow \mathbb{R}^m$ funciones continuas. Una función $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ se dice que es una selección continua de las funciones F^1, \dots, F^p en un conjunto $\Upsilon \subseteq \mathbb{R}^n$ si F es continua en Υ y $F(z) \in \{F^1(z), \dots, F^p(z)\}$ para todo $z \in \Upsilon$. Denotamos por

$$\mathcal{I}(z) = \{i \in \{1, \dots, p\} \mid F(z) = F^i(z)\}$$

el conjunto de índices que aquellas funciones selección que son activas en z . La función F es llamada afín a trozos si $\Upsilon = \mathbb{R}^n$ y las funciones selección son todas afines.

En lo que sigue, explicaremos brevemente la aplicación del Algoritmo 7 a reformulaciones no diferenciables de sistemas KKT.

Sea $H : \mathbb{R}^n \rightarrow \mathbb{R}^n$ diferenciable con derivada localmente Lipschitz continua y $g : \mathbb{R}^n \rightarrow \mathbb{R}^r$ dos veces diferenciable con derivadas de segundo orden localmente Lipschitz continua. Consideremos el sistema KKT

$$H(z) + \nabla g(z)u = 0, \quad g(z) \leq 0, \quad u \geq 0, \quad u^T g(z) = 0 \quad (1.36)$$

el cual ya vimos que puede aparecer de problemas de optimización (donde H es el gradiente de la función objetivo) o desigualdades variacionales (donde H es la función que define las desigualdades variacionales).

Como vimos antes, el sistema (1.36) puede ser reformulado como un sistema de ecuaciones no lineales por medio de funciones de complementariedad. Entonces el sistema (1.36) puede ser escrito de la siguiente forma

$$\Phi(z, u) = \begin{bmatrix} H(z) + \nabla g^T(z)u \\ \psi(u_1, -g_1(z)) \\ \vdots \\ \psi(u_r, -g_r(z)) \end{bmatrix} = 0, \quad (1.37)$$

y además, sus conjuntos solución coinciden.

Tomando ψ como el residuo natural, los autores reformulan el problema (1.37) como un sistema de ecuaciones con restricciones

$$F(z, u, v) = \begin{bmatrix} H(z) + \nabla g^T(z)u \\ g(z) + v \\ \text{mín}\{u_1, v_1\} \\ \vdots \\ \text{mín}\{u_r, v_r\} \end{bmatrix} = 0, \quad (1.38)$$

con $(z, u, v) \in \Omega$, donde Ω denota un conjunto cerrado convexo tal que $\Omega \subseteq \mathbb{R}^n \times \mathbb{R}_+^r \times \mathbb{R}_+^r$.

Es fácil ver que un punto (z^*, u^*) resuelve (1.36) si y solo si $(z^*, u^*, -g(z^*))$ es una solución de (1.38). Esta forma de plantear el problema nos permite asegurarnos la validez de algunas hipótesis ya mencionadas. El próximo teorema [25, Teorema 5] describe varias situaciones donde las Hipótesis 1-4 valen, con lo cual podemos demostrar que la sucesión generada por el Algoritmo 7 converge cuadráticamente a una solución de (1.38).

Teorema 1.3.3. *Sea (z^*, u^*, v^*) una solución arbitraria del problema (1.38) y supongamos que*

$$\Omega = \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_+^m \quad \text{o} \quad \Omega = \{(z, u, v) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \mid u + v \geq 0\}.$$

Si alguna de las condiciones (a)-(d) vale, entonces las Hipótesis 1-4 son satisfechas. Además existe $\varepsilon > 0$ tal que cualquier sucesión infinita $\{z^k\}$ generada por el Algoritmo 7 converge q -cuadráticamente a una solución $(\bar{z}, \bar{u}, \bar{v})$ de (1.38), siempre que $(z^0, u^0, v^0) \in \mathcal{B}_\varepsilon((z^, u^*, v^*) \cap \Omega$ y que $M((z^k, u^k, v^k)) \in \{(F^i)'(z^k, u^k, v^k) \mid i \in \mathcal{I}((z^k, u^k, v^k))\}$ para todo $k \in \mathbb{N}$.*

(a) *Las funciones H y g son afines.*

(b) *La Hipótesis 2 vale y (z^*, u^*, v^*) es una solución no degenerada de (1.38).*

(c) *La Hipótesis 2 vale y si (z^*, u^*, v^*) y (z', u', v') son dos soluciones próximas, entonces $z^* = z'$.*

(d) *La Hipótesis 2 vale y si (z^*, u^*, v^*) y (z', u', v') son dos soluciones próximas, tal que $u^* = v^* = 0$ entonces $u' = v' = 0$.*

En lo que sigue damos un ejemplo extraído de [25, Ejemplo 6]. Este ejemplo es un sistema KKT derivado de un problema de optimización. Este ejemplo es considerado complicado para los métodos existentes dado que las soluciones primal y dual (si el lector no está familiarizado con la reformulación dual de un problema de optimización, ver [3]) son no aisladas y la ecuación proveniente de la reformulación es no diferenciable en el conjunto solución.

Ejemplo 1.3.4. *Consideremos el siguiente problema de optimización.*

$$\begin{aligned} & \underset{z}{\text{minimizar}} && (z_1 + z_2)z_4 + \frac{1}{2}(z_2 + z_3)^2 \\ & \text{s. a.} && z_3 \leq 0, \\ & && z_1 \geq 0, \\ & && z_2 \geq 0, \\ & && z_4 \geq 0, \\ & && z_1 + z_2 + z_3 \geq 0, \end{aligned} \tag{1.39}$$

El sistema KKT asociado a dicho problema puede ser reformulado a través del siguiente

sistema de ecuaciones

$$F(z, u, v) = \begin{bmatrix} z_4 + u_1 - u_2 - u_5 \\ z_4 + z_2 + z_3 - u_3 - u_5 \\ z_2 + z_3 - u_5 \\ z_1 + z_2 - u_4 \\ z_1 + v_1 \\ -z_1 + v_2 \\ 1 - z_2 + v_3 \\ -z_4 + v_4 \\ -z_1 - z_2 - z_3 + v_5 \\ \min\{u_1, v_1\} \\ \min\{u_2, v_2\} \\ \min\{u_3, v_3\} \\ \min\{u_4, v_4\} \\ \min\{u_5, v_5\} \end{bmatrix} = 0, \quad (1.40)$$

con multiplicadores u_1, \dots, u_5 y variables de holgura v_1, \dots, v_5 . El conjunto solución de (1.39) está dado por $\{(0, t, -t, 0) \mid t \geq 1\}$. Para cada una de estas soluciones, el correspondiente conjunto de multiplicadores es $\{(s, s, 0, t, 0) \mid s \geq 0\}$, por lo tanto, el conjunto solución de (1.40) es

$$\mathcal{Z} = \{(0, t, -t, 0, s, s, 0, t, 0, 0, 0, t - 1, 0, 0) \mid t \geq 1, s \geq 0\}.$$

Dado que F es afín a trozos, no es difícil ver que las Hipótesis 1-4 son satisfechas. Entonces el Algoritmo 7 converge localmente con tasa q -cuadrática desde la vecindad de alguna solución.

A continuación mostramos los resultados obtenidos por los autores al aplicar el Algoritmo 7 a este problema con $z^0 = (1, 4, -2, 1, 3, 3, 1, 4, 1, 0, 1, 3, 1, 3)$.

k	$\ F(z^k)\ _\infty$
\vdots	\vdots
6	7.9447e+02
7	2.4470e-04
8	2.3928e-07
9	2.3447e-13

Tabla 1.2: Resultados numéricos para el Ejemplo 1.3.4.

En el próximo capítulo presentamos una estrategia quasi-Newton basada en el método descrito en esta sección, la cual provee una herramienta útil y sencilla para resolver problemas como (1.33).

Capítulo 2

Un Método LP quasi-Newton Modificado

En este capítulo presentamos una adaptación del método LPN introducido anteriormente, obteniendo una velocidad de convergencia q -cuadrática para el algoritmo, para posibles soluciones no aisladas y relajando las hipótesis de suavidad de la función involucrada. En el método que proponemos a lo largo de este capítulo, el cual fue publicado en [63], la información de primer orden es reemplazada por una actualización quasi-Newton secante, la cual provee una herramienta computacionalmente simple y menos costosa. Esta estrategia combina:

- ✓ una actualización secante para un sistema de ecuaciones no lineales con soluciones aisladas y
- ✓ la posibilidad dada por el método LPN para resolver sistemas con restricciones cuyas soluciones pueden ser localmente no aisladas.

Antes de introducir el algoritmo, damos una breve descripción de algunos de los métodos ya existentes en la literatura para la solución de sistemas de ecuaciones. Una vez presentado el método, en la Sección 2.2 analizamos la convergencia local bajo ciertas hipótesis, entre las cuales cabe destacar una condición de cota de error estándar que asegura la convergencia en el caso de soluciones no aisladas. En la Sección 2.3 de este capítulo, presentamos algunos ejemplos numéricos que muestran el desempeño del algoritmo.

2.1. Introducción

Presentamos un método para la solución de sistemas de ecuaciones con restricciones, es decir, para encontrar un punto z tal que

$$F(z) = 0 \quad \text{con} \quad z \in \Omega. \quad (2.1)$$

Nuevamente el conjunto $\Omega \subseteq \mathbb{R}^n$ será cerrado y no vacío. En el caso de la función F , pediremos que $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ sea continuamente diferenciable con F' localmente Lipschitz continua. Como en el caso del algoritmo LPN, nuestra propuesta también contempla $n \neq m$. Esto nos permite tratar con un mayor número de problemas.

La idea de este nuevo método consiste en reemplazar la matriz Jacobiana o algún elemento del Jacobiano generalizado de la función F , el cual está representando como M en el subproblema (1.34). La motivación en este caso proviene del hecho de que el cálculo de este sustituto o aún de la misma matriz Jacobiana, puede introducir error numérico y ser costoso computacionalmente. Estas limitaciones son bien conocidas en la literatura que concierne a los sistemas de ecuaciones no lineales sin restricciones y parte de estos impedimentos han sido manejados a través del uso de estrategias quasi-Newton [5, 11].

Para sistemas sin restricciones los métodos quasi-Newton son usados como herramienta para acoplar el comportamiento local del método de Newton estándar con una consistente estrategia de globalización. En [32] fue mostrado que un método globalmente convergente puede ser formulado tanto para funciones continuamente diferenciables como para funciones pertenecientes a cierta clase continuamente diferenciable a trozos.

Es bien conocido que el problema (2.1) puede ser reformulado como un problema de cuadrados mínimos no lineales. Pero en este caso, los algoritmos de minimización pueden alcanzar puntos estacionarios que no necesariamente resuelven el problema deseado. Más aún, para funciones objetivo no convexas, la unicidad local de las soluciones es una hipótesis estándar. Sin embargo, debemos mencionar el trabajo [56] que estudia métodos de gradiente proyectado y donde la convergencia r -lineal fue probada para puntos estacionarios no aislados. En contraste, el método propuesto en este capítulo no requiere el cálculo de derivadas y una convergencia q -lineal es obtenida para un conjunto solución no aislado.

Para resolver problemas de cuadrados mínimos no lineales, es conocido de la literatura que el método Levenberg-Marquardt es una de las mejores opciones. Bajo ciertas hipótesis éste genera una sucesión que converge q -cuadráticamente a una posible solución no aislada [28, 30, 78]. Sin embargo, nos encontramos nuevamente frente a un método que requiere el cálculo de la matriz Jacobiana exacta.

Otra estrategia quasi-Newton para resolver el problema (2.1) es el propuesto por Ulbrich en [77]. En este trabajo se presenta una reformulación del problema como uno de cuadrados mínimos no lineales, y se intenta resolverlo usando un método de regiones de confianza. En este caso, la convergencia del algoritmo fue demostrada al asumir que los puntos en la preimagen del vector nulo son aislados y que cada una de las componentes de la función (F_i) es continuamente diferenciable fuera de $F_i^{-1}(0)$ y semisuave en $F_i^{-1}(0)$. También deberíamos mencionar métodos desarrollados para sistemas no lineales particulares dados por una reformulación de un sistema KKT [26, 71] y una reformulación de un problema de complementariedad mixto [47].

A continuación presentamos el siguiente procedimiento iterativo para la resolución del problema (2.1).

Algoritmo 8

Paso 0. Elegir $\sigma > 0$, $z^0 \in \Omega$, $M_0 \in \mathbb{R}^{m \times n}$ y fijar $k = 0$.

Paso 1. Definir $z^{k+1} = z^k + d^k$ donde el par (d^k, γ_k) es un solución de

$$\begin{aligned} & \underset{d, \gamma}{\text{minimizar}} \quad \gamma \\ & \text{s. a} \quad \|F(z^k) + M_k d\| \leq \sigma \gamma, \\ & \quad \quad \|d\| \leq \gamma, \\ & \quad \quad z^k + d \in \Omega. \end{aligned}$$

Paso 2. Si $F(z^{k+1}) = 0$, Parar. Sino, calcular $M_{k+1} \in \mathbb{R}^{m \times n}$.

Paso 3. Fijar $k = k + 1$ e ir al Paso 1.

Para simplificar el análisis de convergencia, vamos a considerar una actualización quasi-Newton secante de cambio mínimo de la matriz (para más detalles ver [21]). Además, las líneas del análisis de convergencia siguen las ideas de [6].

En contraste con los resultados estándar quasi-Newton, donde se obtiene convergencia q -superlineal, aquí solo probaremos convergencia q -lineal. A pesar de ello, es un resultado aún significativo dado que podemos resolver problemas con soluciones no aisladas y sin hipótesis de convexidad o monotonicidad, al resolver subproblemas simples y sin calcular información de primer orden. En el capítulo siguiente introduciremos una mejora de este método, donde se puede demostrar convergencia q -superlineal.

Como podemos observar, el subproblema de optimización que involucra el Algoritmo 8, es similar al definido en (1.34), donde M_k es un buen sustituto para $F'(z^k)$. Más aún, es una adaptación de [32, problema (5.9)]. Esta adaptación consiste en tomar un parámetro fijo σ en lugar de uno variable $\|F(z^k)\|$, cambio que fue hecho para obtener un punto factible del subproblema para garantizar el cumplimiento de la cota de error uniforme.

A continuación presentamos el análisis de convergencia para el Algoritmo 8.

2.2. Análisis de Convergencia Local

En esta sección presentamos una serie de resultados que nos serán de utilidad a la hora de demostrar la convergencia local del método propuesto.

Sea $\sigma > 0$, $w \in \Omega$ y $M \in \mathbb{R}^{m \times n}$, consideremos el subproblema de optimización planteado en el Paso 1 del Algoritmo 8,

$$\begin{aligned} & \underset{d, \gamma}{\text{minimizar}} \quad \gamma \\ & \text{s. a} \quad \|F(w) + M d\| \leq \sigma \gamma, \\ & \quad \quad \|d\| \leq \gamma, \\ & \quad \quad w + d \in \Omega. \end{aligned} \tag{2.2}$$

Veamos algunas propiedades importantes del problema (2.2).

Proposición 2.2.1. *Sea $w \in \Omega$, entonces,*

- (a) *el par $(0, \frac{1}{\sigma}\|F(w)\|)$ es factible para el problema (2.2).*
- (b) *El problema de optimización (2.2) tiene una solución con $\gamma \leq \frac{1}{\sigma}\|F(w)\|$.*
- (c) *El valor optimal del problema (2.2) es cero si y solo si w es una solución de (2.1).*

Demostración.

- (a) Si $d = 0$ entonces,

$$\|F(w) + Md\| = \|F(w)\| = \sigma \frac{1}{\sigma} \|F(w)\| \quad \text{y} \quad \|d\| = 0 \leq \frac{1}{\sigma} \|F(w)\|,$$

con $w + d = w \in \Omega$. Entonces, (a) vale.

- (b) Por otro lado, si se suma al problema (2.2) la restricción $\gamma \leq \frac{1}{\sigma}\|F(w)\|$, el conjunto solución no cambia, ya que $(0, \frac{1}{\sigma}\|F(w)\|)$ es factible para el problema. Dado que este nuevo problema tiene un conjunto factible compacto y no vacío, con la función objetivo continua, se puede garantizar la existencia de solución. Luego (b) vale.

- (c) La prueba de (c) es trivial. Si $\gamma = 0$ entonces,

$$\|F(w) + Md\| = 0 \quad \text{y} \quad \|d\| = 0,$$

con $w + d \in \Omega$, luego $\|F(w)\| = 0$ y $w \in \Omega$, es decir, w es una solución de (1.33). Por otro lado, si w es una solución de (1.33) se tiene que el par $(d, \gamma) = (0, 0)$ es una solución de (2.2). ■

Dado que el objetivo en esta sección es definir un algoritmo quasi-Newton, proporcionamos una regla para generar una matriz M adecuada. Consideremos el siguiente conjunto cerrado convexo $\mathcal{X} \subseteq \mathbb{R}^{m \times n}$ de manera que

$$F'(w) \in \mathcal{X}, \quad \forall w \in \Omega.$$

Por otro lado, para puntos $z, w \in \mathbb{R}^n$ y $M \in \mathbb{R}^{m \times n}$ consideremos el siguiente problema

$$\begin{aligned} & \underset{N}{\text{minimizar}} \quad \|N - M\|_*^2 \\ & \text{s. a} \quad N(z - w) = F(z) - F(w), \\ & \quad \quad N \in \mathcal{X}. \end{aligned} \tag{2.3}$$

Dado que (2.3) es un problema de optimización fuertemente convexo, tenemos que la matriz M_+ es la única solución si y solo si

$$\langle M_+ - M, N - M_+ \rangle \geq 0,$$

para todo N tal que

$$N(z - w) = F(z) - F(w) \quad \text{y} \quad N \in \mathcal{X}.$$

Aquí, $\langle \cdot, \cdot \rangle$ es el producto interno asociado a la norma $\|\cdot\|_*$. Entonces podemos ver fácilmente que

$$\|M - M_+\|_*^2 \leq \|M - N\|_*^2 - \|M_+ - N\|_*^2, \quad (2.4)$$

para toda N tal que $N(z - w) = F(z) - F(w)$ y $N \in \mathcal{X}$.

Es valioso mencionar que si $\|\cdot\|_*$ es la norma de Frobenius se puede recobrar la actualización de Broyden cuando $\mathcal{X} = \mathbb{R}^{m \times n}$ y la actualización PSB cuando \mathcal{X} es el espacio de las matrices simétricas. En aquellos casos donde la aproximación de la matriz debe ser simétrica y definida positiva, una pequeña modificación puede ser hecha para incorporar la actualización BFGS (ver [29, 34, 38]).

A continuación establecemos algunas hipótesis que serán de gran utilidad a la hora de estudiar la convergencia. Algunas de ellas que serán usadas tanto en este capítulo como en el próximo, ya han sido definidas en el capítulo anterior, pero serán repetidas aquí para hacer la lectura más ágil.

Recordemos que $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es continuamente diferenciable con F' localmente Lipschitz continua.

Se asume, como en el caso anterior, que el problema (2.1) tiene un conjunto solución \mathcal{Z} no vacío, es decir

$$\mathcal{Z} = \{z \in \Omega \mid F(z) = 0\} \neq \emptyset.$$

Sea $z^* \in \mathcal{Z}$ una solución arbitraria pero fija de (2.2). Asumimos la existencia de una cota de error local alrededor del punto z^* , lo que se refleja en la siguiente hipótesis.

Hipótesis 5. *Existen constantes $\varepsilon_0 > 0$ y $\ell > 0$ tal que*

$$\text{dist}(w, \mathcal{Z}) \leq \ell \|F(w)\| \quad \text{para todo} \quad w \in \mathcal{B}_{\varepsilon_0}(z^*) \cap \Omega.$$

Esta condición puede ser obtenida a partir de ciertas hipótesis que dependen de la estructura de F . Por ejemplo, [31] para sistemas de complementariedades¹ y [44] para problemas de equilibrio de Nash generalizados.

Por otro lado, la hipótesis de suavidad de la función F , asegura que existen constantes $L_0 > 0$ y $L_1 > 0$ tales que

$$\|F(z) - F(w)\| \leq L_0 \|z - w\| \quad \text{y} \quad \|F'(z) - F'(w)\|_* \leq L_1 \|z - w\|,$$

para todo $z, w \in \mathcal{B}_{\varepsilon_0}(z^*)$. Entonces, podemos tomar ε_0 suficientemente pequeño de manera de lograr la validez del siguiente supuesto.

Hipótesis 6. *Existe $\varepsilon_0 > 0$ tal que*

$$\|F(w)\| \leq L_0 \text{dist}(w, \mathcal{Z}) \quad \text{para todo} \quad w \in \mathcal{B}_{\varepsilon_0}(z^*). \quad (2.5)$$

¹En inglés, complementarity system.

Por otro lado, definimos la constante $\beta > 0$ de manera que $\|M\| \leq \beta \|M\|_*$ para toda matriz $M \in \mathbb{R}^{m \times n}$.

Comenzamos presentando el siguiente resultado.

Proposición 2.2.2. *Sea $\sigma > 0$. Entonces existen constantes $\varepsilon_1 > 0$ y $\sigma_1 > 0$ tales que para cualquier $w \in \mathcal{B}_{\varepsilon_1}(z^*)$ y $M \in \mathcal{B}_{\eta_1}^*(F'(z^*))$, si el par $(\bar{d}, \bar{\gamma})$ es una solución del problema (2.2) entonces se cumple*

$$\|F(w) + M\bar{d}\| \leq \sigma \text{dist}(w, \mathcal{Z}), \quad (2.6)$$

$$\|\bar{d}\| \leq \text{dist}(w, \mathcal{Z}). \quad (2.7)$$

Demostración. Si $w \in \mathcal{Z}$, por la Proposición 2.2.1c y la segunda restricción en (2.2), se tiene que $(\bar{d}, \bar{\gamma}) = (0, 0)$. Entonces (2.6) y (2.7) valen.

Se definen

$$\varepsilon_1 = \min \left\{ \frac{\varepsilon_0}{2}, \frac{\sigma}{3\beta L_1} \right\} \quad \text{y} \quad \eta_1 = \frac{\sigma}{2\beta}.$$

Para $w \in \mathcal{B}_{\varepsilon_1}(z^*) \setminus \mathcal{Z}$ sea $\hat{w} \in \mathcal{Z}$ tal que $\|w - \hat{w}\| = \text{dist}(w, \mathcal{Z})$. Entonces, para $d = \hat{w} - w$ tenemos que

$$\|d\| = \text{dist}(w, \mathcal{Z}) \leq \|w - z^*\| \leq \varepsilon_1.$$

Por otro lado, dado que $\|\hat{w} - z^*\| \leq 2\|w - z^*\| \leq \varepsilon_0$ y $M \in \mathcal{B}_{\eta_1}(F'(z^*))$,

$$\begin{aligned} \|F(w) + Md\| &\leq \|F(w) + F'(w)d - F(\hat{w})\| + \|(M - F'(w))d\| \\ &\leq \frac{\beta L_1}{2} \|d\|^2 + \beta (\|F'(z^*) - F'(w)\|_* + \|M - F'(z^*)\|_*) \|d\| \\ &\leq \left(\frac{\beta L_1}{2} \varepsilon_1 + \beta L_1 \varepsilon_1 + \beta \eta_1 \right) \|d\| \\ &\leq \sigma \|d\|. \end{aligned}$$

Debido a la última desigualdad y al hecho que $\|d\| = \text{dist}(w, \mathcal{Z})$, llegamos a

$$\|F(w) + Md\| \leq \sigma \text{dist}(w, \mathcal{Z}).$$

Como $(d, \text{dist}(w, \mathcal{Z}))$ es factible para el problema (2.2), se tiene que $\bar{\gamma} \leq \text{dist}(w, \mathcal{Z})$. Luego \bar{d} satisface (2.6) y (2.7). ■

El próximo resultado es clave para mostrar que la sucesión de puntos $\{z^k\}$ generada por el Algoritmo 3 converge efectivamente a una solución del problema (2.1).

Lema 2.2.3. *Sea la Hipótesis 5 válida. Entonces si $\sigma > 0$, existen $\varepsilon_2 > 0$ y $\eta_2 > 0$ tales que si $w \in \mathcal{B}_{\varepsilon_2}(z^*)$, $M \in \mathcal{B}_{\eta_2}^*(F'(z^*))$ y $(\bar{d}, \bar{\gamma})$ es una solución del problema (2.2) entonces la siguiente desigualdad vale,*

$$\text{dist}(w + \bar{d}, \mathcal{Z}) \leq 2\sigma \ell \text{dist}(w, \mathcal{Z}). \quad (2.8)$$

Demostración. Si $w \in \mathcal{Z}$ entonces $(\bar{d}, \bar{\gamma}) = (0, 0)$, como ya fue mostrado. Luego (2.8) vale. Veamos qué sucede cuando $w \notin \mathcal{Z}$.

Por otro lado, dado que F es continuamente diferenciable, existe $r > 0$ tal que si $w, w + d \in \mathcal{B}_r(z^*)$ lo siguiente vale

$$\|F(w + d) - F(w) - F'(z^*)d\| \leq \frac{\sigma}{2}\|d\|.$$

Definimos las constantes ε_2 y η_2 como,

$$\varepsilon_2 = \min \left\{ \frac{\varepsilon_0}{4}, \varepsilon_1, \frac{r}{2} \right\} \quad \text{y} \quad \eta_2 = \min \left\{ \eta_1, \frac{\sigma}{2\beta} \right\}, \quad (2.9)$$

donde ε_1 y η_1 están dadas por la Proposición 2.2.2.

Ahora, sea $w \in \mathcal{B}_{\varepsilon_2}(z^*) \setminus \mathcal{Z}$ y $M \in \mathcal{B}_{\eta_2}^*(F'(z^*))$. Por (2.7) tenemos que

$$\begin{aligned} \|w + \bar{d} - z^*\| &\leq \|w - z^*\| + \text{dist}(w, \mathcal{Z}) \\ &\leq 2\|w - z^*\| \\ &\leq \min \left\{ \frac{\varepsilon_0}{2}, r \right\}. \end{aligned}$$

De esta forma obtenemos la siguiente desigualdad,

$$\begin{aligned} \|F(w + \bar{d})\| &\leq \|F(w + \bar{d}) - F(w) - F'(z^*)\bar{d}\| + \beta\|F'(z^*) - M\|_*\|\bar{d}\| \\ &\quad + \|F(w) + M\bar{d}\| \\ &\leq \frac{\sigma}{2}\|\bar{d}\| + \beta\eta_2\|\bar{d}\| + \sigma\text{dist}(w, \mathcal{Z}) \\ &\leq 2\sigma\text{dist}(w, \mathcal{Z}). \end{aligned}$$

Y por la Hipótesis 5,

$$\begin{aligned} \text{dist}(w + \bar{d}, \mathcal{Z}) &\leq \ell\|F(w + \bar{d})\| \\ &\leq 2\sigma\ell\text{dist}(w, \mathcal{Z}). \end{aligned}$$

Con lo cual queda demostrado el lema. ■

El siguiente resultado es una adaptación de la propiedad de deterioro acotado para la sucesión de matrices $\{M_k\}$. Algunos ejemplos de esta propiedad pueden verse en [22, Lema 8.2.1]. Dicha condición brinda una herramienta para el análisis local que cuantifica el peor comportamiento de la actualización quasi-Newton.

Proposición 2.2.4. *Suponiendo que la Hipótesis 5 vale, sean $\varepsilon_2 > 0$ y $\eta_2 > 0$ dados por el Lema 2.2.3 para algún $\sigma > 0$. Si $w \in \mathcal{B}_{\varepsilon_2}(z^*)$, $M \in \mathcal{B}_{\eta_2}^*(F'(z^*))$, el par $(\bar{d}, \bar{\gamma})$ es una solución del problema (2.2) y M_+ es la solución del problema (2.3) para $z = w + \bar{d}$, entonces para $c = 2L_1 \left(\frac{1}{2} + \max\{1, 2\sigma\ell\} \right)$ vale que*

$$\|M_+ - F'(\hat{z})\|_* \leq \|M - F'(\hat{w})\|_* + c\text{dist}(w, \mathcal{Z}),$$

donde $\hat{z}, \hat{w} \in \mathcal{Z}$ satisfacen $\|\hat{z} - z\| = \text{dist}(z, \mathcal{Z})$ y $\|\hat{w} - w\| = \text{dist}(w, \mathcal{Z})$.

Demostración. Definimos la matriz $N \in \mathbb{R}^{m \times n}$ como

$$N = \int_0^1 F'(w + t\bar{d}) dt.$$

Es fácil ver que N es factible para el problema (2.3). Entonces, por (2.4) tenemos que

$$\|M_+ - N\|_* \leq \|M - N\|_*. \quad (2.10)$$

Por otro lado, usando (2.7) y (2.9) llegamos a las siguientes desigualdades,

$$\begin{aligned} \|\hat{w} - z^*\| &\leq \|\hat{w} - w\| + \|w - z^*\| \\ &\leq 2\|w - z^*\| \\ &\leq \frac{1}{2}\varepsilon_0, \end{aligned}$$

y

$$\begin{aligned} \|\hat{z} - z^*\| &\leq \|\hat{z} - z\| + \|z - z^*\| \\ &\leq 2\|z - z^*\| \\ &\leq 4\|w - z^*\| \\ &\leq \varepsilon_0. \end{aligned}$$

Entonces $w, z, \hat{w}, \hat{z} \in \mathcal{B}_{\varepsilon_0}(z^*)$. Luego,

$$\begin{aligned} \|N - F'(\hat{z})\|_* &\leq \int_0^1 \|F'(w + t\bar{d}) - F'(\hat{z})\|_* dt \\ &\leq L_1 \int_0^1 \|w + t\bar{d} - \hat{z}\| dt \\ &\leq L_1 \int_0^1 (\|w + t\bar{d} - z\| + \|z - \hat{z}\|) dt \\ &\leq L_1 \int_0^1 (1-t)\|\bar{d}\| dt + L_1 2\sigma\ell \text{dist}(w, \mathcal{Z}) \\ &\leq L_1 \left(\frac{1}{2} + 2\sigma\ell\right) \text{dist}(w, \mathcal{Z}), \end{aligned}$$

donde se usó que $z = w + \bar{d}$, (2.7) y (2.8). De manera similiar, obtenemos

$$\begin{aligned} \|N - F'(\hat{w})\|_* &\leq \int_0^1 \|F'(w + t\bar{d}) - F'(\hat{w})\|_* dt \\ &\leq L_1 \int_0^1 \|w + t\bar{d} - \hat{w}\| dt \\ &\leq L_1 \int_0^1 (\|t\bar{d}\| + \|w - \hat{w}\|) dt \\ &\leq L_1 \int_0^1 t\|\bar{d}\| dt + \text{dist}(w, \mathcal{Z}) \\ &\leq \frac{3}{2}L_1 \text{dist}(w, \mathcal{Z}). \end{aligned}$$

Finalmente, usando (2.10) y las desigualdades anteriores, concluimos que

$$\begin{aligned} \|M_+ - F'(\hat{z})\|_* &\leq \|M_+ - N\|_* + \|N - F'(\hat{z})\|_* \\ &\leq \|M - F'(\hat{w})\|_* + \|F'(\hat{w}) - N\|_* + \|N - F'(\hat{z})\|_* \\ &\leq \|M - F'(\hat{w})\|_* + c \operatorname{dist}(w, \mathcal{Z}), \end{aligned}$$

donde $c = 2L_1 \left(\frac{1}{2} + \max\{1, 2\sigma\ell\} \right)$. ■

A continuación presentamos el siguiente resultado que nos da la convergencia del método propuesto.

Teorema 2.2.5. *Nuevamente consideremos la Hipótesis 5 válida. Sean $\{z^k\}$, $\{d^k\}$ y $\{M_k\}$ sucesiones generadas por el Algoritmo 8, al tomar M_{k+1} en el Paso 2 como la solución del problema (2.3) con $z = z^{k+1}$, $w = z^k$ y $M = M_k$. Si $\sigma \leq \frac{1}{4\ell}$, entonces existen constantes $\varepsilon > 0$ y $\eta > 0$ tales que si*

$$z^0 \in \mathcal{B}_\varepsilon(z^*) \cap \Omega \quad \text{y} \quad M_0 \in \mathcal{B}_\eta^*(F'(z^*)) \cap \mathcal{X},$$

entonces,

(a) la sucesión $\{z^k\}$ converge a algún punto $\bar{z} \in \mathcal{Z}$ y $\{\operatorname{dist}(z^k, \mathcal{Z})\}$ converge a cero,

(b) la sucesión $\{\|M_k - F'(\bar{z})\|_*\}$ converge.

Demostración.

(a) Sean ε_2 y η_2 dados por el Lema 2.2.3. En esta demostración probaremos por inducción que

$$\|z^j - z^*\| \leq \varepsilon_2, \tag{2.11}$$

$$\|M_j - F'(z^*)\|_* \leq \eta_2, \tag{2.12}$$

valen para todo $j \geq 0$.

Definimos

$$\varepsilon = \min \left\{ \frac{\varepsilon_2}{3}, \frac{\eta_2}{28L_1} \right\} \quad \text{y} \quad \eta = \frac{\eta_2}{2}. \tag{2.13}$$

Sean $z^0 \in \mathcal{B}_\varepsilon(z^*)$ y $M_0 \in \mathcal{B}_\eta^*(F'(z^*))$. Luego, (2.11) y (2.12) valen para $j = 0$.

Suponemos ahora que (2.11) y (2.12) valen para todo $j \leq k$. Entonces por la Proposición 2.2.2 tenemos que

$$\|F(z^j) + M_j d^j\| \leq \sigma \operatorname{dist}(z^j, \mathcal{Z}), \tag{2.14}$$

$$\|d^j\| \leq \operatorname{dist}(z^j, \mathcal{Z}), \tag{2.15}$$

para todo $j \leq k$. Así se puede llegar a,

$$\begin{aligned}
\|z^{k+1} - z^*\| &\leq \|z^{k+1} - z^0\| + \|z^0 - z^*\| \\
&\leq \sum_{j=0}^k \|z^{j+1} - z^j\| + \|z^0 - z^*\| \\
&\leq \sum_{j=0}^k \text{dist}(z^j, \mathcal{Z}) + \|z^0 - z^*\| \\
&\leq \sum_{j=0}^k (2\sigma\ell)^j \text{dist}(z^0, \mathcal{Z}) + \|z^0 - z^*\| \\
&\leq \left(\sum_{j=0}^k \frac{1}{2^j} + 1 \right) \|z^0 - z^*\| \\
&\leq 3\|z^0 - z^*\| \\
&\leq 3\varepsilon \\
&\leq \varepsilon_2,
\end{aligned}$$

donde usamos (2.15), (2.8), la desigualdad $2\sigma\ell \leq \frac{1}{2}$ y (2.13). Por otro lado, por la Proposición 2.2.4 y el hecho que $2\sigma\ell < 1$, obtenemos

$$\begin{aligned}
\|M_{k+1} - F'(\hat{z}^{k+1})\|_{\star} &\leq \|M_k - F'(\hat{z}^k)\|_{\star} + 3L_1 \text{dist}(z^k, \mathcal{Z}) \\
&\leq \|M_0 - F'(\hat{z}^0)\|_{\star} + 3L_1 \sum_{j=0}^k \text{dist}(z^j, \mathcal{Z}) \\
&\leq \|M_0 - F'(\hat{z}^0)\|_{\star} + 3L_1 \sum_{j=0}^k \frac{1}{2^j} \text{dist}(z^0, \mathcal{Z}) \\
&\leq \|M_0 - F'(\hat{z}^0)\|_{\star} + 6L_1 \text{dist}(z^0, \mathcal{Z}).
\end{aligned}$$

En consecuencia,

$$\begin{aligned}
\|M_{k+1} - F'(z^*)\|_{\star} &\leq \|M_{k+1} - F'(\hat{z}^{k+1})\|_{\star} + \|F'(\hat{z}^{k+1}) - F'(z^*)\|_{\star} \\
&\leq \|M_0 - F'(\hat{z}^0)\|_{\star} + 6L_1\|z^0 - z^*\| + 2L_1\|z^{k+1} - z^*\| \\
&\leq \|M_0 - F'(z^*)\|_{\star} + \|F'(z^*) - F'(\hat{z}^0)\|_{\star} + 12L_1\|z^0 - z^*\| \\
&\leq \|M_0 - F'(z^*)\|_{\star} + 2L_1\|z^0 - z^*\| + 12L_1\varepsilon \\
&\leq \eta + 14L_1\varepsilon \\
&\leq \eta_2.
\end{aligned}$$

Luego (2.11) y (2.12) valen para $j = k + 1$.

Ahora, dado que (2.11) y (2.12) valen para todo $k \geq 0$, por (2.7) y el Lema 2.2.3

llegamos a

$$\begin{aligned}
\|z^{k+j} - z^j\| &\leq \sum_{i=j}^{k+j-1} \|z^{i+1} - z^i\| \\
&\leq \sum_{i=j}^{k+j-1} \text{dist}(z^i, \mathcal{Z}) \\
&\leq \sum_{i=j}^{k+j-1} \frac{1}{2^{i-j}} \text{dist}(z^j, \mathcal{Z}) \\
&\leq 2 \text{dist}(z^j, \mathcal{Z}) \\
&\leq 2 \frac{1}{2^j} \text{dist}(z^0, \mathcal{Z}).
\end{aligned}$$

Esto nos dice que la sucesión $\{z^k\}$ es una sucesión de Cauchy, y como \mathcal{Z} es un conjunto cerrado por hipótesis, se tiene que $\{z^k\}$ converge a algún $\bar{z} \in \mathcal{Z}$. Puesto que $\text{dist}(z^k, \mathcal{Z}) \leq \|z^k - \bar{z}\|$, se tiene que (a) vale. Además, tomando límite para $k \rightarrow \infty$ se obtiene

$$\|z^j - \bar{z}\| \leq 2 \text{dist}(z^j, \mathcal{Z}), \quad (2.16)$$

para todo $j \geq 0$.

- (b) Con la mente puesta en probar la convergencia de la sucesión $\{\|M_k - F'(\bar{z})\|_\star\}$, notemos que por la Proposición 2.2.4, para todo $k > j$,

$$\begin{aligned}
\|M_k - F'(\hat{z}^k)\|_\star &\leq \|M_j - F'(\hat{z}^j)\|_\star + 3L_1 \sum_{i=j}^{k-1} \text{dist}(z^i, \mathcal{Z}) \\
&\leq \|M_j - F'(\hat{z}^j)\|_\star + 3L_1 \sum_{i=j}^{k-1} \frac{1}{2^{i-j}} \text{dist}(z^j, \mathcal{Z}) \\
&\leq \|M_j - F'(\hat{z}^j)\|_\star + 6L_1 \text{dist}(z^j, \mathcal{Z}).
\end{aligned}$$

Como probamos que $\hat{z}^k \rightarrow \bar{z}$ y F' es continua por hipótesis, entonces

$$\limsup_{k \rightarrow \infty} \|M_k - F'(\bar{z})\|_\star \leq \|M_j - F'(\hat{z}^j)\|_\star + 6L_1 \text{dist}(z^j, \mathcal{Z}).$$

Y tomando límite para $j \rightarrow \infty$ del lado derecho de la desigualdad anterior, llegamos a que

$$\limsup_{k \rightarrow \infty} \|M_k - F'(\bar{z})\|_\star \leq \liminf_{j \rightarrow \infty} \|M_j - F'(\bar{z})\|_\star.$$

Con esto concluimos que (b) vale y así queda demostrado el teorema. ■

Finalmente llegamos al último resultado de esta sección, donde la convergencia lineal del algoritmo queda demostrada.

Corolario 2.2.6. *Bajo las hipótesis del Teorema 2.2.5, la sucesión $\{\text{dist}(z^k, \mathcal{Z})\}$ converge q -linealmente a cero y $\{z^k\}$ converge q -linealmente a \bar{z} .*

Demostración. Para k definimos

$$N_k = \int_0^1 F'(z^k + td^k) dt.$$

Entonces, por (2.4) para $w = z^k$, $z = z^{k+1}$, $M = M_k$ y usando que N_k es factible para este problema, obtenemos

$$\|M_k - M_{k+1}\|_*^2 \leq \|M_k - N_k\|_*^2 - \|M_{k+1} - N_k\|_*^2$$

Debido a la definición de N_k y la continuidad de F' tenemos que $N_k \rightarrow F'(\bar{z})$, concluyendo por el Teorema 2.2.5 que

$$M_{k+1} - M_k \rightarrow 0. \quad (2.17)$$

Además dado que $M_{k+1}d^k = F(z^{k+1}) - F(z^k)$, por la Proposición 2.2.2 llegamos a que

$$\begin{aligned} \|F(z^{k+1})\| &= \|F(z^k) + M_{k+1}d^k\| \\ &\leq \sigma \text{dist}(z^k, \mathcal{Z}) + \|(M_{k+1} - M_k)d^k\| \\ &\leq (\sigma + \|M_{k+1} - M_k\|)\text{dist}(z^k, \mathcal{Z}). \end{aligned}$$

Entonces,

$$\begin{aligned} \frac{\text{dist}(z^{k+1}, \mathcal{Z})}{\text{dist}(z^k, \mathcal{Z})} &\leq \frac{\ell \|F(z^{k+1})\|}{\text{dist}(z^k, \mathcal{Z})} \\ &\leq \ell(\sigma + \|M_{k+1} - M_k\|). \end{aligned}$$

Por (2.17) tenemos que $\|M_{k+1} - M_k\| \rightarrow 0$ entonces para $k \rightarrow \infty$,

$$\frac{\text{dist}(z^{k+1}, \mathcal{Z})}{\text{dist}(z^k, \mathcal{Z})} \rightarrow \sigma \ell \leq \frac{1}{4}.$$

Luego, por (2.16) y el hecho que $\text{dist}(z^k, \mathcal{Z}) \leq \|z^k - \bar{z}\|$,

$$\frac{\|z^{k+1} - \bar{z}\|}{\|z^k - \bar{z}\|} \leq \frac{2\text{dist}(z^{k+1}, \mathcal{Z})}{\text{dist}(z^k, \mathcal{Z})} \rightarrow 2\ell\sigma \leq \frac{1}{2}.$$

De esta forma queda demostrado el corolario. ■

Para resumir, hemos desarrollado un método quasi-Newton para resolver sistemas de ecuaciones con restricciones, basados en los trabajos previos [25, 32]. Claramente el algoritmo propuesto no necesita información de primer orden, lo cual lo hace un método computacionalmente simple que converge al menos, q -linealmente. Y tal convergencia está garantizada aún cuando las soluciones son no aisladas, hecho que hace de este algoritmo una herramienta valiosa a la hora de resolver este tipo de problemas. Tal hecho se debe al requerimiento de la Hipótesis 5.

A continuación damos algunos ejemplos numéricos que muestran el desempeño del algoritmo.

2.3. Ejemplos Numéricos

En la sección anterior analizamos la convergencia del algoritmo bajo ciertas hipótesis. Aquí mostraremos su desempeño, no solo para problemas que satisfacen dichos requerimientos sino también para otros que no los cumplen.

El Algoritmo 8 fue implementado en Octave, usando el método Simplex a través de la función *glpk* para resolver el subproblema (2.2). Las matrices son definidas mediante la actualización de Broyden buena con $M_0 = F'(z^0)$. El criterio de parada usado fue el siguiente

- ✓ $\|F(z^k)\|_\infty < 1e-10$ (error residual),
- ✓ $\|z^{k+1} - z^k\| < 1e-16$ y
- ✓ $itmax = 1500$ (máxima número de iteraciones).

En la mayoría de los casos estudiados el Algoritmo 8 para por llegar a un residuo menor a la tolerancia dada. Los casos donde esto no sucede, son marcados con un asterisco (*). A continuación de cada ejemplo presentamos tablas donde se muestran el residuo final, el número de iteraciones para diferentes valores del parámetro σ y la solución encontrada por el algoritmo. En todos los casos, el conjunto solución \mathcal{Z} es no aislado, y si bien comenzamos desde el mismo punto inicial z^0 , al cambiar los valores del parámetro σ , el algoritmo puede encontrar diferentes soluciones al problema.

Ejemplo 2.3.1. Consideremos el siguiente sistema de desigualdades

$$\begin{aligned} z_1^2 + z_2^2 - 1 &\leq 0, \\ (z_1 - 1)^2 + z_2^2 - 1 &\leq 0. \end{aligned}$$

Tomando $\Omega = \mathbb{R}^2 \times \mathbb{R}_+^2$, este sistema puede ser reescrito en el formato del problema (2.1), es decir,

$$F(z) = \begin{bmatrix} z_1^2 + z_2^2 - 1 + z_3 \\ (z_1 - 1)^2 + z_2^2 - 1 + z_4 \end{bmatrix} \quad \text{con } z \in \Omega,$$

donde z_3 y z_4 son variables de holgura.

Este problema fue extraído de [25, Ejemplo 1]. Dado que el sistema satisface MFCQ en cualquier punto factible, es fácilmente visible que F satisface la Hipótesis 5 en cualquier $z^* \in \mathcal{Z}$. Además, F es diferenciable y su derivada es localmente Lipschitz continua, entonces todos los requerimientos de la Sección 2.2 se cumplen.

En las Tabla 2.1 y 2.2 se muestran algunos de los resultados numéricos obtenidos para distintos puntos iniciales.

σ	# Iteraciones	Residuo
0.1	14 $\bar{z} = (0.8796, 0, 0.2261, 0.9855)$	2.7756e-17
1e-2	11 $\bar{z} = (0.9663, 0, 0.0662, 0.9988)$	2.2204e-16
1e-3	10 $\bar{z} = (0.9626, 0, 0.0733, 0.9986)$	4.1453e-11
1e-4	9 $\bar{z} = (0.9622, 0, 0.0741, 0.9985)$	8.4163e-12
1e-5	8 $\bar{z} = (0.9622, 0, 0.0743, 0.9985)$	3.0211e-11

Tabla 2.1: Resultados para el Ejemplo 2.3.1 tomando $z^0 = (2, 0, 0, 1)$.

σ	# Iteraciones	Residuo
0.1	30 $\bar{z} = (0.2684, -0.3840, 0.7804, 0.3172)$	6.4690e-11
1e-2	19 $\bar{z} = (0.2496, -0.2841, 0.8569, 0.3563)$	4.4540e-11
1e-3	17 $\bar{z} = (0.2448, -0.2810, 0.8610, 0.3507)$	1.0314e-13
1e-4	18 $\bar{z} = (0.2443, -0.2807, 0.8614, 0.3502)$	3.3307e-16
1e-5	17 $\bar{z} = (0.2443, -0.2806, 0.8615, 0.3501)$	1.5410e-13

Tabla 2.2: Resultados para el Ejemplo 2.3.1 tomando $z^0 = (0, 1, 2.1, 0.1)$.

Ejemplo 2.3.2. Este ejemplo, como el anterior, fue extraído de [25, Ejemplo 5]. Sin bien en este caso no se satisface la Hipótesis 5, la intención es mostrar que aún así el algoritmo puede generar una sucesión de iterados $\{z^k\}$ que converge a una solución.

Consideremos el problema de complementariedad dado por

$$z \geq 0, \quad T(z) \geq 0 \quad \text{y} \quad z^T T(z) = 0,$$

donde T es diferenciables con derivada localmente Lipschitz continua. Entonces, para

$$T(z) = \begin{bmatrix} z_1 z_2 \\ z_1^2 + z_2 - 1 \end{bmatrix}$$

el problema puede ser formulado como

$$F(z) = \begin{bmatrix} z_1 z_2 - z_3 \\ z_1^2 + z_2 - 1 - z_4 \\ z_1 z_3 \\ z_2 z_4 \end{bmatrix} \quad \text{con } z \in \Omega,$$

donde $\Omega = \mathbb{R}_+^2 \times \mathbb{R}_+^2$ y z_3, z_4 son variables de holgura.

En las Tablas 2.3 y 2.4 se muestran algunos resultados.

σ	‡ Iteraciones	Residuo
0.1	1101	3.6278e-10* $\bar{z} = (2e-5, 1, 2e-5, 0)$
1e-2	119	3.9433e-12 $\bar{z} = (0, 1, 0, 0)$
1e-3	28	9.9180e-11 $\bar{z} = (0, 1, 0, 0)$
1e-4	41	2.9631e-11 $\bar{z} = (0, 1, 0, 0)$
1e-5	44	1.6547e-11 $\bar{z} = (0, 1, 0, 0)$

Tabla 2.3: Resultados para el Ejemplo 2.3.2 tomando $z^0 = (1, 1, 1, 1)$. Para $\sigma = 0.1$, el algoritmo para por el segundo criterio.

σ	‡ Iteraciones	Residuo
0.1	18	2.2204e-16 $\bar{z} = (1, 0, 0, 0)$
1e-2	16	2.3921e-11 $\bar{z} = (1, 0, 0, 0.0001)$
1e-3	16	7.4402e-12 $\bar{z} = (1.0001, -1.22e-12, -3.67e-13, 1.99e-04)$
1e-4	15	2.9631e-11 $\bar{z} = (1.0001, 4.37e-12, 1.47e-12, 2.00e-04)$
1e-5	14	2.4669e-11 $\bar{z} = (1.0001, -7.15e-12, 5.32e-12, 2.01e-04)$

Tabla 2.4: Resultados para el Ejemplo 2.3.2 tomando $z^0 = (2, 1, 1, 0)$.

El resto de los ejemplos son problemas de punto factible tomados de la Colección de Hock-Schittkowski (para más detalles ver [40]). En estos casos, se reformularon los conjuntos factibles al considerarlos como en (2.1), a pesar de la posible violación de la Hipótesis 5.

Ejemplo 2.3.3. El conjunto factible del problema HS19 puede ser escrito como (2.1) tomando

$$F(z) = \begin{bmatrix} -(z_1 - 5)^2 - (z_2 - 5)^2 + 100 + z_3 \\ (z_1 - 6)^2 + (z_2 - 5)^2 - 82.81 + z_4 \end{bmatrix} \quad \text{con } z \in \Omega,$$

donde $\Omega = [13, 100] \times [0, 100] \times \mathbb{R}_+^2$.

En las Tablas 2.5 y 2.6 mostramos los resultados obtenidos por el Algoritmo 8.

σ	# Iteraciones	Residuo
0.1	15 $\bar{z} = (1.49\text{e}+01, 3.1117, 1.6137, 1.51\text{e}-04)$	2.6512e-13
1e-2	13 $\bar{z} = (1.48\text{e}+01, 3.0846, 1.6022, 9.52\text{e}-05)$	8.6597e-15
1e-3	10 $\bar{z} = (1.48\text{e}+01, 3.0819, 1.6011, 7.28\text{e}-06)$	2.8451e-15
1e-4	9 $\bar{z} = (14.8954, 3.0816, 1.6009, 0)$	1.4140e-11
1e-5	8 $\bar{z} = (14.8954, 3.0815, 1.6009, 0)$	4.5191e-12

Tabla 2.5: Resultados para el Ejemplo 2.3.3 tomando $z^0 = (20, 5, 0, 0)$.

σ	# Iteraciones	Residuo
0.1	14 $\bar{z} = (14.095, 0.8429, 0, 0)$	5.6787e-11
1e-2	8 $\bar{z} = (14.095, 0.8429, 0, 0)$	6.2016e-11
1e-3	7 $\bar{z} = (1.40\text{e}+01, 8.46\text{e}-01, 1.42\text{e}-03, 1.74\text{e}-03)$	2.0145e-12
1e-4	7 $\bar{z} = (1.40\text{e}+01, 8.46\text{e}-01, 1.74\text{e}-03, 1.77\text{e}-03)$	3.0760e-14
1e-5	6 $\bar{z} = (1.40\text{e}+01, 8.46\text{e}-01, 1.77\text{e}-03, 1.77\text{e}-03)$	1.3751e-11

Tabla 2.6: Resultados para el Ejemplo 2.3.3 tomando $z^0 = (13, 1, 1, 0)$.

Ejemplo 2.3.4. El conjunto factible del problema HS23 puede ser escrito como (2.1) tomado

$$F(z) = \begin{bmatrix} -z_1 - z_2 + 1 + z_3 \\ -z_1^2 - z_2^2 + 1 + z_4 \\ -9z_1^2 - z_2^2 + 9 + z_5 \\ -z_1^2 + z_2 + z_6 \\ -z_2^2 + z_1 + z_7 \end{bmatrix} \quad \text{con } z \in \Omega,$$

donde $\Omega = [-50, 50] \times [-50, 50] \times \mathbb{R}_+^5$.

En la Tablas 2.7 y 2.8 se muestran algunos resultados obtenidos.

σ	Iteraciones	Residuo
0.1	48 $\bar{z} = (1.0462, 1.0787, 1.1250, 1.2583, 2.0160, 0.0159, 0.1174)$	6.1963e-11
1e-2	14 $\bar{z} = (1.1223, 1.1794, 1.3017, 1.6506, 3.7275, 0.0801, 0.2687)$	2.8282e-11
1e-3	13 $\bar{z} = (1.1425, 1.1946, 1.3372, 1.7326, 4.1767, 0.1108, 0.2845)$	5.7998e-13
1e-4	11 $\bar{z} = (1.1448, 1.1963, 1.3412, 1.7420, 4.2280, 0.1143, 0.2864)$	1.5838e-11
1e-5	12 $\bar{z} = (1.1451, 1.1965, 1.3416, 1.7429, 4.2332, 0.1147, 0.2865)$	6.5103e-13

Tabla 2.7: Resultados para el Ejemplo 2.3.4 tomando $z^0 = (2, 2, 0, 0, 0, 0, 0)$.

σ	Iteraciones	Residuo
0.1	31 $\bar{z} = (1, 1, 1, 1, 1, 0, 0)$	4.8342e-11
1e-2	6 $\bar{z} = (1, 1, 1, 1, 1, 0, 0)$	3.3411e-11
1e-3	4 $\bar{z} = (1, 1, 1, 1, 1, 0, 0)$	4.0501e-13
1e-4	6 $\bar{z} = (1, 1, 1, 1, 1, 0, 0)$	2.8422e-14
1e-5	2 $\bar{z} = (1, 1, 1, 1, 1, 0, 0)$	6.8342e-11

Tabla 2.8: Resultados para el Ejemplo 2.3.4 tomando $z^0 = (1, 1, 0, 0, 2, 0, 0)$.

Ejemplo 2.3.5. El conjunto factible del problema HS24 puede ser escrito como (2.1) tomando

$$F(z) = \begin{bmatrix} -z_1/\sqrt{3} + z_2 + z_3 \\ -z_1 - \sqrt{3}z_2 + z_4 \\ z_1 + \sqrt{3}z_2 - 6 + z_5 \end{bmatrix} \quad \text{con } z \in \Omega,$$

donde $\Omega = \mathbb{R}_+^5$.

Algunos resultados son mostrados en las Tablas 2.9 y 2.10.

σ	# Iteraciones	Residuo
0.1	15 $\bar{z} = (1.5186, 0.8611, 0.0156, 3.0101, 2.9898)$	5.7732e-15
1e-2	8 $\bar{z} = (1.5, 0.8660, 0, 3, 3)$	1.3323e-15
1e-3	6 $\bar{z} = (1.5021, 0.86479, 0.00246, 3, 3)$	8.8818e-16
1e-4	4 $\bar{z} = (1.5001, 8.65e-01, 1.16e-04, 3.0001, 2.9999)$	3.1264e-13
1e-5	3 $\bar{z} = (1.5, 8.66e-01, 6.72e-06, 3, 3)$	4.9195e-11

Tabla 2.9: Resultados para el Ejemplo 2.3.5 tomando $z^0 = (2, 1, 0, 0, 0)$.

σ	# Iteraciones	Residuo
0.1	10 $\bar{z} = (0.75, 0.43302, 0, 1.50001, 4.4999)$	8.8818e-16
1e-2	7 $\bar{z} = (0.75, 4.33e-01, 1.51e-09, 1.5, 4.5)$	4.5901e-12
1e-3	6 $\bar{z} = (0.75, 0.43301, 0, 1.5, 4.5)$	3.2710e-11
1e-4	7 $\bar{z} = (0.75, 0.43301, 0, 1.5, 4.5)$	1.6236e-12
1e-5	3 $\bar{z} = (0.75, 0.43301, 0, 1.5, 4.5)$	3.3307e-16

Tabla 2.10: Resultados para el Ejemplo 2.3.5 tomando $z^0 = (0, 0, 1, 0, 3)$.

Ejemplo 2.3.6. El conjunto factible del problema HS34 puede ser escrito como (2.1) tomando

$$F(z) = \begin{bmatrix} -z_2 + \exp(z_1) + z_4 \\ -z_3 + \exp(z_2) + z_5 \end{bmatrix} \quad \text{con } z \in \Omega,$$

donde $\Omega = [0, 100] \times [0, 100] \times [0, 10] \times \mathbb{R}_+^2$.

Las Tablas 2.11 y 2.12 muestran los resultados obtenidos.

σ	# Iteraciones	Residuo
0.1	13 $\bar{z} = (0.0111, 1.0155, 2.8025, 0.0043, 0.0416)$	4.8242e-12
1e-2	10 $\bar{z} = (4.16e-02, 1.0427, 2.8887, 1.33e-04, 5.19e-02)$	3.4037e-11
1e-3	11 $\bar{z} = (5.74e-02, 1.0594, 2.9427, 2.95e-04, 5.78e-02)$	7.5634e-16
1e-4	11 $\bar{z} = (0.3693, 1.4567, 4.6391, 0.0100, 0.3469)$	5.5511e-17
1e-5	10 $\bar{z} = (5.92e-02, 1.0614, 2.9489, 3.21e-04, 5.85e-02)$	7.5562e-11

Tabla 2.11: Resultados para el Ejemplo 2.3.6 tomando $z^0 = (0, 0, 0, 0, 0)$.

σ	# Iteraciones	Residuo
0.1	11 $\bar{z} = (0.3572, 1.4329, 4.5954, 0.0034, 0.4045)$	8.6900e-11
1e-2	10 $\bar{z} = (0.3681, 1.4542, 4.6476, 0.0091, 0.3661)$	1.7764e-15
1e-3	10 $\bar{z} = (0.3692, 1.4565, 4.6386, 0.0099, 0.3474)$	4.9960e-16
1e-4	10 $\bar{z} = (0.3693, 1.4567, 4.6391, 0.0100, 0.3469)$	1.1657e-15
1e-5	10 $\bar{z} = (0.3693, 1.4568, 4.6391, 0.0100, 0.3468)$	6.6613e-16

Tabla 2.12: Resultados para el Ejemplo 2.3.6 tomando $z^0 = (0, 1, 5, 0, 0)$.

Ejemplo 2.3.7. El conjunto factible para el problema HS60 puede ser escrito como (2.1) tomando

$$F(z) = z_1(1 + z_2^2) + z_3^4 - 4 - 3\sqrt{2}, \quad \text{con } z \in \Omega,$$

donde $\Omega = [-10, 10] \times [-10, 10] \times [-10, 10]$.

En la Tabla 2.13 se muestran algunos resultados.

σ	# Iteraciones	Residuo
0.1	11 $\bar{z} = (1.4142, 1.4142, 1.4142)$	3.5527e-15
1e-2	10 $\bar{z} = (1.4142, 1.4142, 1.4142)$	1.9540e-14
1e-3	8 $\bar{z} = (1.4142, 1.4142, 1.4142)$	4.4409e-14
1e-4	8 $\bar{z} = (1.4142, 1.4142, 1.4142)$	2.8422e-14
1e-5	8 $\bar{z} = (1.4142, 1.4142, 1.4142)$	6.5725e-14

Tabla 2.13: Resultados para el Ejemplo 2.3.7 tomando $z^0 = (1, 1, 1)$.

σ	# Iteraciones	Residuo
0.1	11 $\bar{z} = (-0.2839, -0.7160, -1.7160)$	5.5067e-14
1e-2	11 $\bar{z} = (-0.2839, -0.7160, -1.7160)$	1.0658e-14
1e-3	12 $\bar{z} = (-0.2839, -0.7160, -1.7160)$	7.1054e-15
1e-4	12 $\bar{z} = (-0.2839, -0.7160, -1.7160)$	4.3343e-13
1e-5	11 $\bar{z} = (-0.2839, -0.7160, -1.7160)$	1.2434e-14

Tabla 2.14: Resultados para el Ejemplo 2.3.7 tomando $z^0 = (-1, 0, -1)$.

Ejemplo 2.3.8. El conjunto factible del problema HS74 puede ser escrito como (2.1) tomando

$$F(z) = \begin{bmatrix} -z_4 + z_3 - 0.55 + x_5 \\ -z_3 + z_4 - 0.55 + x_6 \\ 1000 \operatorname{sen}(-z_3 - 0.25) + 1000 \operatorname{sen}(-z_4 - 0.25) + 894.8 - z_1 \\ 1000 \operatorname{sen}(z_3 - 0.25) + 1000 \operatorname{sen}(z_3 - z_4 - 0.25) + 894.8 - z_2 \\ 1000 \operatorname{sen}(z_4 - 0.25) + 1000 \operatorname{sen}(z_4 - z_3 - 0.25) + 1294.8 \end{bmatrix}$$

con $\Omega = [0, 1200] \times [0, 1200] \times [-0.55, 0.55] \times [-0.55, 0.55] \times \mathbb{R}_+^2$.

En las Tablas 2.15 y 2.16 pueden observarse los resultados obtenidos.

σ	Iteraciones	Residuo
0.1	342	5.9799e-11
	$\bar{z} = (800.56, 900.21, 3.41e-02, -4.38e-01, 7.86e-02, 1.0214)$	
1e-2	29	5.9711e-11
	$\bar{z} = (800.39, 900.39, 3.42e-02, -4.37e-01, 7.85e-02, 1.0214)$	
1e-3	16	1.7280e-11
	$\bar{z} = (800.39, 900.39, 3.42e-02, -4.37e-01, 7.85e-02, 1.0214)$	
1e-4	15	6.1164e-11
	$\bar{z} = (800.39, 900.39, 3.42e-02, -4.37e-01, 7.85e-02, 1.0214)$	
1e-5	13	2.6603e-11
	$\bar{z} = (800.39, 900.39, 3.43e-02, -4.37e-01, 7.85e-02, 1.0214)$	

Tabla 2.15: Resultados para el Ejemplo 2.3.8 tomando $z^0 = (800, 900, 0, 0, 0, 0)$.

σ	Iteraciones	Residuo
0.1	104	7.4799e-11
	$(800.3, 900.4, 3.43e-02, -4.37e-01, 7.85e-02, 1.0214)$	
1e-2	26	5.7526e-11
	$(800.3, 900.3, 3.42e-02, -4.37e-01, 7.86e-02, 1.0214)$	
1e-3	21	4.5475e-13
	$(800.3, 900.3, 3.42e-02, -4.37e-01, 7.86e-02, 1.0214)$	
1e-4	11	2.9559e-12
	$(800.3, 900.3, 3.42e-02, -4.37e-01, 7.86e-02, 1.0214)$	
1e-5	21	4.0700e-11
	$(800.3, 900.3, 3.42e-02, -4.37e-01, 7.86e-02, 1.0214)$	

Tabla 2.16: Resultados para el Ejemplo 2.3.8 tomando $z^0 = (800, 900, -0.5, -0.5, 0, 10)$.

Mostramos que el Algoritmo 8 tiene una buena performance cuando todas las hipótesis requeridas en la Sección 2.2 son satisfechas, convergiendo q -linealmente como es esperado, aún cuando las soluciones no son localmente aisladas. Sin embargo, si tales

hipótesis no se cumplen, el método todavía es capaz de encontrar una solución, como muestra la mayoría de los ejemplos. En general, notamos que los resultados de los experimentos numéricos mejoran notablemente a medida que el parámetro σ decrece. Además, observamos que el subproblema (2.2) es numéricamente estable (como fue sugerido en [32]).

2.3.1. Comentarios

Como último comentario, queremos agregar la motivación del trabajo presentado en el próximo capítulo. Como ya vimos, la sucesión generada por el Algoritmo 8 converge q -linealmente a alguna solución del problema (2.1). Sin embargo, al analizar cada uno de los ejemplos estudiados anteriormente, pudimos observar un cierto comportamiento superlineal, para algunos de ellos y ciertos valores del parámetro σ .

A continuación presentamos algunos resultados donde el comportamiento superlineal puede ser identificado.

k	$\ F(z^k)\ _\infty$	k	$\ F(z^k)\ _\infty$
\vdots	\vdots	\vdots	\vdots
6	1.4393e-05	5	1.1289e-03
7	1.2403e-07	6	1.3102e-05
8	8.1232e-10	7	1.2309e-08
9	8.4163e-12	8	3.0211e-11

Tabla 2.17: Resultados para el Ejemplo 2.3.1 con $z^0 = (2, 0, 0, 1)$ para $\sigma = 1e-4$ y $\sigma = 1e-5$.

k	$\ F(z^k)\ _\infty$	k	$\ F(z^k)\ _\infty$
\vdots	\vdots	\vdots	\vdots
7	1.4393e-05	5	9.8500e-05
8	1.1978e-06	6	6.4755e-08
9	4.4806e-10	7	5.4180e-10
10	2.8451e-15	8	4.5191e-12

Tabla 2.18: Resultados para el Ejemplo 2.3.3 con $z^0 = (20, 5, 0, 0)$ para $\sigma = 1e-3$ y $\sigma = 1e-5$.

k	$\ F(z^k)\ _\infty$	k	$\ F'(z^k)\ _\infty$
\vdots	\vdots	\vdots	\vdots
4	9.5478e-07	3	3.8272e-03
5	1.9086e-08	4	4.2395e-07
6	3.8110e-10	5	3.5187e-09
7	3.0760e-14	6	1.3751e-11

Tabla 2.19: Resultados para el Ejemplo 2.3.3 con $z^0 = (13, 1, 1, 0)$ para $\sigma = 1e-4$ y $\sigma = 1e-5$.

Estos resultados fueron un atractivo importante a la hora de buscar una forma de modificar el algoritmo presentado en este capítulo para obtener una mayor velocidad de convergencia. La estrategia planteada para lograr este propósito es presentada en el próximo capítulo.

Capítulo 3

Un Método LP-Newton Secante Modificado

Siguiendo con el espíritu del capítulo anterior, el objetivo es resolver un sistema de ecuaciones con restricciones y la estrategia propuesta logra convergencia q -superlineal, a diferencia del método anterior, donde solo convergencia q -lineal fue obtenida. Un análisis del comportamiento local del método y ejemplos numéricos son presentados a lo largo de las secciones de este capítulo. Además se presenta una comparación de performance entre ambos métodos propuestos y el LPN.

Los resultados mostrados en este capítulo fueron presentados en [64].

3.1. Introducción

En este capítulo presentamos una variante del método qNLP para encontrar una solución de un sistema de ecuaciones no lineales con restricciones, donde la función involucrada es nuevamente continuamente diferenciable con derivada localmente Lipschitz continua.

Esta aproximación es una nueva modificación del método LPN. Recordemos que este método es capaz de encontrar una solución de un sistema de ecuaciones con posibles soluciones no aisladas, posibles funciones no suaves y alcanzando una tasa de convergencia q -cuadrática. Sin embargo, el cálculo de la derivada exacta o una aproximación de primer orden es requerida. A diferencia de ello, nuestro método evita cálculos de este tipo y utiliza aproximaciones secantes de cambio mínimo.

Como ya fue mencionado, este tipo de aproximaciones no alcanza velocidad de convergencia q -cuadrática como el método de Newton exacto, pero si convergencia q -superlineal, si y solo si la condición de Dennis-Moré se cumple. En particular esta condición es satisfecha por las matrices de actualización secante. Para más detalles teóricos y prácticos sobre los métodos quasi-Newton alentamos al lector a leer el survey [60].

La estrategia elegida es una adaptación de la regla estándar referida a los métodos de Newton truncados presentes en la literatura, que mide en algún sentido la inexactitud de la

aproximación lineal. En nuestro caso, el problema consiste en un sistema de ecuaciones sin restricciones donde la sucesión generada por el método satisface una condición de error dependiente de una sucesión forzante. Como podemos ver en [43], estos métodos son localmente q -superlinealmente convergentes a una solución aislada, si la sucesión forzante converge a cero y q -cuadráticamente convergente, si decrece con el mismo orden que el residuo.

Alternativamente, sistemas del tipo (2.1) pueden ser reformulados como problemas de cuadrados mínimos no lineales con restricciones. En estos casos, métodos de convergencia cuadrática fueron propuestos para soluciones aisladas [46, 50, 77] y no aisladas [28, 30, 78].

A continuación describimos el método propuesto. Aclaramos que la notación será la misma que en el capítulo anterior y las hipótesis para el conjunto Ω y la función involucrada son las mismas, es decir, la Hipótesis 5 del Capítulo 2 es satisfecha.

Algoritmo 9

Paso 0. Elegir $\sigma_0 > 0$, $z^0 \in \Omega$, $M_0 \in \mathbb{R}^{m \times n}$ y fijar $k = 0$.

Paso 1. Definir $z^{k+1} = z^k + d^k$ donde el par (d^k, γ_k) es un solución de

$$\begin{aligned} & \underset{d, \gamma}{\text{minimizar}} \quad \gamma \\ & \text{s. a.} \quad \|F(z^k) + M_k d\| \leq \sigma_k \gamma, \\ & \quad \quad \|d\| \leq \gamma, \\ & \quad \quad z^k + d \in \Omega. \end{aligned}$$

Paso 2. Si $F(z^{k+1}) = 0$, parar. Sino, calcular $M_{k+1} \in \mathbb{R}^{m \times n}$ y σ_{k+1} .

Paso 3. Fijar $k = k + 1$ e ir al Paso 1.

Este algoritmo se basa en el trabajo previo presentado en [63], donde el parámetro σ_k permanece fijo. Como antes, el subproblema involucrado en el algoritmo anterior es una adaptación de [32, problema (5.9)], que fue introducido para obtener un mejor escalado del problema.

En la siguiente sección desarrollamos la teoría de convergencia para el Algoritmo 9.

3.2. Análisis de Convergencia

Antes de comenzar con el análisis de convergencia, presentamos algunas hipótesis que serán necesarias para el desarrollo teórico.

Las matrices M_{k+1} involucradas en el Paso 2 del Algoritmo 9, serán actualizadas de la misma manera que en el Capítulo 2, es decir, son solución del problema de minimización (2.3), con $M = M_k$, $z = z^{k+1}$ y $w = z^k$. Además, asumiremos una propiedad de localización resumida en la siguiente hipótesis.

Hipótesis 7. Dada M_{k+1} definida en el Paso 2 del Algoritmo 9, existen constantes $c > 0$ y $0 < \alpha \leq 1$ tal que

$$\|N_{k+1} - M_{k+1}\| \leq c \|z^{k+1} - z^k\|^\alpha, \quad (3.1)$$

donde,

$$\begin{aligned} N_{k+1} &= \int_0^1 F'(z^k + t(z^{k+1} - z^k)) dt \\ &= \int_0^1 F'(z^{k+1} + s(z^k - z^{k+1})) ds. \end{aligned} \quad (3.2)$$

La matriz promedio (3.2) satisface $N_{k+1}(z^{k+1} - z^k) = F(z^{k+1}) - F(z^k)$, como ya habíamos mencionado anteriormente, y el siguiente error de linealización

$$\begin{aligned} \|F(z) - F(z^{k+1}) - N_{k+1}(z - z^{k+1})\| &\leq L \int_0^1 s \|z - z^k\| \|z - z^{k+1}\| ds \\ &= \frac{L_1}{2} \|z - z^k\| \|z - z^{k+1}\|, \end{aligned} \quad (3.3)$$

para cualquier z en un conjunto acotado B , donde L_1 es la constante de Lipschitz para F' .

De (3.1), claramente vemos que si la sucesión $\{z^k\}$ converge a un punto z^* , entonces la sucesión de matrices $\{M_k\}$ en el Algoritmo 9 deben aproximar asintóticamente a $F'(z^*)$. Es bien conocido que las actualizaciones secantes estándar pueden no satisfacer esta propiedad. Además (3.1) es más fuerte que la condición de Dennis-Moré. Esta condición nos asegura una aproximación asintótica en ciertas direcciones.

A continuación mostramos dos ejemplos extraídos de [22, Ejemplos 8.2.6 y 8.1.3] que muestran el comportamiento de las aproximaciones a la matriz Jacobiana de una función con respecto a la matriz Jacobiana en el punto solución.

Ejemplo 3.2.1. Consideremos la siguiente función $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$,

$$F(z) = \begin{bmatrix} z_1^2 + z_2^2 - 2 \\ e^{z_1-1} + z_2^3 - 2 \end{bmatrix}.$$

Esta función tiene una raíz en $z^* = (1, 1)$. Cuando aplicamos el método secante con la actualización de Broyden, comenzando en $z^0 = (1, 5, 2)$ y $M_0 = F'(z^0)$, la aproximación final a la matriz Jacobiana generada por este método es

$$M_{10} = \begin{bmatrix} 1.999137 & 2.021829 \\ 0.9995643 & 3.011004 \end{bmatrix}.$$

Mientras que

$$F'(z^*) = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}.$$

En el próximo ejemplo tal similitud no es obtenida.

Ejemplo 3.2.2. Sea $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por ,

$$F(z) = \begin{bmatrix} z_1 + z_2 - 3 \\ z_1^2 + z_2^2 - 9 \end{bmatrix}.$$

Las raíces en este caso son $z^* = (0, 3)$ y $z^{**} = (3, 0)$. Al aplicar el método de Broyden con $z^0 = (1, 5)$ y $M_0 = F'(z^0)$, se puede demostrar que la sucesión de matrices generadas tiende a M dada por

$$M = \begin{bmatrix} 1 & 1 \\ 1.5 & 7.5 \end{bmatrix}.$$

Mientras que

$$F'(z^*) = \begin{bmatrix} 1 & 1 \\ 0 & 6 \end{bmatrix}.$$

La condición de error (3.1) será usada solo para propósitos teóricos en la prueba de convergencia. En la Sección 3.3 mostramos algunos ejemplos numéricos, donde se puede alcanzar convergencia superlineal a pesar de la posible violación de (3.1).

En cuanto a la actualización de σ_{k+1} en el Paso 2, proponemos lo siguiente.

$$\sigma_{k+1} = \min \left\{ \sigma_0, \kappa \max \left\{ \|F(z^k)\|^\alpha, \|F(z^{k+1})\|^\alpha \right\} \right\} \quad (3.4)$$

para todo $k \geq 0$, con $\sigma_0 > 0$, $\kappa > 0$ y α satisfaciendo (3.1).

En lo que sigue, presentamos algunos resultados que nos llevarán a la convergencia superlineal del Algoritmo 9. Si bien siguen el espíritu de los resultados del capítulo anterior, hay modificaciones importantes en la estructuras de las demostraciones. La siguiente proposición nos da una condición de deterioro acotado para la sucesión de matrices, como la Proposición 2.2.4.

Proposición 3.2.3. Supongamos que la Hipótesis 5 vale. Sean $\{z^k\}$ y $\{M_k\}$ generadas por el Algoritmo 9. Si $z^{k+1}, z^k \in \mathcal{B}_{\frac{\varepsilon_0}{2}}(z^*)$ y $\|z^{k+1} - z^k\| \leq \text{dist}(z^k, \mathcal{Z})$ entonces

$$\|M_{k+1} - F'(\hat{z}^{k+1})\|_* \leq \|M_k - F'(\hat{z}^k)\|_* + 2L_1 \text{dist}(z^k, \mathcal{Z}) + L_1 \text{dist}(z^{k+1}, \mathcal{Z}),$$

donde $\hat{z} \in \mathcal{Z}$ satisface $\|\hat{z} - z\| = \text{dist}(z, \mathcal{Z})$.

Demostración. Definimos N_{k+1} como en (3.2). Como ya dijimos N_{k+1} es factible para el problema (2.3). Entonces por (2.4), tomando $M = M_k$, $M_+ = M_{k+1}$ y $N = N_{k+1}$ obtenemos

$$\|M_{k+1} - N_{k+1}\|_* \leq \|M_k - N_{k+1}\|_*. \quad (3.5)$$

Por otro lado, tenemos,

$$\|\hat{z}^k - z^*\| \leq \varepsilon_0 \quad \text{y} \quad \|\hat{z}^{k+1} - z^*\| \leq \varepsilon_0,$$

entonces, $z^k, z^{k+1}, \hat{z}^k, \hat{z}^{k+1} \in \mathcal{B}_{\varepsilon_0}(z^*)$. Luego,

$$\begin{aligned} \|N_{k+1} - F'(\hat{z}^{k+1})\|_* &\leq L_1 \int_0^1 \left(\|z^{k+1} - \hat{z}^{k+1}\| + s \|z^k - z^{k+1}\| \right) ds \\ &\leq L_1 \text{dist}(z^{k+1}, \mathcal{Z}) + \frac{1}{2} L_1 \text{dist}(z^k, \mathcal{Z}). \end{aligned}$$

De manera similar, obtenemos

$$\begin{aligned} \|N_{k+1} - F'(\hat{z}^k)\|_* &\leq L_1 \int_0^1 \left(\|z^k - \hat{z}^k\| + t\|z^{k+1} - z^k\| \right) dt \\ &\leq \frac{3}{2}L_1 \text{dist}(z^k, \mathcal{Z}). \end{aligned}$$

Entonces, usando (3.5) y las desigualdades de arriba, concluimos que

$$\begin{aligned} \|M_{k+1} - F'(\hat{z}^{k+1})\|_* &\leq \|M_{k+1} - N_{k+1}\|_* + \|N_{k+1} - F'(\hat{z}^{k+1})\|_* \\ &\leq \|M_k - F'(\hat{z}^k)\|_* + \|F'(\hat{z}^k) - N_{k+1}\|_* \\ &\quad + L_1 \text{dist}(z^{k+1}, \mathcal{Z}) + \frac{1}{2}L_1 \text{dist}(z^k, \mathcal{Z}) \\ &\leq \|M_k - F'(\hat{z}^k)\|_* + 2L_1 \text{dist}(z^k, \mathcal{Z}) + L_1 \text{dist}(z^{k+1}, \mathcal{Z}). \end{aligned}$$

■

A continuación presentamos el siguiente resultado, el cual es central a la hora de demostrar la convergencia del algoritmo.

Lema 3.2.4. *Supongamos que la Hipótesis 5 se cumple. Sean $\{z^k\}$, $\{M_k\}$ y $\{\sigma_k\}$ generadas por el Algoritmo 9 con M_{k+1} satisfaciendo (3.1) y σ_{k+1} satisfaciendo (3.4) para $\kappa \geq (c + L_1)\ell^\alpha$. Entonces existen constantes $r > 0$, $\varepsilon > 0$ y $\delta > 0$ tal que si*

$$\sigma_0 < \frac{1}{6\ell}, \quad z^0 \in \mathcal{B}_r(z^*) \cap \Omega \quad \text{y} \quad M_0 \in \mathcal{B}_{\frac{\delta}{2}}^*(F'(z^*)) \cap \mathcal{X},$$

entonces para todo $j \geq 0$,

- (a) $\|z^j - z^*\| \leq \varepsilon$,
- (b) $\|M_j - F'(z^*)\|_* \leq \delta$,
- (c) $\|F(z^j) + M_j(z^{j+1} - z^j)\| \leq \sigma_j \text{dist}(z^j, \mathcal{Z})$,
- (d) $\|z^{j+1} - z^j\| \leq \text{dist}(z^j, \mathcal{Z})$,
- (e) $\text{dist}(z^{j+1}, \mathcal{Z}) \leq \frac{1}{2} \text{dist}(z^j, \mathcal{Z})$.

Demostración. Dado que F es continuamente diferenciable, existe $\varepsilon_2 > 0$ tal que, si $w, z \in \mathcal{B}_{\varepsilon_2}(z^*)$ vale que

$$\|F(z) - F(w) - F'(z^*)(z - w)\| \leq \frac{1}{6\ell} \|z - w\|. \quad (3.6)$$

Por la Proposición 2.2.2 presentada en el capítulo anterior, dado σ_0 existen ε_1 y δ_1 constantes positivas, tal que

$$\|F(z^0) + M_0 d^0\| \leq \sigma_0 \text{dist}(z^0, \mathcal{Z}), \quad (3.7)$$

$$\|d^0\| \leq \text{dist}(z^0, \mathcal{Z}), \quad (3.8)$$

vale para $z^0 \in \mathcal{B}_{\varepsilon_1}(z^*)$, $M_0 \in \mathcal{B}_{\delta_1}^*(F'(z^*))$ y (d^0, γ_0) solución del problema (2.2). Además, recordemos que $\|M\| \leq \beta \|M\|_*$ para todo $M \in \mathbb{R}^{m \times n}$.

Definimos

$$\varepsilon = \min \left\{ 1, \varepsilon_2, \varepsilon_1, \frac{\varepsilon_0}{2} \right\}, \quad \delta = \min \left\{ \delta_1, \frac{1}{6\beta\ell} \right\} \quad \text{y} \quad r = \min \left\{ \frac{\varepsilon}{3}, \frac{\delta}{26L_1} \right\}. \quad (3.9)$$

Sea $z^0 \in \mathcal{B}_r(z^*) \cap \Omega$ y $M_0 \in \mathcal{B}_{\frac{\delta}{2}}^*(F'(z^*)) \cap \mathcal{X}$. Procediendo por inducción, tenemos que para $j = 0$ (a) y (b) valen, (c) y (d) valen por (3.7) y (3.8), respectivamente. Ahora, dado que

$$\begin{aligned} \|z^1 - z^*\| &\leq \|z^0 - z^*\| + \|z^1 - z^0\| \\ &\leq 2\|z^0 - z^*\| \\ &\leq 2r \\ &\leq \varepsilon_2, \end{aligned} \quad (3.10)$$

de (3.6) y (3.7) obtenemos

$$\begin{aligned} \|F(z^1)\| &\leq \|F(z^1) - F(z^0) - F'(z^*)(z^1 - z^0)\| + \|(M_0 - F'(z^*))(z^1 - z^0)\| \\ &\quad + \|F(z^0) + M_0(z^1 - z^0)\| \\ &\leq \frac{1}{6\ell} \|z^1 - z^0\| + \beta\delta \|z^1 - z^0\| + \sigma_0 \text{dist}(z^0, \mathcal{Z}) \\ &\leq \left(\frac{1}{6\ell} + \beta\delta + \sigma_0 \right) \text{dist}(z^0, \mathcal{Z}) \\ &\leq \frac{1}{2\ell} \text{dist}(z^0, \mathcal{Z}). \end{aligned}$$

Luego, dado que $z^1 \in \mathcal{B}_{\varepsilon_0}(z^*) \cap \Omega$, (v) se sigue de

$$\text{dist}(z^1, \mathcal{Z}) \leq \ell \|F(z^1)\| \leq \frac{1}{2} \text{dist}(z^0, \mathcal{Z}).$$

Con lo cual, (a) - (e) vale para $j = 0$.

Ahora, supongamos que (a) - (e) vale para todo $j \leq k - 1$. En lo que sigue, usaremos la siguiente relación: para cualquier $0 \leq p \leq q \leq k - 1$, vale que

$$\begin{aligned} \sum_{i=p}^q \|z^{i+1} - z^i\| &\leq \sum_{i=p}^q \text{dist}(z^i, \mathcal{Z}) \\ &\leq \sum_{i=p}^q \frac{1}{2^{i-p}} \text{dist}(z^p, \mathcal{Z}) \\ &\leq 2 \text{dist}(z^p, \mathcal{Z}). \end{aligned} \quad (3.11)$$

Entonces, usando (3.11) para $p = 0$ y $q = k - 1$,

$$\begin{aligned} \|z^k - z^*\| &\leq \|z^k - z^0\| + \|z^0 - z^*\| \\ &\leq \sum_{i=0}^{k-1} \|z^{i+1} - z^i\| + \|z^0 - z^*\| \\ &\leq 2 \text{dist}(z^0, \mathcal{Z}) + \|z^0 - z^*\| \\ &\leq 3\|z^0 - z^*\| \\ &\leq \varepsilon. \end{aligned} \quad (3.12)$$

Por lo tanto, (a) vale. Además, por la Proposición 2.2.4, (e), y (3.11) con $p = 0$ y $q = k - 1$, obtenemos

$$\begin{aligned}
\|M_k - F'(\hat{z}^k)\|_* &\leq \|M_{k-1} - F'(\hat{z}^{k-1})\|_* + 2L_1 \text{dist}(z^{k-1}, \mathcal{Z}) + L_1 \text{dist}(z^k, \mathcal{Z}) \\
&\leq \|M_{k-1} - F'(\hat{z}^{k-1})\|_* + \frac{5}{2}L_1 \text{dist}(z^{k-1}, \mathcal{Z}) \\
&\leq \|M_0 - F'(\hat{z}^0)\|_* + \frac{5}{2}L_1 \sum_{i=0}^{k-1} \text{dist}(z^i, \mathcal{Z}) \\
&\leq \|M_0 - F'(\hat{z}^0)\|_* + 5L_1 \text{dist}(z^0, \mathcal{Z}).
\end{aligned}$$

Entonces,

$$\begin{aligned}
\|M_k - F'(z^*)\|_* &\leq \|M_k - F'(\hat{z}^k)\|_* + \|F'(\hat{z}^k) - F'(z^*)\|_* \\
&\leq \|M_0 - F'(\hat{z}^0)\|_* + 5L_1 \|z^0 - z^*\| + 2L_1 \|z^k - z^*\| \\
&\leq \|M_0 - F'(z^*)\|_* + \|F'(z^*) - F'(\hat{z}^0)\|_* + 11L_1 \|z^0 - z^*\| \\
&\leq \|M_0 - F'(z^*)\|_* + 2L_1 \|z^0 - z^*\| + 11L_1 \|z^0 - z^*\| \\
&\leq \frac{\delta}{2} + 13L_1 r \\
&\leq \delta,
\end{aligned}$$

donde usamos que $\|\hat{z} - z^*\| \leq 2\|z - z^*\|$ en la segunda y cuarta desigualdad y (3.12) en la tercera. Luego (b) vale para $j = k$.

Ahora, por (3.3) y (3.1) tenemos que

$$\begin{aligned}
\|F(z^k) + M_k(\hat{z}^k - z^k)\| &= \|F(\hat{z}^k) - F(z^k) - M_k(\hat{z}^k - z^k)\| \\
&\leq \|F(\hat{z}^k) - F(z^k) - N_k(\hat{z}^k - z^k)\| \\
&\quad + \|N_k - M_k\| \|\hat{z}^k - z^k\| \\
&\leq \left(\frac{L_1}{2} \|\hat{z}^k - z^{k-1}\| + c \|z^k - z^{k-1}\|^\alpha \right) \|\hat{z}^k - z^k\| \\
&\leq \left(\frac{L_1}{2} \text{dist}(z^k, \mathcal{Z}) + \frac{L_1}{2} \text{dist}(z^{k-1}, \mathcal{Z}) \right. \\
&\quad \left. + c \text{dist}(z^{k-1}, \mathcal{Z})^\alpha \right) \text{dist}(z^k, \mathcal{Z}) \\
&\leq \kappa \max \left\{ \|F(z^k)\|^\alpha, \|F(z^{k-1})\|^\alpha \right\} \text{dist}(z^k, \mathcal{Z}) \\
&\leq \sigma_k \text{dist}(z^k, \mathcal{Z}).
\end{aligned}$$

Entonces,

$$\|F(z^k) + M_k(\hat{z}^k - z^k)\| \leq \sigma_k \text{dist}(z^k, \mathcal{Z}) \quad \text{y} \quad \|\hat{z}^k - z^k\| \leq \text{dist}(z^k, \mathcal{Z}).$$

Luego, (c) y (d) se cumplen, ya que z^{k+1} satisface

$$\|F(z^k) + M_k(z^{k+1} - z^k)\| \leq \sigma_k \text{dist}(z^k, \mathcal{Z}) \quad \text{y} \quad \|z^{k+1} - z^k\| \leq \text{dist}(z^k, \mathcal{Z}).$$

Veamos que (e) también vale. Procediendo como en el caso $j = 0$ y usando que $\sigma_k \leq \sigma_0$, tenemos que

$$\|F(z^{k+1})\| \leq \frac{1}{2\ell} \text{dist}(z^k, \mathcal{Z}).$$

Por otro lado, notemos que

$$\begin{aligned} \|z^{k+1} - z^*\| &\leq \|z^k - z^*\| + \|z^{k+1} - z^k\| \\ &\leq 2\|z^k - z^*\| \\ &\leq 2\varepsilon \\ &\leq \varepsilon_0. \end{aligned}$$

Entonces, dado $z^{k+1} \in \mathcal{B}_{\varepsilon_0}(z^*) \cap \Omega$,

$$\text{dist}(z^{k+1}, \mathcal{Z}) \leq \frac{1}{2} \text{dist}(z^k, \mathcal{Z}).$$

Finalmente, (a)-(e) valen para todo $j \geq 0$. ■

Los resultados anteriores junto con las hipótesis mencionadas nos permiten llegar al siguiente teorema, que demuestra la convergencia del algoritmo a un punto solución de (2.1) y posteriormente, verificar que la velocidad de convergencia es q -superlineal.

Teorema 3.2.5. *Tomemos $\sigma_0 < \frac{1}{6\ell}$. Bajo las hipótesis del Lema 3.2.4, si*

$$z^0 \in \mathcal{B}_r(z^*) \cap \Omega \quad \text{y} \quad M_0 \in \mathcal{B}_{\frac{\delta}{2}}^*(F'(z^*)) \cap \mathcal{X},$$

entonces

(a) *la sucesión $\{z^k\}$ converge a algún punto $\bar{z} \in \mathcal{Z}$,*

(b) *la sucesión $\{\|M_k - F'(\bar{z})\|_*\}$ converge,*

(c) *la sucesión forzante $\{\sigma_k\}$ converge a cero .*

Demostración.

(a) Por (d) y (e) del Lema 3.2.4 y (3.11) tenemos que

$$\begin{aligned} \|z^{k+j} - z^j\| &\leq \sum_{i=j}^{k+j-1} \|z^{i+1} - z^i\| \\ &\leq 2 \text{dist}(z^j, \mathcal{Z}) \\ &\leq 2 \frac{1}{2^j} \text{dist}(z^0, \mathcal{Z}). \end{aligned}$$

Entonces $\{z^k\}$ es de Cauchy y dado que \mathcal{Z} es un conjunto cerrado, la sucesión converge a algún $\bar{z} \in \mathcal{Z}$, por lo tanto (a) vale. Además, tomando límite para $k \rightarrow \infty$ obtenemos

$$\|z^j - \bar{z}\| \leq 2 \text{dist}(z^j, \mathcal{Z}), \quad (3.13)$$

para todo $j \geq 0$.

(b) Para probar la convergencia de $\{\|M_k - F'(\bar{z})\|_*\}$, notemos que por la Proposición 3.2.3 y (3.11), para $k > j$,

$$\begin{aligned} \|M_k - F'(\hat{z}^k)\|_* &\leq \|M_j - F'(\hat{z}^j)\|_* + \frac{5}{2}L_1 \sum_{i=j}^{k-1} \text{dist}(z^i, \mathcal{Z}) \\ &\leq \|M_j - F'(\hat{z}^j)\|_* + 5L_1 \text{dist}(z^j, \mathcal{Z}). \end{aligned}$$

Dado que $\hat{z}^k \rightarrow \bar{z}$ y F' es continua, tenemos que

$$\limsup_{k \rightarrow \infty} \|M_k - F'(\bar{z})\|_* \leq \|M_j - F'(\hat{z}^j)\|_* + 5L_1 \text{dist}(z^j, \mathcal{Z}).$$

Ahora, tomando límite para $j \rightarrow \infty$ en el lado derecho de la desigualdad anterior llegamos a

$$\limsup_{k \rightarrow \infty} \|M_k - F'(\bar{z})\|_* \leq \liminf_{j \rightarrow \infty} \|M_j - F'(\bar{z})\|_*.$$

Entonces (b) vale.

(c) Para la última prueba, como $z^k \rightarrow \bar{z} \in \mathcal{Z}$ y F es una función continua, entonces $\max\{\|F(z^k)\|, \|F(z^{k-1})\|\}^\alpha \rightarrow 0$. Lo cual implica que, $\sigma_k \rightarrow 0$ como $k \rightarrow \infty$. ■

Finalmente, podemos mostrar la convergencia superlineal del método.

Corolario 3.2.6. *Adicional a los resultados del Teorema 3.2.5, vale que $\{z^k\}$ converge q -superlinealmente a \bar{z} y $\{\|F(z^k)\|\}$ converge q -superlinealmente a cero.*

Demostración. Por (2.4), la matriz N_{k+1} definida en (3.2) satisface

$$\|M_k - M_{k+1}\|_*^2 \leq \|M_k - N_{k+1}\|_*^2 - \|M_{k+1} - N_{k+1}\|_*^2.$$

Por definición de N_{k+1} y la continuidad de F' tenemos que $N_k \rightarrow F'(\bar{z})$, concluyendo por el Teorema 3.2.5 (b) que

$$M_{k+1} - M_k \rightarrow 0.$$

Por otro lado, como $M_{k+1}(z^{k+1} - z^k) = F(z^{k+1}) - F(z^k)$, entonces

$$\begin{aligned} \|F(z^{k+1})\| &= \|F(z^k) + M_{k+1}(z^{k+1} - z^k)\| \\ &\leq \sigma_k \text{dist}(z^k, \mathcal{Z}) + \|(M_{k+1} - M_k)(z^{k+1} - z^k)\| \\ &\leq (\sigma_k + \|M_{k+1} - M_k\|)\text{dist}(z^k, \mathcal{Z}) \\ &\leq (\sigma_k + \|M_{k+1} - M_k\|)\ell\|F(z^k)\|. \end{aligned}$$

Luego,

$$\frac{\|F(z^{k+1})\|}{\|F(z^k)\|} \leq \ell(\sigma_k + \|M_{k+1} - M_k\|).$$

Como $\sigma_k \rightarrow 0$ y $\|M_{k+1} - M_k\| \rightarrow 0$ entonces

$$\frac{\|F(z^{k+1})\|}{\|F(z^k)\|} \rightarrow 0.$$

Lo cual nos dice que la sucesión $\|F(z^k)\|$ tiende a cero q -superlinealmente.

Por otro lado, por (3.13) y el hecho de que $\|F(z^k)\| = \|F(z^k) - F(\bar{z})\| \leq L_0\|z^k - \bar{z}\|$, llegamos a

$$\frac{\|z^{k+1} - \bar{z}\|}{\|z^k - \bar{z}\|} \leq \frac{2L_0 \text{dist}(z^{k+1}, \mathcal{Z})}{\|F(z^k)\|} \leq \frac{2L_0 \ell \|F(z^{k+1})\|}{\|F(z^k)\|}.$$

Por lo tanto,

$$\frac{\|z^{k+1} - \bar{z}\|}{\|z^k - \bar{z}\|} \rightarrow 0.$$

Concluyendo que $z^k \rightarrow \bar{z}$ q -superlinealmente. Y así queda demostrado el último resultado de esta sección. ■

En resumen, hemos propuesto un algoritmo para resolver sistemas de ecuaciones no lineales con restricciones y posibles soluciones no aisladas, a través de una estrategia quasi-Newton, la cual no requiere el cálculo de derivadas de la función involucrada. Además mostramos resultados de convergencia superlineal para el algoritmo. En lo que sigue, presentamos algunos ejemplos numéricos para verificar la performance del método propuesto.

3.3. Ejemplos Numéricos

En esta sección analizamos numéricamente el comportamiento local del método propuesto y finalizamos con una comparación entre nuestro algoritmo y otros presentados en este trabajo.

En algunos casos, las hipótesis del Teorema 3.2.5 no son satisfechas, pero mostramos que de igual forma el Algoritmo 9 posee un buen desempeño. Como la mayoría de los problemas de la Sección 2.3, los que presentamos aquí tienen un conjunto solución de puntos localmente no aislados, además hay casos donde el conjunto \mathcal{Z} coincide con la frontera del conjunto factible, o no es convexo.

Nuevamente, el Algoritmo 9 fue implementado en Octave, usando el método simplex de la función *glpk* para resolver el subproblema (2.2). Los criterios de parada son los ya mencionados en el Capítulo 2. Las matrices son actualizadas según Broyden, con $M_0 = F'(z^0)$ y dado σ_0 usamos

$$\sigma_k = \min \left\{ \sigma_0, 100 \max \left\{ \|F(z^{k-1})\|, \|F(z^k)\| \right\} \right\}$$

y fijamos $\alpha = 1$. Una buena elección general para σ_0 , en términos del punto inicial z^0 y la dimensión del problema, es $\sigma_0 = \frac{\|F(z^0)\|}{n^2}$, cuya motivación viene dada principalmente de experimentos numéricos.

En cada ejemplo mostramos tablas que muestran el residuo en las últimas iteraciones para diferentes valores de σ_0 y puntos iniciales z^0 . Más aún, la solución \bar{z} encontrada por el algoritmo es presentada en cada caso.

En el primer problema, usamos diferentes valores de σ_0 para mostrar que el algoritmo tiene un comportamiento similar en todos estos casos.

Ejemplo 3.3.1. Consideremos el problema (2.1) con

$$F(z) = \begin{bmatrix} z_2^2 \\ z_2(1 + z_1^2) \end{bmatrix} \quad \text{con } z \in \Omega,$$

donde $\Omega = \mathbb{R} \times \mathbb{R}_+$. En este caso, F provee una cota de error local en una vecindad de cualquier solución \bar{z} (ver [27]). Para este ejemplo, el conjunto solución esta dado por

$$\mathcal{Z} = \{(u, 0) \mid u \in \mathbb{R}\}.$$

Como podemos ver, el conjunto solución coincide con la frontera de Ω y aún así, el algoritmo es capaz de encontrar una solución.

La Tabla 3.1 muestra las últimas iteraciones del algoritmo considerando $z^0 = (1, 0.5)$ (izq.) y $z^0 = (2, 0.9)$ (der.) con $\sigma_0 = \frac{\|F(z^0)\|}{2^2}$. Los resultados obtenidos son $\bar{z} = (0.4808, 0)$ y $\bar{z} = (1.1, 0)$ respectivamente.

k	$\ F(z^k)\ _\infty$	k	$\ F(z^k)\ _\infty$
\vdots	\vdots	\vdots	\vdots
7	1.9799e-05	9	3.1113e-04
8	3.4451e-07	10	2.2326e-05
9	4.6686e-10	11	1.6952e-07
10	0	12	0

Tabla 3.1: Resultados para el Ejemplo 3.3.1.

En la Tabla 3.2 mostramos los resultados para $z^0 = (1, 0.5)$ con $\sigma_0 = 0.2$ (izq.) y $\sigma_0 = 0.8$ (der.). Las soluciones encontradas son $\bar{z} = (0.4561, 0)$ y $\bar{z} = (0.5, 0)$.

k	$\ F(z^k)\ _\infty$	k	$\ F(z^k)\ _\infty$
\vdots	\vdots	\vdots	\vdots
6	1.2846e-04	9	6.1965e-05
7	8.8638e-06	10	2.3462e-06
8	1.0073e-07	11	8.7003e-09
9	0	12	0

Tabla 3.2: Resultados para el Ejemplo 3.3.1.

La Tabla 3.3 muestra los resultados para $z^0 = (2, 0.9)$ con $\sigma_0 = 0.1$ (izq.) y $\sigma_0 = 1.8$ (der.). Las soluciones encontradas por el algoritmo son $\bar{z} = (7.5378e-01, 1.0001e-11)$ y $\bar{z} = (1.1, 0)$.

k	$\ F(z^k)\ _\infty$	k	$\ F(z^k)\ _\infty$
\vdots	\vdots	\vdots	\vdots
7	2.9894e-04	11	9.1703e-05
8	3.1200e-05	12	3.0476e-06
9	1.0125e-06	13	7.0199e-09
10	1.5684e-11	14	0

Tabla 3.3: Resultados para el Ejemplo 3.3.1.

De aquí en adelante, se fija $\sigma_0 = \frac{\|F(z^0)\|}{n^2}$ para todos los ejemplos.

Ejemplo 3.3.2. Consideremos el problema (2.1) donde

$$F(z) = \begin{bmatrix} 100 - (z_1 - 5)^2 - (z_2 - 5)^2 + z_3 \\ (z_1 - 6)^2 + (z_2 - 5)^2 - 82.81 + z_4 \end{bmatrix} \quad \text{con } z \in \Omega,$$

donde $\Omega = [13, 100] \times [0, 100] \times \mathbb{R}_+^2$.

Los resultados obtenidos son mostrados en la Tabla 3.4 para $z^0 = (15, 2, 1, 0)$ (izq.) y $z^0 = (16, 1, 1, 1)$ (der.). En este caso, las soluciones encontradas por el algoritmo son $\bar{z} = (14.69560, 2.31737, 1.20121, 0)$ y $\bar{z} = (14.74775, 2.49264, 1.30550, 0)$, respectivamente.

k	$\ F(z^k)\ _\infty$	k	$\ F(z^k)\ _\infty$
\vdots	\vdots	\vdots	\vdots
7	1.0246e-05	20	3.4446e-05
8	1.6960e-07	21	1.2344e-06
9	2.2727e-10	22	5.6915e-09
10	1.0214e-14	23	9.5213e-13

Tabla 3.4: Resultados para el Ejemplo 3.3.2.

Ejemplo 3.3.3. Consideremos el siguiente problema escrito como (2.1) con

$$F(z) = \begin{bmatrix} -z_1 - z_2 + 1 + z_3 \\ -z_1^2 - z_2^2 + 1 + z_4 \\ -9z_1^2 - z_2^2 + 9 + z_5 \\ -z_1^2 + z_2 + z_6 \\ -z_2^2 + z_1 + z_7 \end{bmatrix} \quad \text{con } z \in \Omega,$$

donde $\Omega = [-50, 50] \times [-50, 50] \times \mathbb{R}_+^5$. En este caso, el conjunto solución no solo es no aislado sino que tampoco es conexo.

Algunos resultados numéricos son mostrados en la Tabla 3.5, para los puntos iniciales $z^0 = (1, 1, 1.5, 0, 0, 0, 0)$ (izq.) y $z^0 = (1, 1, 1, 1.5, 0, 0, 0)$ (der.). En ambos casos, la solución encontrada por el algoritmo es $\bar{z} = (1, 1, 1, 1, 1, 0, 0)$.

k	$\ F(z^k)\ _\infty$	k	$\ F(z^k)\ _\infty$
\vdots	\vdots	\vdots	\vdots
9	5.6581e-06	8	1.5720e-06
10	3.7141e-07	9	4.7926e-08
11	4.3604e-09	10	1.5759e-10
12	3.8156e-12	11	2.7889e-13

Tabla 3.5: Resultados para el Ejemplo 3.3.3.

Ejemplo 3.3.4. Consideremos el problema de encontrar las raíces del polinomio multivariable

$$F(z) = z_1(1 + z_2^2) + z_3^4 - 4 - 3\sqrt{2} \quad \text{con } z \in \Omega,$$

en la región $\Omega = [-10, 10] \times [-10, 10] \times [-10, 10]$.

Claramente, es un problema del tipo (2.1). En la Tabla 3.6 se muestran los resultados obtenidos tomando $z^0 = (2, 1, 1)$ (izq.) y $z^0 = (1, 1, 1.2)$ (der.). Las soluciones encontradas son $\bar{z} = (2.2522, 1.2522, 1.2522)$ y $\bar{z} = (1.2836, 1.2836, 1.4836)$, respectivamente.

k	$\ F(z^k)\ _\infty$	k	$\ F(z^k)\ _\infty$
\vdots	\vdots	\vdots	\vdots
4	1.6499e-04	4	2.2919e-04
5	3.4913e-06	5	5.0863e-06
6	3.5509e-09	6	6.0718e-09
7	9.7700e-14	7	1.5277e-13

Tabla 3.6: Resultados para el Ejemplo 3.3.4.

Ejemplo 3.3.5. Este problema es una nueva formulación del Ejemplo 1.3.4, presentado en la Sección 1.3. Esta nueva formulación fue hecha para que la función F cumpla con la hipótesis de diferenciabilidad que requerimos en el análisis de convergencia del algoritmo.

Consideremos el problema (2.1) con

$$F(z) = \begin{bmatrix} z_4 + z_5 - z_6 - z_9 \\ z_4 + z_2 + z_3 - z_7 - z_9 \\ z_2 + z_3 - z_9 \\ z_1 + z_2 - z_8 \\ z_1 + z_{10} \\ -z_1 + z_{11} \\ 1 - z_2 + z_{12} \\ -z_4 + z_{13} \\ -z_1 - z_2 - z_3 + z_{14} \\ z_5 z_{10} \\ z_6 z_{11} \\ z_7 z_{12} \\ z_8 z_{13} \\ z_9 z_{14} \end{bmatrix} \quad \text{con } z \in \Omega,$$

$$y \Omega = \mathbb{R}^4 \times \mathbb{R}_+^5 \times \mathbb{R}_+^5.$$

Los resultados numéricos son mostrados en la Tabla 3.7. Fijado como punto inicial $z^0 = \frac{3}{2}(0, 1.1, -1, 0, 0.1, 0.1, 0, 1, 0, 0, 0, 0.1, 0, 0)$ (izq.) la solución encontrada por el algoritmo fue $\bar{z} = (0, t, -t, 0, 0, 0, 0, t, 0, 0, 0, t - 1, 0, 0)$ donde $t = 1.40037$. Para el punto $z^0 = 4(0, 1, -0.6, 0, 0.2, 0.3, 0, 1, 0, 0, 0, 0, 0, 0)$ (der.) la solución encontrada fue $\bar{z} = (0, s, -s, 0, 0, 0, 0, s, 0, 0, 0, s - 1, 0, 0)$ donde $s = 2.51554$.

k	$\ F(z^k)\ _\infty$	k	$\ F(z^k)\ _\infty$
1	7.4264e-04	2	6.3738e-04
2	2.2028e-06	3	1.3071e-05
3	6.5338e-09	4	2.6806e-07
4	2.2204e-16	5	4.4409e-16

Tabla 3.7: Resultados para el Ejemplo 3.3.5.

Ejemplo 3.3.6. El siguiente ejemplo fue elegido de una lista propuesta en [68] sobre problemas de complementariedad y proviene de una aplicación en economía al problema de equilibrio económico [49].

El problema fue reformulado como en (2.1) con

$$F(z) = \begin{bmatrix} 3z_1^2 + 2z_1z_2 + 2z_2^2 + z_3 + 3z_4 - 6 - z_5 \\ 2z_1^2 + z_2^2 + z_1 + 10z_3 + 2z_4 - 2 - z_6 \\ 3z_1^2 + z_1z_2 + 2z_2^2 + 2z_3 + 9z_4 - 9 - z_7 \\ z_1^2 + 3z_2^2 + 2z_3 + 3z_4 - 3 - z_8 \\ z_1z_5 \\ z_2z_6 \\ z_3z_7 \\ z_4z_8 \end{bmatrix} \quad \text{con } z \in \Omega,$$

donde $\Omega = \mathbb{R}_+^8$.

A continuación presentamos en la Tabla 3.8 los resultados obtenidos por el Algoritmo 9 para $z^0 = (1, 0, 0, 0, 0, 1, 0, 0)$ (izq.) y $z^0 = (1.5, 0, 0, 0, 0.5, 0.3, 0.8, 0, 0)$ (der.). Los resultados obtenidos para estos casos son $\bar{z} = (1.22474, 0, 0, 0.5, 0, 3.22474, 0, 0)$ y $\bar{z} = (1.2247, 3.75\text{e-}14, 9.18\text{e-}19, 0.5, 4.89\text{e-}19, 3.2247, 6.31\text{e-}13, 1.39\text{e-}13)$.

k	$\ F(z^k)\ _\infty$	k	$\ F(z^k)\ _\infty$
\vdots	\vdots	\vdots	\vdots
17	4.4815e-08	11	1.0745e-07
18	4.4251e-09	12	1.5531e-08
19	3.8971e-10	13	6.2675e-10
20	2.0015e-11	14	4.8505e-12

Tabla 3.8: Resultados para el Ejemplo 3.3.6.

A continuación realizamos una comparación de algunos de los métodos presentados en este trabajo.

3.3.1. Comentarios

Presentamos un último ejemplo donde comparamos el tiempo computacional para los método LP quasi-Newton secante modificado presentado en este capítulo, el método LP-Newton modificado y el LP-Newton propuesto en [25].

Ejemplo 3.3.7. Consideremos el problema (2.1) donde $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ está dada por

$$F_i(z) = z_i - \frac{2}{n} \sum_{j=1}^n z_j - 1 + (z_i + 1)^2,$$

para todo $i = 1, \dots, n$, en la región $\Omega = [-10, 10]^n$. En este caso, comparamos el tiempo computacional tomado por los dos métodos descritos en los Capítulos 2 y 3, y el presentado en [25]. Para ello, tomamos diferentes valores de n , para SMLPN fijamos $\kappa = 10$ y $\sigma = 1e - 2$ para qNLP. Los resultados son mostrados en la Tabla 3.9.

Método	LPN	qNLP	SMLPN
n	Tiempo (s)		
400	11.236	15.680	13.252
700	102.98	109.18	90.332
900	155.40	156.10	142.18
1100	297.26	330.71	245.55
1500	888.11	964.07	672.26

Tabla 3.9: Tiempo.

En todos los casos los algoritmos encontraron alguna solución del problema. Como podemos ver, cuando la dimensión del problema no es demasiado grande, el método LP-Newton toma menos tiempo en encontrar una solución. Cuando la dimensión comienza a aumentar, el método SMLPN lo hace en menor tiempo. La principal razón por lo que esto ocurre es que este algoritmo no necesita calcular la matriz Jacobiana de la función en cada iteración sino que utiliza la actualización de Broyden (1.24), lo que lo hace computacionalmente menos costoso. Con respecto al método qNLP, el tiempo computacional es siempre mayor que en el resto de los algoritmos. Si bien, los métodos presentados en los Capítulos 2 y 3 comparten el mismo espíritu, la ventaja del algoritmo SMLPN respecto a qNLP es la posibilidad de actualizar el parámetro σ_k en cada iteración.

Los resultados presentados en esta sección muestran que el método SMLPN es una valiosa herramienta para la solución de sistemas de ecuaciones no lineales, aún cuando algunas de las hipótesis requeridas para el análisis teórico de convergencia puedan no cumplirse.

Con este ejemplo concluimos el Capítulo 3 de este trabajo.

Capítulo 4

Conclusiones Generales

En este trabajo hemos presentado una serie de métodos basados en el método clásico de Newton para la solución de sistemas de ecuaciones y problemas de optimización. Como pudimos ver en la Subsección 1.1.3, el método de Newton tal cual lo conocemos puede ser aplicado a problemas de optimización sin restricciones o con restricciones de igualdad. Cuando nos encontramos frente a problemas con restricciones mixtas como (1.16), dicho método debe ser modificado y terminamos resolviendo en cada iteración un problema de programación cuadrática, como fue mostrado

Otra forma de resolver estos problemas es planteando el sistema KKT asociado y pensarlo como uno de ecuaciones no lineales sin restricciones (como en (1.20)) o con restricciones (como en (1.38)). Recordemos que en cualquiera de los casos el problema es no diferenciable, por lo tanto, una de las opciones para resolverlo es a través de los métodos de Newton Generalizados, que hacen uso de derivadas generalizadas de la función involucrada. En la Sección 1.3 presentamos el método LPN propuesto en [25] para la solución de sistemas de ecuaciones no lineales con restricciones y mostramos que es una herramienta poderosa para resolver sistemas KKT que surgen, entre otros, de plantear las condiciones de optimalidad de problemas de optimización.

Nuestra principal contribución fue el desarrollo de dos métodos para la solución de sistemas de ecuaciones con restricciones y posibles soluciones no aisladas. En el Capítulo 2 presentamos el método qNLP, el cual se basa en una estrategia quasi-Newton del método LPN. En este caso, la derivada o derivada generalizada es reemplazada por aproximaciones secantes en cada iteración del algoritmo. Al tratar con posibles soluciones locales no aisladas, perdemos la no singularidad de la matriz Jacobiana en alguna solución, por lo que este supuesto es reemplazado por una condición de cota de error local, que nos da una estimación de la distancia entre un punto y el conjunto solución. Presentamos el análisis de convergencia local del método y demostramos que posee convergencia q -lineal.

En el Capítulo 3 introducimos el método SMLPN, que intenta mejorar la velocidad de convergencia del qNLP. En este caso obtuvimos convergencia q -superlineal, sin embargo necesitamos de una hipótesis extra. Para la actualización de las matrices debimos pedir una condición de localización dada por (3.1). Si bien esta hipótesis es muy fuerte, y como

ya comentamos, muchas de las actualizaciones de cambio mínimo no la cumplen, solo fue utilizada con fines teóricos. Tanto para este método como para el anterior, presentamos algunos resultados numéricos. Al finalizar mostramos una tabla comparando la performance de los algoritmos LPN, qNLP y SMLPN. Si bien no se trata de una función demasiado complicada, el problema aquí son las dimensiones. Corrimos los algoritmos para diferentes valores de n (dimesión del problema) y pudimos observar que mientras mayor era la dimensión, el método SMLPN podía encontrar una solución en menor tiempo, como esperábamos. Estos resultados son mostramos en la Tabla 3.9.

Bibliografía

- [1] J. Avila and P. Concus. Update methods for highly structured systems of nonlinear equations. *SIAM Journal on Numerical Analysis*, 16(2):260–269, 1979.
- [2] J. Barzilai and M. A. Dempster. Measuring rates of convergence of numerical algorithms. *Journal of Optimization Theory and Applications*, 78(1):109–125, 1993.
- [3] D. P. Bertsekas. *Nonlinear programming*. Athena scientific Belmont, 1999.
- [4] P. T. Boggs and J. W. Tolle. Sequential quadratic programming. *Acta numerica*, 4:1–51, 1995.
- [5] P. T. Boggs, J. W. Tolle, and P. Wang. On the local convergence of quasi-Newton methods for constrained optimization. *SIAM Journal on Control and Optimization*, 20(2):161–171, 1982.
- [6] J. F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29(2):161–186, 1994.
- [7] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical optimization: theoretical and practical aspects*. Springer Science & Business Media, 2006.
- [8] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2013.
- [9] C. G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of computation*, 19(92):577–593, 1965.
- [10] C. G. Broyden. The convergence of a class of double-rank minimization algorithms 1. General Considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, 1970.
- [11] C. G. Broyden, J. Dennis Jr, and J. J. Moré. On the local and superlinear convergence of quasi-Newton methods. *IMA Journal of Applied Mathematics*, 12(3):223–245, 1973.
- [12] S. Buhmiller and N. Krejić. A new smoothing quasi-Newton method for nonlinear complementarity problems. *Journal of Computational and Applied Mathematics*, 211(2):141–155, 2008.

- [13] J. S. Chen and S. Pan. A family of NCP functions and a descent method for the nonlinear complementarity problem. *Computational Optimization and Applications*, 40(3):389–404, 2008.
- [14] T. F. Coleman and J. J. More. Software for estimating sparse Jacobian matrices. Technical report, CM-P00069197, 1982.
- [15] T. F. Coleman and J. J. Moré. Estimation of sparse Jacobian matrices and graph coloring blems. *SIAM Journal on Numerical Analysis*, 20(1):187–209, 1983.
- [16] A. R. Curtis, M. J. Powell, and J. K. Reid. On the estimation of sparse Jacobian matrices. *J. Inst. Math. Appl*, 13(1):117–120, 1974.
- [17] W. C. Davidon. Variable metric method for minimization. *SIAM Journal on Optimization*, 1(1):1–17, 1991.
- [18] D. Decker, H. Keller, and C. Kelley. Convergence rates for Newton’s method at singular points. *SIAM Journal on Numerical Analysis*, 20(2):296–314, 1983.
- [19] D. Decker and C. Kelley. Newton’s method at singular points. I. *SIAM Journal on Numerical Analysis*, 17(1):66–70, 1980.
- [20] R. S. Dembo, S. C. Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM Journal on Numerical Analysis*, 19(2):400–408, 1982.
- [21] J. E. Dennis, Jr and R. B. Schnabel. Least change secant updates for quasi-Newton methods. *SIAM Review*, 21(4):443–459, 1979.
- [22] J. E. Dennis Jr and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16. SIAM, 1996.
- [23] P. Deuffhard, R. Freund, and A. Walter. Fast secant methods for the iterative solution of large nonsymmetric linear systems. *IMPACT of Computing in Science and Engineering*, 2(3):244–276, 1990.
- [24] A. L. Dontchev and R. T. Rockafellar. Implicit functions and solution mappings. *Springer Monogr. Math.*, 2009.
- [25] F. Facchinei, A. Fischer, and M. Herrich. An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. *Mathematical Programming*, 146(1-2):1–36, 2014.
- [26] F. Facchinei, A. Fischer, C. Kanzow, and J. M. Peng. A simply constrained optimization reformulation of KKT systems arising from variational inequalities. *Applied Mathematics and Optimization*, 40(1):19–37, 1999.
- [27] J. Fan. Convergence rate of the trust region method for nonlinear equations under local error bound condition. *Computational Optimization and Applications*, 34(2):215–227, 2006.

- [28] J. Fan and Y. Yuan. On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. *Computing*, 74(1):23–39, 2005.
- [29] D. Fernández. A quasi-Newton strategy for the sSQP method for variational inequality and optimization problems. *Mathematical Programming*, 137(1-2):199–223, 2013.
- [30] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Mathematical Programming*, 94(1):91–124, 2002.
- [31] A. Fischer, M. Herrich, A. F. Izmailov, and M. V. Solodov. Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions. *Computational Optimization and Applications*, 63(2):425–459, 2016.
- [32] A. Fischer, M. Herrich, A. F. Izmailov, and M. V. Solodov. A globally convergent LP-Newton method. *SIAM Journal on Optimization*, 26(4):2012–2033, 2016.
- [33] R. Fletcher. A new approach to variable metric algorithms. *The computer journal*, 13(3):317–322, 1970.
- [34] R. Fletcher. A new variational result for quasi-Newton formulae. *SIAM Journal on Optimization*, 1(1):18–21, 1991.
- [35] D. Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of computation*, 24(109):23–26, 1970.
- [36] M. A. Gomes-Ruggiero and J. M. Martínez. The column-updating method for solving nonlinear equations in Hilbert space. *ESAIM: Mathematical Modelling and Numerical Analysis*, 26(2):309–330, 1992.
- [37] I. Griva, S. G. Nash, and A. Sofer. *Linear and nonlinear optimization*, volume 108. SIAM, 2009.
- [38] O. Güler, F. Gürtuna, and O. Shevchenko. Duality in quasi-Newton methods and new variational characterizations of the DFP and BFGS updates. *Optimization Methods & Software*, 24(1):45–62, 2009.
- [39] W. Hart and F. Soesianto. On the solution of highly structured nonlinear equations. *Journal of computational and applied mathematics*, 40(3):285–296, 1992.
- [40] W. Hock and K. Schittkowski. Test examples for nonlinear programming codes. *Journal of Optimization Theory and Applications*, 00:127–129, 1980.
- [41] G. Hughes and J. Dunn. Newton–Goldstein convergence rates for convex constrained minimization problems with singular solutions. *Applied Mathematics and Optimization*, 12(1):203–230, 1984.
- [42] A. Izmailov and M. Solodov. *Otimização*. IMPA, 2005.

- [43] A. F. Izmailov and M. V. Solodov. *Newton-type methods for optimization and variational problems*. Springer, 2014.
- [44] A. F. Izmailov and M. V. Solodov. On error bounds and Newton-type methods for generalized nash equilibrium problems. *Computational Optimization and Applications*, 59(1-2):201–218, 2014.
- [45] G. W. Johnson and N. H. Austria. A quasi-Newton method employing direct secant updates of matrix factorizations. *SIAM Journal on Numerical Analysis*, 20(2):315–325, 1983.
- [46] C. Kanzow. An active set-type Newton method for constrained nonlinear systems. In *Complementarity: applications, algorithms and extensions*, pages 179–200. Springer, 2001.
- [47] C. Kanzow. Strictly feasible equation-based methods for mixed complementarity problems. *Numerische Mathematik*, 89(1):135–160, 2001.
- [48] C. Kelley and E. Sachs. A quasi-Newton method for elliptic boundary value problems. *SIAM Journal on Numerical Analysis*, 24(3):516–531, 1987.
- [49] M. Kojima and S. Shindo. Extension of Newton and quasi-Newton methods to systems of PC 1 equations. *Journal of the Operations Research Society of Japan*, 29(4):352–375, 1986.
- [50] D. N. Kozakevich, J. M. Martínez, S. A. Santos, et al. Solving nonlinear systems of equations with simple constraints. *Computational and Applied Mathematics*, 1997.
- [51] B. Kummer et al. Newton’s method for non-differentiable functions. *Advances in mathematical optimization*, 45:114–125, 1988.
- [52] D.-H. Li and M. Fukushima. Globally convergent Broyden-like methods for semismooth equations and applications to VIP, NCP and MCP. *Annals of Operations Research*, 103(1-4):71–97, 2001.
- [53] G. Li. Successive column correction algorithms for solving sparse nonlinear systems of equations. *Mathematical Programming*, 43(1-3):187–207, 1989.
- [54] V. L. R. Lopes, J. M. Martínez, and R. Pérez. On the local convergence of quasi-Newton methods for nonlinear complementarity problems. *Applied Numerical Mathematics*, 30(1):3–22, 1999.
- [55] L. Lukšan and J. Vlček. Computational experience with globally convergent descent methods for large sparse systems of nonlinear equations. *Optimization Methods and Software*, 8(3-4):201–223, 1998.
- [56] Z. Q. Luo and P. Tseng. Error bounds and convergence analysis of feasible descent methods: a general approach. *Annals of Operations Research*, 46(1):157–178, 1993.

- [57] C. Ma. A new smoothing quasi-Newton method for nonlinear complementarity problems. *Applied mathematics and computation*, 171(2):807–823, 2005.
- [58] J. M. Martínez. A quasi-Newton method with modification of one column per iteration. *Computing*, 33(3-4):353–362, 1984.
- [59] J. M. Martínez. A family of quasi-Newton methods for nonlinear equations with direct secant updates of matrix factorizations. *SIAM Journal on Numerical Analysis*, 27(4):1034–1049, 1990.
- [60] J. M. Martínez. Practical quasi-Newton methods for solving nonlinear systems. *Journal of Computational and Applied Mathematics*, 124(1-2):97–121, 2000.
- [61] J. M. Martínez and L. Ochi. Sobre dois métodos de Broyden. *Matemática Aplicada e Computacional*, 1(2):135–141, 1982.
- [62] J. M. Martínez and M. C. Zambaldi. An inverse column-updating method for solving large-scale nonlinear systems of equations. *Dynamical Systems*, 1(2):129–140, 1992.
- [63] M. Martínez and D. Fernández. A quasi-Newton modified LP-Newton method. *Optimization Methods and Software*, 00:1–16, 2017.
- [64] M. Martínez and D. Fernández. On the local and superlinear convergence of a secant modified Linear-Programming-Newton method. *Journal of Optimization Theory and Applications*, 180(3):993–1010, 2019.
- [65] H. Matthies and G. Strang. The solution of nonlinear finite element equations. *International journal for numerical methods in engineering*, 14(11):1613–1626, 1979.
- [66] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.
- [67] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*, volume 30. SIAM, 1970.
- [68] J. S. Pang and L. Qi. Nonsmooth equations: motivation and algorithms. *SIAM Journal on optimization*, 3(3):443–465, 1993.
- [69] R. Pérez and V. L. R. Lopes. Recent applications and numerical implementation of quasi-Newton methods for solving nonlinear systems of equations. *Numerical Algorithms*, 35(2-4):261–285, 2004.
- [70] M. J. Powell. A new algorithm for unconstrained optimization. In *Nonlinear programming*, pages 31–65. Elsevier, 1970.
- [71] H. Qi, L. Qi, and D. Sun. Solving Karush–Kuhn–Tucker systems via the trust region and the conjugate gradient methods. *SIAM Journal on Optimization*, 14(2):439–463, 2003.

- [72] L. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of operations research*, 18(1):227–244, 1993.
- [73] L. Qi and J. Sun. A nonsmooth version of Newton’s method. *Mathematical programming*, 58(1-3):353–367, 1993.
- [74] D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Mathematics of computation*, 24(111):647–656, 1970.
- [75] A. H. Sherman. On Newton-iterative methods for the solution of systems of nonlinear equations. *SIAM Journal on Numerical Analysis*, 15(4):755–771, 1978.
- [76] D. Sun and L. Qi. On NCP-functions. *Computational Optimization and Applications*, 13(1-3):201–220, 1999.
- [77] M. Ulbrich. Nonmonotone trust-region methods for bound-constrained semismooth equations with applications to nonlinear mixed complementarity problems. *SIAM Journal on Optimization*, 11(4):889–917, 2001.
- [78] N. Yamashita and M. Fukushima. On the rate of convergence of the Levenberg-Marquardt method. In *Topics in numerical analysis*, pages 239–249. Springer, 2001.