



Fundamentos de mínimos cuadrados restringidos a través de la factorización QR generalizada

Sergio Andrés Cabrera Miranda

Universidad ECCI
Departamento de Ciencias Básicas
Bogotá, Colombia
2022

Fundamentos de mínimos cuadrados restringidos a través de la factorización QR generalizada

Sergio Andrés Cabrera Miranda

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Profesional en Estadística

Director(a):
Ph.D. Juan Gabriel Triana Laverde

Universidad ECCI
Departamento de Ciencias Básicas
Bogotá, Colombia
2021

¡Oh alma mía, no aspire a la vida inmortal,
pero agota el campo de lo posible!

Píndaro, Píticas III.

Yo no soy un hombre, soy un campo de batalla.

Friedrich Nietzsche

Agradecimientos

El gran reto personal durante mi vida ha sido maximizar la virtuosidad y la belleza a través de una creación. Particularmente le he dedicado cuanto mis fuerzas internas me han permitido postrar en este trabajo, pero nunca como un trabajo individual. Debo iniciar agradeciendo a cuán maravillosas mentes durante siglos han permitido en un esfuerzo colectivo forjar con tanta solidez y con una belleza inmaculada, Newton diría “Si he visto más lejos es porque estoy sentado sobre los hombros de gigantes”.

En mi esfera personal, debo agradecer en primera instancia a mi familia, mi madre Ángela, mi padre Guillermo, mi hermana Diana y todos aquellos familiares que vieron mi proceso durante los últimos años. En tercera instancia al Doctor Juan Gabriel Triana Laverde por su acompañamiento, su paciencia y el esfuerzo que le ha dedicado al proyecto para que saliera a flote con la magnitud que se ha demostrado, y por darme el gran reto de exigir mi capacidad al límite para desbordar cuanta ciencia he podido entender y desarrollar en pro de un objetivo más grande que lo que demuestra este trabajo específico.

Le agradezco a todos los docentes que con su pasión han logrado irradiar cuanto conocen para nuestra formación, en especial a Alexander Fuentes y Alexandra Esteban por su labor tan profesional y generosa, cuyo crecimiento en mí han podido evidenciar en el transcurrir del pregrado. En siguiente instancia, a mis queridos compañeros con quienes he logrado compartir experiencias que me han enriquecido a nivel educativo y personal, incluyendo aquellos que hoy no están pero que han acompañado el proceso que he realizado, en especial y entre todos ellos a Ana Ortiz por su apoyo tan incondicional, a todos les agradezco y los recuerdo con afecto.

Fuera del ámbito académico a mis queridos amigos Anly y Julián por estar cuando los necesitaba para darme la motivación que requería. Finalmente le agradezco con profundo cariño a Amandine, quien más ánimo y motivación me brindó durante este periodo para luchar por mis deseos e impulsarme más allá de mí mismo para lograr un proyecto de la más alta calidad.

Esta creación vive y refleja haces de luz de tal intensidad como lo fue el arduo proceso que ha engrandecido mi visión profesional; este proyecto ha maximizado mi pasión por las ciencias exactas que me alumbraron en su inicio gracias a ustedes.

Los frutos recopilados en este trabajo pertenecen a ustedes y lo que han sembrado en mí.

Resumen

La explicación de diversos fenómenos se puede desarrollar a través de un modelamiento de funciones que permiten caracterizar su comportamiento. Una de las metodologías utilizadas para este fin requiere de sistemas de ecuaciones lineales, en los cuales el método de mínimos cuadrados restringidos optimiza la solución que ajusta una función a los datos respecto del método de mínimos cuadrados ordinarios en aspectos de eficiencia de un estimador y la interpretabilidad de los resultados. La solución del modelo que presenta restricciones puede estar basada en la factorización QR. En este trabajo se formulan las bases teóricas de una variación de dicho método, conocida como factorización QR generalizada para ser aplicada cuando se incluyen restricciones de tipo igualdad y posteriormente estos fundamentos son aplicados al problema de mínimos cuadrados restringidos con restricciones de igualdad en cuaterniones.

Palabras clave: Factorización GQR, mínimos cuadrados, mínimos cuadrados restringidos con restricciones de igualdad, cuaterniones.

Abstract

Diverse phenomena can be explained through functions that characterizing their behavior. Some of these can be expressed using systems of linear equations whose solutions are obtained through least-squares methods. Constrained least squares (CLS) method are better than ordinary least squares method for some inferential features. The solution of CLS can be reached with some method that guarantee the numerical stability of the estimated values, this numerical stability varies according to the calculation method, among them QR factorization is the standard method for solving least square problems due to numerical stability in their results. In this work, the theoretical bases of a variation of QR factorization, known as generalized QR factorization, are formulated. In addition, this method is applied for solving least squares problems with equality constraints in quaternions.

Keywords: GQR factorization, least-square problems, linear equality-constrained least-square problems, quaternion

Lista de Figuras

| | |
|---|----|
| 2-1. Proyección ortogonal de un vector respuesta b sobre el espacio generado por $C(A)$ | 12 |
| 3-1. Valores críticos de la prueba del error cuadrático medio para restricciones lineales. | 33 |
| 4-1. Representación del plano de Argand para números complejos. | 49 |
| 4-2. Representación gráfica para un cuaternión Q en \mathbb{R}^3 | 50 |
| 5-1. Residual $Ax - b$ para 4 métodos de solución | 56 |
| 5-2. Residual $Cx - d$ para 4 métodos de solución | 56 |
| 5-3. Residual $Ax - b$ para 4 métodos de solución | 58 |
| 5-4. Residual $Cx - d$ para 4 métodos de solución | 58 |
| 5-5. Residual $Ax - b$ para 4 métodos de solución | 59 |
| 5-6. Residual $Cx - d$ para 4 métodos de solución | 60 |
| 5-7. Residuales para la colinealidad en A con $m = 10$ | 61 |
| 5-8. Residuales para la colinealidad en A y C con $m = 10$ | 62 |
| 5-9. Residuales para la colinealidad en C con $m = 10$ | 62 |
| 5-10. Residuales para la colinealidad en A con $m = 100$ | 63 |
| 5-11. Residuales para la colinealidad en A y C con $m = 100$ | 64 |
| 5-12. Residuales para la colinealidad en C con $m = 100$ | 64 |
| 5-13. Residuales para los métodos de solución para cuaterniones | 66 |
| 5-14. Residuales para los métodos de solución para cuaterniones | 67 |

Lista de Tablas

| | |
|---|----|
| 4-1. Multiplicación de cuaterniones | 50 |
| 5-1. Comparación de 4 métodos de solución | 55 |
| 5-2. Comparación de 4 métodos de solución | 57 |
| 5-3. Comparación de 4 métodos de solución | 59 |

1. Introducción

En el contexto de la optimización matemática, se pueden modelar funciones que se ajusten adecuadamente al comportamiento de algún fenómeno, la herramienta insigne es el método de mínimos cuadrados, la cual busca minimizar la distancia de los errores entre unos valores reales y los valores que se estiman.

Al considerar sistemas de ecuaciones lineales, es usual representar el problema mediante matrices [37]. En estos problemas, la gran cantidad de datos puede generar sistemas de gran tamaño; por lo anterior, una estrategia que se ha empleado para facilitar el tratamiento de los procesos es la factorización de matrices, que consiste en descomponer una matriz como producto de matrices especiales [54].

A través de la descomposición matricial, la consistencia de la solución del método de mínimos cuadrados se evalúa según particularidades como la exactitud de las estimaciones, la estabilidad numérica y/o el tiempo de ejecución del método [56]. El caso de la factorización QR, permite obtener resultados con alta estabilidad numérica en las estimaciones de los parámetros que caracterizan la función de ajuste para el problema de mínimos cuadrados; por lo anterior, la factorización QR y sus variantes siguen siendo un área de estudio de interés tanto matemático como estadístico [58].

Además de la vertiente de mínimos cuadrados ordinarios, han surgido otras alternativas que buscan suplir las deficiencias y que se adaptan mejor a los datos obtenidos, particularmente se puede evitar la incompatibilidad entre los resultados del modelo y las interpretaciones prácticas, además de aumentar criterios de eficiencia para las predicciones, esto se puede lograr imponiendo restricciones basada en información a priori. Entre las múltiples alternativas para resolver problemas de mínimos cuadrados con restricciones de tipo igualdad, se puede encontrar la factorización QR [58], métodos basados en regularización [59], entre otros [9], [5].

Con el propósito de exponer las diferencias del método de mínimos cuadrados ordinarios y restringidos, y tras una revisión histórica del método de mínimos cuadrados, se exponen los fundamentos de las ecuaciones normales con la interpretación geométrica, el teorema de Gauss-Markov y un análisis detallado sobre las implicaciones del uso de este teorema en efectos teóricos y prácticos, para finalizar con la solución del método a través del uso de la factorización QR.

En la segunda parte del trabajo, se fundamenta la validez para el uso de mínimos cuadrados restringidos mediante el desarrollo del estimador propuesto por [50] mediante multiplicadores

de Lagrange, además de los momentos del estimador sobre el cual se realizar estimaciones a partir de la esperanza y la varianza. Dos criterios de comparación entre el estimador restringido y el estimador de mínimos cuadrados ordinarios son expuestos a través de pruebas de hipótesis, uno de ellos está basado en la prueba t y F, para determinar la existencia de diferencias significativas entre el modelo con y sin restricciones [57], y el segundo criterio basado en el error cuadrático medio que evalúa las implicaciones al imponer las restricciones en el estimador [52]. Para finalizar con el desarrollo del modelo restringido a través de otros dos métodos de solución con sus respectivos algoritmos de solución en el software R que se pueden encontrar en el Apéndice.

Dada la importancia de las descomposiciones matriciales, en el siguiente capítulo se desarrolla un cuarto método de solución basado en la factorización QR generalizada, ésta es definida con los conceptos matriciales que subyacen y la sustentan, junto con la debida programación para ser aplicada como se evidencia en el Apéndice B, posteriormente la teoría desarrollada se acopla al problema del modelo restringido que presenta restricciones de igualdad. Las aplicaciones del método desarrollado son expuestas en cuaterniones una vez definidos los conceptos relacionados con los números complejos, como las propiedades básicas de los cuaterniones, interpretación geométrica y las transformaciones necesarias para ser aplicada la factorización QR generalizada.

Por último, a través de una simulación con matrices aleatorias, se evalúan los 4 métodos de solución expuestos a partir de matrices sobredeterminadas, matrices sobredeterminadas que presentan colinealidad y la aplicación de los algoritmos en cuaterniones, mediante la comparación de los dos residuales que se evalúan para el modelo restringido.

2. Mínimos cuadrados

Todos ellos recibieron honores de sus contemporáneos y fueron la gloria de su tiempo.

Algunos dejaron un nombre famoso que será conservado por sus herederos.

Y hay otros a los que ya nadie recuerda, que terminaron cuando terminó su vida, que existieron como si no hubieran existido, y después pasó lo mismo con sus hijos.

Eclesiástes 44:7-9.

Lo nuevo se apoya en lo viejo, y lo viejo aflora en lo nuevo: no hay tradición sin progreso, pero tampoco hay progreso sin tradición.

Erasmus de Rotterdam, Elogio de la locura.

2.1. Historia del método de mínimos cuadrados

Un primer precursor moderno del uso de herramientas estadísticas es probablemente Galileo [22], quien primero sugiere que toda observación está sujeta a errores, y segundo que es más frecuente encontrar errores pequeños que grandes, pero que de igual modo, éstos se distribuyen simétricamente alrededor del valor cero. Un esbozo de una teoría sobre los errores se puede encontrar en su versión traducida [49], cuyo aparte dice:

“Entonces, siendo capaces estos observadores, y habiéndose equivocado en todas formas, sus errores deben corregirse para poder obtener la mejor información posible de sus observaciones, será apropiado para nosotros aplicar las mínimas correcciones y que sean lo más pequeñas que se pueda; solo lo suficiente para eliminar las observaciones de la imposibilidad y restaurarlas a la posibilidad”. p. (290)

Posteriormente, en el siglo XVIII hubo una gran cantidad de avances matemáticos, se iniciaron teorías que con el tiempo se reforzaron y se continuaron desarrollos que iniciaron con los aportes probabilísticos de Cardano, Fermat y Pascal. Los desarrollos colectivos logrados en distintos ámbitos de la experiencia humana, ya sean artísticos o científicos no se pueden disociar de la época en la que se desarrollan, para este caso hubo dos razones fundamentales.

La primer razón es que en el siglo XVIII, conocido como el siglo de la ilustración, la ciencia tomó un papel central. La búsqueda del conocimiento llevó a la financiación de diversas expediciones, tomando gran relevancia la cartografía ya que además de aportar conocimientos, permitía desarrollar mejoras en la navegación; esta área se vio beneficiada por los nuevos enfoques geométricos, con los cuales se lograba mayor precisión en los mapas [16]; las técnicas de localización no podía depender netamente de las observaciones terrestres, razón por la cual el comportamiento de los cuerpos celestes ayudaba en la navegación de los buques en mares abiertos [28], todos estos estudios generaban retos ante los cuales los científicos buscaban soluciones.

La segunda razón es debida al legado de Newton, como las tres leyes que llevan su nombre, la ley de gravitación universal, el desarrollo del cálculo matemático, los avances en el campo de la óptica, etc. Los desarrollos provocaron un interés asiduo en el mundo científico por el estudio de la mecánica celeste [6], es decir, un interés en el estudio de los cuerpos celestes, de la forma y del comportamiento del planeta.

Entre la diversidad de técnicas que permitían la localización basada en los cuerpos celestes, los métodos desarrollados involucraban el uso de la media aritmética de un conjunto de mediciones en condiciones de repetibilidad ¹ [49]. Una variación del método usual se dió en 1722, al publicar una obra póstuma de Roger Cotes, titulada “Aestimatio Errorum in Mixta

¹Condición de medición, dentro de un conjunto de condiciones que incluye el mismo procedimiento de medida, los mismos operadores, el mismo sistema de medida, las mismas condiciones de operación y el mismo lugar, así como mediciones repetidas del mismo objeto o de un objeto similar en un periodo corto de tiempo. Fuente: Vocabulario Internacional de Metrología JCGM 200:2012.

Mathesi”, en esta obra el autor expone una regla cuya interpretación es semejante al de una media ponderada [49], aunque tras la publicación tuvo poca influencia en la matemática posterior, pues la explicación de Cotes en su obra es ambigua y no ejemplificó el método de uso de su regla.

Estudiando los movimientos de Júpiter y Saturno, Leonhard Euler publicó en 1749 la obra “Recherches sur la question des inégalités du mouvement de Saturne et de Jupiter” donde estudia el movimiento de la órbita de Saturno, con solución insatisfactoria. Pero a través de su trabajo, en la astronomía se hizo la transición del uso de promediar simples observaciones a combinar ecuaciones de observaciones, mediante el método de los promedios para el ajuste de ecuaciones lineales [47].

Por otra parte, el astrónomo y cartógrafo Tobias Mayer, en 1750 publicó “Abhandlung uber die Umwalzung des Monds um seine Axe und die scheinbare Bewegung der Mondsfecten”, donde trabajó de manera independiente de los avances de Euler, recolectó observaciones sobre el problema de la libración de la luna entre 1748 y 1749. Al igual que Euler desarrolló e implementó el método de los promedios y le da solución al problema; sin embargo, se concluyó que los resultados obtenidos dependen de la elección de los grupos [47]. A pesar de ello, el avance de Mayer fue abordar el problema con un enfoque estadístico más desarrollado que el de Euler. Tobias Mayer destacó por su trabajos de cartografía de la luna, desarrolló tablas lunares midiendo posiciones y movimientos precisos de los cráteres de la luna, pues, aunque se suele decir que la luna muestra el mismo hemisferio hacia la tierra, ésta oscila ligeramente de dos maneras. Al respecto, Rambaux [43] dice:

“Uno está en longitud, con rotación paralela al plano ecuatorial de la Luna, y tiene un período de 2,9 años. Para el segundo modo, el modo de latitud, el eje normal al plano ecuatorial lunar traza un pequeño cono. Ese movimiento en el espacio es retrógrado con un período cercano a los 81 años.”

A este hecho que se describe, se le denomina libración -o titubeo- de la luna. El planteamiento de Mayer se basó en este hecho, pudiendo generar 27 ecuaciones con 3 parámetros, divididos en tres grupos de 9 ecuaciones, el primer grupo consistía en las ecuaciones con los valores más altos para una variable independiente, el segundo grupo con los siguientes valores de la misma variable y continuaba sucesivamente hasta los valores menores, luego sumaba en cada grupo las ecuaciones componente a componente, terminando con tres ecuaciones de tres incógnitas, cuya resolución presenta mayor sencillez que el planteamiento inicial [41], [28]. Así, podemos observar las clase de herramientas que se utilizaban en la época y la dimensión de la clase de problemas que resolvían con herramientas estadísticas básicas.

Debido a la sencillez del método de Mayer, éste fue muy usado por astrónomos y geodestas; sin embargo, debido a que su desarrollo fue el de un método ad hoc, no era capaz de generalizar la solución a cualquier tipo de problema, por lo tanto y como se mencionó previamente, este método no encontraba la mejor solución a los problemas. Basado en este método y en las

implicaciones que tenía consigo, Boscovich, Simpson y Maire desarrollaron nuevos métodos. Roger Boscovich en 1755 propuso el método de mínimas desviaciones absolutas estudiando la forma de la tierra [28], con el cual dió solución al considerado tercer mayor problema científico del siglo XVIII.

En 1787, Laplace desarrolló métodos para ajustar curvas y superficies tanto para problemas relacionados con astronomía, como en geodesia. En los métodos desarrollados, Laplace minimizó el residual máximo, el residual absoluto promedio y el residuo cuadrado promedio, de un modelo lineal [40], las ideas las recopiló en “Exposition du système du monde” publicado en 1796.

En 1799, fue presentada a la Asamblea Francesa el patrón de una nueva unidad de medida, conocida como metro, definida como la diez millonésima parte de la longitud de un cuadrante terrestre, esto fue posible gracias a una extrapolación realizada a partir de la medición de la distancia desde Dunkerque hasta Barcelona, a través del meridiano de París; en este mismo año, Gauss publica “Allgemeine Geographische Ephemeriden 4” en donde se muestran los datos obtenidos durante dichas mediciones y se realiza el cálculo de esta distancia empleando un método precursor al método de mínimos cuadrados [18].

En 1805, Legendre fue el primero en publicar un desarrollo del método de mínimos cuadrados en el apéndice de “Nouvelles méthodes pour la détermination des orbites des comètes” aunque sin fundamentos sólidos desde un enfoque probabilístico. No obstante, es explícito en el proceso que se busca realizar:

“En la mayoría de las investigaciones donde el objetivo es deducir los resultados más precisos posibles de las mediciones observadas, nos dirigimos a un sistema de ecuaciones de la forma:

$$E = ax + by + fz + \&c.,$$

En donde a, b, f y $\&c.$ son coeficientes conocidos que varían de una ecuación a otra, y x, y, z y $\&c.$ son cantidades desconocidas a ser determinadas con la condición de que cada valor de E debe ser cercano a cero o una cantidad muy pequeña.” p. (13).

Una primera prueba es dada por Dr. Robert Adrain en 1808 bajo el nombre de “Research concerning the probabilities of the errors which happen in making observations”. En 1809 Gauss de manera independiente publica “Theoria motus corporum coelestium” dando una justificación probabilística a través de máxima verosimilitud. Con el paso de los años continúan diferentes demostraciones del método propuestas por Laplace, Gauss, Ivory, Hagen, Bessel y más autores, a los que el lector puede referirse en [38].

Sin embargo, era usual en Gauss no publicar sus descubrimientos de inmediato, hasta tener contenido suficiente [1] así, a pesar de la publicación de Legendre, Gauss se atribuyó el desarrollo del método, generando una disputa entre ambos que terminaría favoreciendo a Gauss, quien indica el inicio de su desarrollo desde 1795, la evidencia ampliamente estudiada,

véase [46] muestra 4 piezas de estudio, de las cuatro, la más importante se encuentra detallada en [6] sobre la predicción de la aparición de Ceres.

El astrónomo italiano Joseph Piazzi en 1801 descubrió un cuerpo celeste al que posteriormente se le llamaría Ceres, él observó la trayectoria del cuerpo durante 40 días, correspondientes a 9 grados de su órbita, luego, desapareció tras el sol y surgió el interés de la predicción de la posición al reaparecer.

Franz von Zach, un astrónomo alemán que Gauss conoció dos años antes, sugirió el lugar de aparición de Ceres, Gauss también realizó predicciones obteniendo resultados diferentes a las de Von Zach y de otros interesados en el evento; medio año después, Zach redescubrió al cuerpo en diciembre de 1801, el cual estaba casi exactamente donde Gauss lo predijo [46].

En 1823 Gauss comparó su publicación de 1809 con una publicación de Laplace de 1812 y, al ver que aún no se tenía un método completamente satisfactorio, demostró el teorema de Gauss-Markov, el cual menciona que a través de mínimos cuadrados las estimaciones tienen varianza mínima sin importar cuál sea la distribución de los errores como se evidenciará adelante.

Quienes más contribuyeron al avance desde el enfoque estadístico de mínimos cuadrados fueron Francis Galton, Francis Edgeworth, y Karl Pearson. Galton aplicó el método de mínimos cuadrados desde un enfoque estadístico, en el artículo “Regression towards mediocrity in hereditary stature” publicado en 1886 y en “Hereditability Genius” de 1869, denominando análisis de regresión al realizar investigaciones sobre la herencia genética entre padres e hijos [20], siendo una continuación de los planteamientos de Lambert Adolphe Quetelet.

Galton pudo desarrollar sus planteamientos basado en datos de 930 padres y sus hijos, encontrando una relación visual mediante un gráfico estadístico, actualmente conocido como gráfico de dispersión, empleando además la distribución acumulada normal inversa. En 1889, con el libro “Natural Inheritance”, tras veinte años de trabajo consigue reconciliar la teoría de los errores y la herencia, ese trabajo lo dividió en 4 etapas [47]: una investigación inicial sobre las condiciones en las que se produce la ley de errores en 1874-1875; la experimentación con guisantes que le permitió formular de manera empírica la regresión en 1877; luego su trabajo de un marco de referencia matemático de regresión en 1885 y finalmente la recopilación en 1889 publicando el libro.

Las contribuciones de Galton inspiraron a autores como Francis Edgeworth, Karl Pearson, Edne Yule, entre otros. Edgeworth pese a que estudió literatura clásica, fue un teórico que generalizó formalmente ideas de Galton y contribuyó con la aplicación en economía, en 1883 publicó “The method of least squares” donde discute los enfoques desarrollados por Gauss y Laplace e indicó planteamientos sobre la bondad de ajuste. Contribuyó al estudio de la regresión cuantílica, basado en el enfoque de Boscovich [34], en donde estudió la desviación absoluta media y propuso el uso de la mediana sobre la media en ajustes mediante mínimos cuadrados para distribuciones con colas pesadas; no obstante, no pudo encontrar un

fundamento matemático.

El matemático Karl Pearson se valió de la originalidad de los planteamientos de Galton y de la falta de profundidad en los desarrollos de Edgeworth -debido a que no contaba con formación como matemático- para desarrollar una metodología sólida [47]. Pearson encontró el ajuste mediante una línea recta a través de las estimaciones de mínimos cuadrados para el problema de Galton, encontrando que en efecto la línea tiende hacia la media aritmética. Además, Pearson generalizó las ideas de Galton a múltiples dimensiones y estudió la distribución normal multivariada [48]. Posteriormente continuó implementando medidas como el coeficiente de correlación y desarrollando un método más robusto de medición. También, indujo avances en inferencia estadística como la distribución Chi Cuadrado por ejemplo, mientras dirigía la Escuela Biométrica del University College de Londres tras haberla creado.

Las bases modernas del método se deben a Ronald Fisher, quien consolidó las teorías de errores de Gauss y la teoría de correlación de Pearson formulando las estimaciones de la regresión lineal [2]. Promovió avances del principio de máxima verosimilitud [7], generó métodos de bondad de ajuste basado en la teoría de la distribución Chi-cuadrado de Pearson publicándolo en 1922 “The goodness of fit of regression formulae, and the distribution of regression coefficients”. Realizó al igual que Pearson, comparaciones para el mejor ajuste comparando el método de mínimos cuadrados con el método de los momentos [2].

El desarrollo de las técnicas ad hoc explicadas previamente, una vez abstraídas al campo matemático, lograron una consistencia y solidez lógica, consolidándose como concepto formal cada método, y trayendo a colación a Charles Darwin, los métodos más adaptables al cambio sobrevivieron, así, de estas bases matemáticas pudieron evolucionar otros, y fueron aplicados nuevamente en campos específicos.

Debido a todos los avances realizados y al desarrollo continuo de métodos, surge en la estadística una rama de estudio, regresión lineal, que goza de herramientas heredadas de las aplicaciones prácticas y de los fundamentos matemáticos; para efectos prácticos, su esencia recae en interpolaciones y predicciones, y desde el enfoque teórico la potencia surge en optimizar y garantizar cualidades óptimas de las estimaciones además de características apropiadas con las que debe contar un resultado de medición.

En síntesis, se generaron diversos métodos para el ajuste de funciones dada información previamente obtenida (mínimos cuadrados, máxima verosimilitud y basada en la estadística bayesiana), y métodos algorítmicos para la ejecución de los métodos de ajuste. Este trabajo considera el enfoque de regresión lineal a través del método de mínimos cuadrados.

2.2. Mínimos cuadrados

El análisis de regresión es una técnica que permite ajustar un modelo, explicando el comportamiento de un fenómeno a partir de un conjunto de observaciones. Las áreas de aplicación abarcan múltiples disciplinas, debido a que desde sus primeras formulaciones se ha logrado aplicar la teoría a problemas de otras áreas; esto ha permitido que la regresión lineal sea abordada desde diferentes perspectivas, lo cual ha llevado a nuevas preguntas que han permitido enriquecer la teoría [45].

El modelamiento de funciones buscar encontrar una función que se adapte mejor a las observaciones, determinar la que se denomina “mejor función” implica basarse en distintos criterios que garanticen la existencia de una solución satisfactoria; según cada problema, existen métodos que aumentan la eficacia de los resultados, ya sea a través de la exactitud de las estimaciones, el tiempo de ejecución del método, la estabilidad numérica y/o el costo computacional.

Mediante el método de mínimos cuadrados se garantiza obtener la función que mejor se ajusta a los datos bajo ciertas condiciones, para que se pueda explicar con fiabilidad el comportamiento de una variable de interés a través de una variable observada. El error entre las observaciones y la función de ajuste es la base sobre la cual se minimizan las diferencias, disminuyendo el sesgo, y en consecuencia, se garantiza mayor exactitud en las estimaciones realizadas. La solución de este método depende de las características de los datos obtenidos, en ese orden de ideas, se pueden clasificar distintos tipos de soluciones a través de métodos analíticos, matriciales, de regularización o basados en métodos numéricos.

Definición 1 (Modelo lineal). *Sean n observaciones a_1, a_2, \dots, a_n una m.a. de una población, además, sea b una variable aleatoria relacionada con el conjunto de variables a_i como sigue*

$$\hat{b} \approx x_0 + x_1 a_1 + x_2 a_2 + \dots + x_n a_n$$

donde $x = (x_0, x_1, x_2, \dots, x_n)$ son parámetros desconocidos fijos, $a = (a_1, a_2, a_3, \dots, a_n)$ son variables independientes, endógenas u observadas y, b es una variable dependiente, respuesta, regresora o exógena.

Este modelo representa una relación estocástica entre una variable dependiente y otra independiente; es una relación estocástica debido a que se encuentra sujeta a condiciones aleatorias de las variables de medición, de allí que las estimaciones se realizan en un campo probabilístico a diferencia de un sistema determinista, en el cual el comportamiento a estudiar es predecible. De esta manera, la relación de las variables expresada anteriormente, puede ser generalizada a través de un sistema de ecuaciones lineales, que a su vez, puede ser reexpresado a través de una representación matricial donde las variables independientes son agrupadas en una matriz A de tamaño $m \times n$, los vectores x y b son de tamaño $n \times 1$ y $m \times 1$ respectivamente, de la manera:

$$Ax \approx b \quad (2-1)$$

Las distintas formas matriciales que puede adquirir la matriz A tiene como consecuencia distintos desarrollos de las formulaciones matemáticas. En el caso más común de las aplicaciones prácticas, podemos encontrar un sistema de ecuaciones donde la matriz A conduce a tener un sistema sobredeterminado, se caracteriza por tener más ecuaciones que incógnitas ($m \geq n$), no tiene solución debido a que el sistema no es consistente si b no pertenece al espacio columna de A , sin embargo, se puede realizar una aproximación según la minimización de los residuales.

$$\sum \varepsilon_i^2 = \min \|Ax - b\|_2^2 \quad (2-2)$$

Se emplea la norma euclidiana en la ecuación 2-2 por las propiedades que se mencionan en el Apéndice A, el uso de esta norma se justificará al utilizar transformaciones ortogonales debido a su invarianza; en esta sección esta norma es útil para la construcción de las ecuaciones normales.

De la ecuación 2-2 surge el planteamiento del método de mínimos cuadrados ordinarios, su solución se determina a través de las ecuaciones normales, que se derivan de las condiciones de primer orden. Para la optimización de dicha ecuación no se trabaja con las desviaciones absolutas, pues la suma de estos valores es cero, en cambio, se trabaja con los cuadrados de dichas diferencias, de este planteamiento inicial se realiza la representación matricial de las ecuaciones normales, en el cual se reitera que el mejor ajuste por obtener es aquel en el cual se minimizan los valores de los residuales.

Proposición 2.2.1 (Ecuación normal). *La minimización de la suma de cuadrados de los residuales $Ax - b$, se satisface cuando $A^T Ax = A^T b$.*

Demostración. Partiendo de buscar minimizar los valores de los residuales

$$\sum_{i=0}^n \|\varepsilon_i\|_2^2 = \min (b - Ax)^T (b - Ax)$$

A través de las condiciones de primer orden

$$\begin{aligned} \frac{\partial}{\partial x} (b - Ax)^T (b - Ax) &= 0 \\ -2A^T b + 2A^T Ax &= 0 \end{aligned}$$

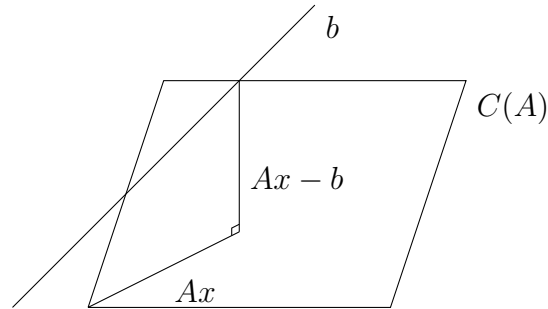
De aquí se concluye que $A^T Ax = A^T b$. □

Las ecuaciones que se obtienen de la expresión matricial

$$A^T Ax = A^T b \quad (2-3)$$

se denominan ecuaciones normales, el nombre hace referencia a la normalidad (u ortogonalidad), pues a nivel geométrico la solución ofrecida por este método es de proyecciones ortogonales sobre subespacios, como consecuencia de la ortogonalidad se puede afirmar que se minimizan las distancias; esta optimización es la esencia del método de mínimos cuadrados, pues en estas ecuaciones se afirma que $A^T(Ax - b) = 0$, véase la figura 2-1.

Figura 2-1.: Proyección ortogonal de un vector respuesta b sobre el espacio generado por $C(A)$.



La formulación de mínimos cuadrados ordinarios tiene solución, en principio, si las ecuaciones normales tienen solución, en cuyo caso dichas soluciones coinciden y son dadas por

$$x = (A^T A)^{-1} A^T b. \quad (2-4)$$

Una condición que permite garantizar la unicidad en la solución en la Ecuación 2-3, se deriva del rango de la matriz A . En caso de que el rango sea completo, es decir, $\text{rango}(A) = m$, la matriz es invertible (no-singular) por lo tanto existe la inversa de $A^T A$. Por el contrario, si no es completo el rango de la matriz A , existe un subespacio de valores para los cuales x admite solución en las ecuaciones normales, esto se debe a que existe dependencia entre las columnas de A , nos extenderemos adelante en este caso, al hablar de multicolinealidad en la matriz de diseño A .

Sin embargo, en caso de no ser posible calcular la inversa de la matriz $A^T A$, surgen complicaciones para estimar los valores que garanticen una solución óptima; no obstante, dado que es un método altamente utilizado en aplicaciones teóricas y prácticas, las limitaciones se pueden sortear de distintas maneras según una literatura ampliamente diversificada en distintas áreas, c.f. [44].

A pesar de esto, ciertas limitaciones del método son evidentes, una de ellas es la pérdida de

la estabilidad cuando la matriz A está mal condicionada². Un método basado en el cálculo de $A^T A$ es más costoso que los métodos basados en factorizaciones matriciales, entonces el método es más propenso a tener falta de precisión; sin embargo, [29] menciona que puede ser dos veces más rápido este método para el caso sobredeterminado cuando la matriz A es densa³. Otros detalles sobre el método pueden ser consultados en [29] y [8].

Una alternativa usual al usar las ecuaciones normales se obtiene al observar que $A^T A$ es una matriz simétrica y semidefinida positiva, por lo cual admite una factorización de Cholesky [23]; así, las propiedades en casos donde no es eficiente el método, pueden mejorar, entonces se induce que es un caso común el uso de factorizaciones matriciales, en aquellos casos donde los algoritmos basados en la matriz $A^T A$ presentan inconvenientes al trabajar con los datos de la matriz de diseño por la cantidad de información o por las características de los resultados en términos de exactitud numérica, de la estabilidad de las estimaciones y tiempos de ejecución del método.

Las ecuaciones anteriores adquieren otra connotación cuando se les busca una aplicación estadística, este hecho depende de que no se trabaja solamente con ecuaciones, en palabras de [21]:

“Un modelo es más que una ecuación, pueden haber declaraciones asociadas sobre la distribución de la variables aleatoria o sobre la naturaleza de X o y .”

Y en el caso en donde las declaraciones de las ecuaciones sobre las variables aleatorias cumplen las pautas del teorema de Gauss-Markov, el estimador x cumple las propiedades estadísticas de ser el mejor estimador lineal insesgado óptimo (MELIO).

2.3. Teorema de Gauss-Markov aplicado al modelo lineal

Teorema 2.3.1 (Gauss Markov). *Sea el modelo considerado en 2-1 donde $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ y $x \in \mathbb{R}^n$. El mejor estimador lineal insesgado óptimo (MELIO) se obtiene minimizando $\min_x \|Ax - b\|_2^2$. Y las condiciones de los residuales en las estimaciones deben seguir:*

- *la variable x es no estocástica o fija.*
- $E[\varepsilon|x] = 0$
- $V[\varepsilon|x] = \sigma^2$

²El condicionamiento de una matriz indica una medida de la perturbación de un resultado. Si un pequeño cambio en los valores de entrada produce un cambio pequeño en los valores de salida, la matriz está bien condicionada; pero, si en los valores de salida la variación es grande, se obtiene una matriz mal condicionada. Véase [29].

³Se dice matriz densa en casos donde la matriz presenta poca cantidad de valores cero, en caso contrario la denominación es matriz dispersa.

- $Corr(\varepsilon_i, \varepsilon_j) \neq 0, \forall i \neq j$

Establezcamos con detalle las condiciones mencionadas en el teorema, pues de ellas se derivan las implicaciones de carácter teórico y práctico por las cuales este teorema es de alta importancia.

A pesar de garantizar las mejores estimaciones cumpliendo los supuestos, cuyas bases son algebraicas, en muchos casos prácticos se incurre en la violación de los mismos. A pesar de incumplirse, éstos no son extremadamente limitantes, existen maneras de sortearlos sin afectar de gran manera la validez de los resultados [13]. No obstante, es de carácter imperativo analizar dichos diagnósticos para evitar asumir incorrectamente su cumplimiento o incumplimiento.

La linealidad es generada a través de una función lineal en sus parámetros, entonces, en caso de haber no-linealidad las estimaciones se tornan inválidas, las pruebas visuales sobre la relación de las variables pueden ayudar a indicar el tipo de comportamiento que presentan.

La insesgadez de un estimador, establece que el valor esperado coincide con el valor del parámetro, es decir, el parámetro θ es estimado a través de $\hat{\theta}$, y se relacionan con el operador de la esperanza matemática $E[\hat{\theta}] = \theta$, la ausencia de sesgo de un estimador ayuda a garantizar la validez de las estimaciones realizadas en torno al valor verdadero y desconocido del parámetro.

No obstante, aunque en algún caso puede existir ausencia de sesgo, no se puede garantizar que sea un buen estimador, incluso, no se puede garantizar en algunos casos su existencia, siguiendo a [36] “Aunque el sesgo es una condición atractiva, después de que un mejor estimador insesgado se ha encontrado, se debe investigar su desempeño y no descartar la posibilidad de que exista un estimador ligeramente sesgado con un riesgo mucho menor” debido a que “incluso cuando el estimador es insesgado no hay garantía de que el estimador sea deseables de otras formas y, en cambio, se puede preferir estimador que tiene algún sesgo.”

“Por otro lado, un gran sesgo generalmente se considera un inconveniente y se han desarrollado métodos especiales de reducción del sesgo.” Y en ese caso, el uso de métodos de remuestreo como el uso del método jackknife o bootstrap pueden ser de utilidad para la corrección del sesgo como menciona enseguida [36]. Así, el estudio detallado del desempeño de un estimador es un estudio en conjunto dado que debe cumplir alternamente con las características que afirma el teorema de Gauss Markov.

En caso de existir el estimador insesgado, el valor de $E(\varepsilon) = 0$, está centrado alrededor del valor del parámetro x , esto es, en condiciones de repetibilidad, el valor de \hat{x} varía debido a errores aleatorios, pero el componente sistemático es el que se mantiene alrededor del valor del parámetro. Entre los estimadores lineales e insesgados, la mejor estimación está garantizada

a través de este teorema, sin embargo, esto no significa que sea la mejor estimación entre todos los posibles estimadores como se mencionó anteriormente. Basados en 2-4 surge la demostración de la afirmación de insesgadez y de las propiedades de los residuales, teniendo en cuenta $E[\varepsilon] = 0$ se procede a continuación así

$$\begin{aligned}
 E[x] &= E[(A^T A)^{-1} A^T b] \\
 &= E[(A^T A)^{-1} A^T (Ax + \varepsilon)] \\
 &= E[(A^T A)^{-1} A^T Ax] + E[(A^T A)^{-1} \varepsilon] \\
 &= E[x] + E[(A^T A)^{-1} A^T \varepsilon] \\
 &= \hat{x}
 \end{aligned}$$

Otra propiedad a estudiar es la varianza de un estimador, también llamada eficiencia, ésta queda determinada entre la clase de estimadores lineales e insesgados. Eficiente es aquel que presenta una menor varianza respecto de los demás estimadores, un criterio para evaluar es el error cuadrático medio (ECM), la varianza del estimador \hat{x} queda determinada por la matriz de varianzas y covarianzas Σ_x tal que

$$\Sigma_x = E [(\hat{x} - x) (\hat{x} - x)']$$

Reexpresando el valor de \hat{x} según 2-4 tenemos

$$\begin{aligned}
 \Sigma_x &= E \left[(A^T A)^{-1} A^T (Ax - \varepsilon) - x \right] \left[(A^T A)^{-1} A^T (Ax - \varepsilon) - x \right]' \\
 &= E \left[((A^T A)^{-1} (A^T A)x - x - (A^T A)^{-1} A^T \varepsilon) \left((A^T A)^{-1} (A^T A)x - x - (A^T A)^{-1} A^T \varepsilon \right)' \right] \\
 &= E \left[(A^T A)^{-1} A^T \varepsilon \right] \left[(A^T A)^{-1} A^T \varepsilon \right]' \\
 &= E \left[(A^T A)^{-1} A^T \varepsilon \varepsilon (A^T A)^{-1} A^T \right]
 \end{aligned}$$

Bajo el supuesto de cumplir con la varianza constante se tiene que $E(\varepsilon) = \sigma^2$, así

$$\begin{aligned}
 \Sigma_x &= (A^T A)^{-1} A^T \sigma^2 I (A^T A)^{-1} A^T \\
 &= \sigma^2 (A^T A)^{-1}.
 \end{aligned}$$

En ésta expresión, se puede observar que en la demostración no se hace necesario el uso de propiedades como la insesgadez del estimador, ni del comportamiento normal de los residuales, y por lo tanto no se ve afectado en la práctica por estos supuestos.

Se espera que se dé cumplimiento si en cada observación la varianza permanece en un rango constante, grandes alteraciones afectan el estimador y con él, su validez. En la práctica es plausible el incumplimiento de la varianza constante debido a las condiciones de experimentación u obtención de la información, transformaciones erróneas de la variable, la asimetría

en una distribución, la presencia de datos atípicos, mezcla de subpoblaciones en los datos de medición, interacciones entre variables que no se han contemplado en el modelo -omisión de la variable y/o la interacción-, o que los datos pueden diferir en contextos temporales o espaciales, c.f. [26] y [13].

Aunque la heterocedasticidad no afecta la insesgadez del estimador o la consistencia, si la eficiencia, es decir, existe otro estimador con una varianza mínima que no corresponde con el de mínimos cuadrados ordinarios, por lo tanto, el cálculo de la matriz de varianzas y covarianzas no queda apropiadamente estimada, tampoco los cálculos de errores estándar, que a su vez, su uso incurre en predicciones, intervalos de confianza, y pruebas de hipótesis.

Dado que se puede sobrestimar el valor de la varianza, no se garantiza que en un intervalo de confianza se cubra el nivel nominal de cobertura propuesto. Decimos que la varianza esta sobreestimada debido a que al no haber una varianza constante, la alta dispersión entre las varianzas permiten la existencia de otro estimador \hat{x}_2 tal que $var[\hat{x}_1] \geq var[\hat{x}_2]$.

En pruebas de hipótesis, dado que se suelen establecer los resultados a través del uso del error estándar, los cuales están sobreestimados, entonces, por ejemplo, en un modelo a través de la prueba F o prueba t para un ANOVA, no quedará evidentemente bien estimado, incurriendo probablemente en decisiones erróneas como lo es rechazar la significancia de un modelo o la contribución de una variable independiente en el modelo cuando si es significativa, o viceversa.

Algunas alternativas para tratar la heterogeneidad en las varianzas pueden estar relacionadas con el uso de errores robustos, la técnica es conocida como el estimador de White, o estimador sandwich, que según [30] “El estimador de White para los errores estándar nos ayuda a evitar calcular estimaciones de intervalo incorrectas o valores incorrectos para las estadísticas de prueba en presencia de heterocedasticidad.” El autor añade a continuación que de ninguna manera esto garantiza que deje de existir otro estimador que sea el mejor entre la clase de estimadores lineales e insesgados.

Otras técnicas pueden estar relacionadas con el uso de mínimos cuadrados ponderados, pues en caso de conocer las varianzas puede generar estimadores MELI; uso de transformación de las variables; también el uso de mínimos cuadrados generalizados en presencia de heterocedasticidad pueden ser una alternativa adecuada, en tal caso se puede demostrar que el estimador es MELI a través de este modelo; Se aclara que son técnicas que sirven como alternativas en caso de presentar heterocedasticidad, es decir, previo a este proceso hubo un estudio sobre su detección como la visualización gráfica de las observaciones versus los residuales que aunque subjetiva y limitada puede ofrecer información para su detección, o las pruebas formales como Breuch-Pagan, White, de Levene, Goldfelt-Quandt, etc, c.f. [26].

Otro supuesto es la autocorrelación de una serie de datos en un modelo lineal, según [26], trae problemas al igual que la heterocedasticidad, en que a pesar de no perder la insesgadez, ni la consistencia, sucede que el aumento en el número de la muestra no es garantía de la

solución de la autocorrelación, además se ve afectada la eficiencia del estimador. Un modelo que usa la información otorgada sobre la autocorrelación es el modelo de mínimos cuadrados generalizados, y según *idem.*, se puede obtener así un estimador que puede ser MELI a diferencia del estimador de mínimos cuadrados ordinarios.

Medidas de detección de la autocorrelación puede ser el uso de un correlograma, la prueba de Durbin Watson o aunque menos conocida, se encuentra la prueba de Breusch Godfrey. Al igual que el caso de heterocedasticidad, la variedad de las pruebas es de vital importancia para reconocer de manera adecuada el tipo de prueba que se adecua mejor a los datos obtenidos, solo así, se garantiza una consistencia y validez de la información que se analiza.

Tras su detección, como algunas posibles alternativas se encuentra la transformación de variables, el uso de mínimos cuadrados generalizados, o una nueva especificación del modelo, aunque el nuevo planteamiento propone problemas de otro tipo y que no garantizan que se solucione la autocorrelación de las variables, debido a que puede ser por la naturaleza de la información a tratar, tal es el caso en estadística espacial o series de tiempo.

Aunque en los enunciados sobre el teorema de Gauss-Markov no se hace referencia sobre la normalidad de los residuales, se realiza la especificación dentro del mismo, o la concatenación de ambos hechos, para simplificar y generalizar el comportamiento que éstos deben tener, pues mediante la distribución normal, la asignación de la distribución de probabilidad sobre los estimadores \hat{x}_i y sobre la varianza $\hat{\sigma}^2$ pueden declararse, desencadenando en la distribución para las pruebas de hipótesis e intervalos de confianza, este hecho ampliamente estudiado se ve estructurado por el teorema de límite central, las demás ventajas ofrecidas por esa suposición son las bien conocidas, facilidad de su tratamiento a partir de valores como la media y la varianza, y la conveniencia de que una gran multitud de fenómenos a estudiar tienen un comportamiento aproximadamente normal.

Se puede derivar otro inconveniente en el ajuste del modelo, esto es cuando se presenta una relación entre las variables independientes, denominamos colinealidad o multicolinealidad cuando una de las variables independientes es combinación lineal de otra variable independiente o más de dos, respectivamente, las consecuencias afectan la estimación de \hat{b} ; la matriz $(A^T A)^{-1}$, tendrá una columna que es combinación lineal de otras, como consecuencia la matriz es singular y no se puede estimar el valor de \hat{b} apropiadamente [15].

Al haber multicolinealidad, la especificación del modelo evidencia información redundante, es de alta importancia lidiar con dicho problema para no invalidar las estimaciones, la presencia de valores de correlación altos entre las variables puede ser prueba para su detección, aunque no sea exacta la correlación entre las variables, una casi combinación lineal exacta, genera una matriz casi singular, y aunque de este modo si hay solución del sistema, la matriz pasa a ser de tipo mal condicionada.

Se ve generada la multicolinealidad debido al sobreajuste del modelo, esto es, cuando en el modelo se incluye una alta cantidad de variables, entonces aumentan las posibilidades de que

se encuentren correlacionadas, un tamaño de muestra pequeño puede ser causante de este problema, o en estudios transversales o longitudinales, se puede presentar que con el tiempo ciertas variables se comporten de manera similar [26].

Métodos para la solución de la multicolinealidad siguiendo a [17] son el cálculo de las correlaciones entre las variables dependientes, también el cálculo de los autovalores de la matriz A , y así, aquellos cuyo valor sea cercano a cero indican la presencia de multicolinealidad, o a través del número de condición de la matriz A , el cual es un indicador cuando se presentan grandes perturbaciones en los valores de entrada para determinar el impacto en los valores de salida. Otros estadísticos comunes para la detección de la multicolinealidad son el factor de inflación de la varianza (FAV), el cual mide el cambio de la varianza de algún coeficiente de regresión cuando se presenta la multicolinealidad entre las variables, entonces para un coeficiente a_i , el FAV se expresa como

$$FAV_i = \frac{1}{1 - R_i^2}$$

Aunque no se encuentra establecido, desde enfoques teóricos e incluso desde enfoques prácticos se sugiere la presencia de multicolinealidad si $FAV \geq 10$.

En [25] se tienen técnicas para su solución como la regresión Ridge, que hace parte de las técnicas de regularización, en ésta técnica se imponen penalidades en el vector de coeficientes, para este caso particular se busca eliminar la dependencia entre las variables, si se identifican, se busca la omisión de la variable en el modelo a través de $(A^T A + \lambda I)^{-1} A^T b$, donde λ es un término de penalidad tal que, en general $\lambda \geq 0$; este método funciona como un control sobre las variables de entrada, para que pasen a hacer parte de un modelo de estudio y análisis.

Siguiendo a [26] “Si la multicolinealidad es perfecta, los coeficientes de regresión de las variables x son indeterminados, y sus errores estándar, infinitos. Si la multicolinealidad es menos que perfecta, los coeficientes de regresión, aunque sean determinados, poseen grandes errores estándar (en relación con los coeficientes mismos), lo cual significa que los coeficientes no pueden ser estimados con gran precisión o exactitud”.

El análisis de los supuestos que se generan a partir del teorema de Gauss Markov, que garantiza la optimalidad de las estimaciones para el valor de \hat{b} a través de \hat{x} , mediante las ecuaciones normales, si bien abren un panorama de fiabilidad estadística, no es la única perspectiva por analizar, pues aunque se cumplan todos los supuestos, otras maneras de donde la influencia de los resultados puede ser sesgada o alterada es a través de los algoritmos que lo generan y de la fundamentación matemática sobre la cual se desarrolló.

El motivo de la mención de este factor es que se debe tener en cuenta la naturaleza de los datos de estudio, pues el uso de las ecuaciones normales, con todo lo que conlleva su uso, es aplicable convenientemente para algunos casos, si se trabaja con una gran cantidad de información el cálculo de la matriz $(A^T A)^{-1}$ puede ser de un trabajo ineficiente, respecto de otras técnicas que no involucran la propensión a errores de redondeo. A continuación,

y siguiendo con un enfoque matricial, se aborda el problema a través de descomposiciones matriciales.

2.4. Descomposición del problema de mínimos cuadrados

La solución a la estimación de los parámetros se puede encontrar mediante la descomposición de una matriz, permite facilitar los procesos buscando eficacia sin dejar de tener en cuenta la veracidad, exactitud y precisión de las estimaciones, una amplia gama de técnicas ha sido de alta implementación e investigación. Éstos métodos se basan en su mayoría en la noción de transformación ortogonal, véase Apéndice A, que se basan en la descomposición de una matriz para facilitar la ejecución de los procesos.

En este trabajo nos centraremos en la descomposición QR, donde las transformaciones matriciales realizadas en este método son tan precisas como en métodos numéricos o basados en la descomposición de valores singulares (DVS) [56].

Sin embargo, cuando la matriz no tiene rango completo una buena alternativa es usar DVS, tiene la ventaja de que a través de este método se puede calcular la matriz pseudoinversa en esta factorización matricial, aunque carece de la eficiencia y de la estabilidad numérica de otros métodos además de posibles costos computacionales altos [8], [56] en razón de no garantizar la unicidad de las matrices en las cuales se descompone A . Otro tipo de factorización como LU o Choleski, dependen de que la matriz tenga características de simetría y/o que sea definida no negativa, por lo tanto es una restricción que no permite ser implementada de manera general.

A pesar del gran uso e implementación de los modelos de mínimos cuadrados ordinarios es necesario tener en cuenta la falta de fiabilidad, que puede deberse a un escenario relacionado con el incumplimiento de los supuestos que se indican en el teorema de Gauss-Markov, su incumplimiento acarrea invalidez desde distintas perspectivas como se explicaron previamente, de allí, el surgimiento de alternativas como mínimos cuadrados ponderados, robustos, totales, parciales, restringidos, etc., que permiten tratar conforme la naturaleza de la medición de los datos, de las condiciones experimentales, como la aparición de valores atípicos e influyentes, de la información adicional, u otros factores a tener en cuenta en el análisis por realizar.

Otras alternativas que han surgido con el objetivo de garantizar estimaciones fiables en mínimos cuadrados, podemos encontrar los métodos de regularización. En estos métodos se imponen penalizaciones en los estimadores de la función a estimar ayudando a reducir la varianza, como era el caso de la regresión Ridge tratada anteriormente, método que también pretenden solucionar problemas de colinealidad de las variables.

Otro tipo de solución para los métodos es a través de métodos numéricos, bien sea basado en eliminación gaussiana, mínimos cuadrados recursivos y otros métodos a los que el lector puede referirse en [8].

Entre todos los métodos, son de amplia utilidad e interés aquellos basados en descomposiciones matriciales, debido a la alta eficiencia, exactitud, facilidad de implementar computacionalmente. Podemos encontrar la descomposición de valores singulares, factorización LU, Choleski o factorización QR, éste último a través de 3 métodos esencialmente: reflexiones de Householder, rotaciones de Givens o el proceso de ortogonalización de Gram-Schmidt. En el Apéndice A se exponen los fundamentos relacionados con la factorización QR para los tres métodos, y en adelante se delimita el uso de esta factorización para la teoría expuesta.

2.4.1. Factorización QR en la solución de mínimos cuadrados

Sea $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^{m \times 1}$, además sea una matriz ortogonal $Q \in \mathbb{R}^{m \times m}$ y una matriz triangular superior $R \in \mathbb{R}^{m \times n}$, el problema de mínimos cuadrados lineales es

$$\min_x \|Q^T (Ax - b)\|$$

Como se indica en el Apéndice A, una transformación ortogonal es invariante al utilizar la norma euclidiana. Siguiendo a [21] en el caso donde en A es de rango completo la descomposición de A mediante matrices por bloque es de la forma

$$A = Q \begin{pmatrix} R_1 & 0 \end{pmatrix}$$

Sea c un vector definido a través de la reexpresión

$$Q^T y = c = \begin{pmatrix} c_1 & c_2 \end{pmatrix}$$

donde $c_1 \in \mathbb{R}^m$ y $c_2 \in \mathbb{R}^{n-m}$, en palabras, la partición del vector c se realiza conforme la partición de la matriz A en términos de R_1 y del vector 0, que a su vez, depende estrictamente del rango de la matriz A , pues $\rho(A) = \rho(R_1) = m$

Por lo tanto, la minimización 2-2 se expresa

$$\begin{aligned} (b - Ax)^T (b - Ax) &= (b - QR_1x)^T (b - QR_1x) \\ &= (Q^T b - R_1x)^T (Q^T b - R_1x) \\ &= (c_1 - R_1x)^T (c_1 - R_1x) + c_2^T c_2 \end{aligned}$$

Debido a que esta última ecuación es no-negativa, el valor mínimo de la norma residual al cuadrado se da cuando $(c_1 - R_1x)^T (c_1 - R_1x) = 0$. De donde se obtiene la solución

$$x = R^{-1}c_1$$

Así las cosas, la solución a través de la factorización QR tiene una solución sencilla para el caso de una matriz A de rango completo, lo restante dependerá pues, del método para el cálculo de la factorización QR, sea reflexiones de Householder, rotaciones de Givens u ortogonalización de Gram Schmidt.

3. Mínimos cuadrados restringidos

Qué es el hombre se pregunta Pascal: Una potencia de exponente cero. Nada si se compara con el Todo. Todo si se compara con la nada: Nacimiento más muerte: Ruido multiplicado por silencio: Medio aritmético entre el todo y la nada.

Nicanor Parra. Pensamientos

El modelo de mínimos cuadrados ordinarios, como bien se indicó en el capítulo pasado, puede presentar inconvenientes al incumplir los supuestos de los residuales, obteniéndose un estimador sin las propiedades MELIO u otras consecuencias prácticas.

Como consecuencia práctica, es posible que las estimaciones no se acoplen adecuadamente a las condiciones del problema por solucionar, dando como resultado, coeficientes que no se ajusten a los valores prácticos adecuados, sin factibilidad alguna de interpretación.

En caso de ser posible que se cuente con información auxiliar para no depender únicamente de los datos obtenidos, la estructura del modelo, las consecuencias teóricas en las propiedades de los estimadores, los métodos de solución y las consecuencias prácticas se pueden ver alteradas drásticamente respecto del modelo de mínimos cuadrados ordinarios.

La información a priori “puede provenir de diferentes fuentes como la experiencia del pasado o largas asociaciones del experimentador con el experimento, u otros tipos de conductas en el pasado” [42]. La inclusión de la información a priori en el modelo, puede ser añadida a través de restricciones que al imponerlas en el modelo lineal, pueden suplir las desventajas del estimador de mínimos cuadrados ordinarios, además de que se adecuan a soluciones que si tienen interpretación en el contexto del problema.

Distintos tipos de información a priori en forma de restricciones pueden ser incluidos en el modelo,

- **Igualdad**
- Desigualdad
- No lineal
- Estocástico

A través de igualdades, en forma de restricciones de un solo componente $x_1 = 0$ o de una relación entre componentes $x_2 = 3x_3$, de manera general nos referiremos a este sistema de ecuaciones lineales de la forma $Cx = d$, véase Idem. Nótese que si $d = 0$, el modelo de mínimos cuadrados restringidos coincide con el modelo de mínimos cuadrados ordinarios, en otras palabras, esté ultimo es un caso particular de mínimos cuadrados restringidos.

A través de desigualdades pueden ser encontradas en su forma más común como no-negatividad $x \geq 0$ o convexidad $\lambda x + (1 - \lambda)x$; también a través de restricciones de tipo no lineal xAx^T , u otro tipo de restricciones como combinaciones entre éstas $\|x\|_2 \leq \alpha$, etc. Sin embargo, la inferencia cambia para el caso de restricciones de tipo desigualdad o no lineal, además no hay una formula general para su solución, pues depende de la estructura de la restricción impuesta en el modelo [21]. Finalmente para las restricciones estocásticas se considera que la información a priori se satisface de manera aproximada $Cx + \phi = d$ con $\phi \approx (0, \sigma^2 V)$, sus fundamentos pueden ser vistos en [42].

Debido a la inferencia estadística que fundamenta de manera muy particular a aquellas restricciones que son de tipo no igualdad, el alcance de este trabajo se limita ostentosamente, así, se añade más detalle y fundamentación en el modelo de mínimos cuadrados restringidos a través de restricciones de igualdad.

Definición 2 (Mínimos cuadrados restringidos). *La representación del modelo de mínimos cuadrados restringidos a través de restricciones de igualdad es*

$$\min_{Cx=d} \|Ax - b\| \quad (3-1)$$

Siendo $A \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^{m \times 1}$ y $d \in \mathbb{R}^{p \times 1}$.

Las restricciones $Cx = d$ son otro sistema de ecuaciones lineales, del cual ya se conoce su propósito al añadirse en el modelo lineal, si este sistema de ecuaciones lineales cumple la pauta de que C tenga rango completo $\rho(C) = n$, entonces en tal caso, el sistema tiene solución única, y así, también el modelo de mínimos cuadrados restringidos, la notación aunque cambie más adelante en términos del rango y la nulidad, expresará la misma idea recién expuesta.

3.1. Estimador para mínimos cuadrados restringidos

En 1961 Henri Theil [50] propuso un estimador directo que permite el cálculo del vector que satisfaga la ecuación 3-1 [51].

Definición 3 (Estimador para MCR). *Basado en el uso de multiplicadores de Lagrange, una expresión para la obtención del vector de estimadores x_{MCR} es:*

$$x_{MCR} = \hat{x} + (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right]^{-1} [d - C\hat{x}] \quad (3-2)$$

El cálculo se basa en el uso de multiplicadores de Lagrange, de tal manera que las restricciones se incorporan al modelo y la solución se genera a partir de las condiciones de primer orden en el nuevo modelo de n variables originales más k restricciones. Se usa indistintamente x y x_{MCR} como la solución del modelo y \hat{x} como el estimador de mínimos cuadrados ordinarios

Demostración.

$$\begin{aligned} \min \varepsilon^T \varepsilon &= (Ax - b)^T (Ax - b) - 2\lambda (Cx - d) \\ \min \varepsilon^T \varepsilon &= b^T b - 2x^T A^T b + x^T A^T A x - 2\lambda^T (Cx - d) \end{aligned}$$

derivando respecto a x y λ tenemos

$$\frac{\partial}{\partial x} = -2A^T b + 2A^T A x_{MCR} + C^T \lambda = 0 \quad (3-3)$$

$$\frac{\partial}{\partial \lambda} = 2Cx_{MCR} - 2d = 0 \quad (3-4)$$

Este sistema puede escribirse a través de matrices [24], donde se evidencia que el sistema tiene solución si tiene rango completo la primer matriz

$$\begin{bmatrix} A^T A & C \\ C & 0 \end{bmatrix} \begin{bmatrix} x_{MCR} \\ \lambda \end{bmatrix} = \begin{bmatrix} A^T b \\ d \end{bmatrix} \quad (3-5)$$

Al multiplicar 3-3 por el factor $C(A^T A)^{-1}$ se obtiene

$$\begin{aligned} -2C(A^T A)^{-1}A^T b + 2C(A^T A)^{-1}A^T Ax_{MCR} + C(A^T A)^{-1}C^T \lambda &= 0 \\ -2C(A^T A)^{-1}A^T b + 2Cx_{MCR} + C(A^T A)^{-1}C^T \lambda &= 0 \end{aligned}$$

Si la restricciones se suponen ciertas, entonces de 3-4 se obtiene $Cx = d$, así

$$-2C(A^T A)^{-1}A^T b + 2d + C(A^T A)^{-1}C^T \lambda = 0$$

Por lo tanto el valor de λ es

$$\lambda = -2 \left[C(A^T A)^{-1}C^T \right]^{-1} \left[d - C(A^T A)^{-1}A^T b \right]$$

Este valor se sustituye en 3-3

$$-2A^T b + 2A^T Ax_{MCR} - 2C^T \left[C(A^T A)^{-1}C^T \right]^{-1} \left[d - C(A^T A)^{-1}A^T b \right] = 0$$

Finalmente, el valor de x_{MCR} se obtiene al premultiplicar el factor $(A^T A)^{-1}$ con el fin de despejar dicha variable, entonces

$$-2(A^T A)^{-1}A^T b + 2(A^T A)^{-1}A^T Ax_{MCR} - 2(A^T A)^{-1}C^T \left[C(A^T A)^{-1}C^T \right]^{-1} \left[d - C(A^T A)^{-1}A^T b \right] = 0$$

$$-2(A^T A)^{-1}A^T b + 2x_{MCR} - 2(A^T A)^{-1}C^T \left[C(A^T A)^{-1}C^T \right]^{-1} \left[d - C(A^T A)^{-1}A^T b \right] = 0$$

Al reemplazar el estimador de mínimos cuadrados ordinarios $\hat{x} = (A^T A)^{-1}A^T b$, se obtiene el estimador

$$x_{MCR} = \hat{x} + (A^T A)^{-1}C^T \left[C(A^T A)^{-1}C^T \right]^{-1} [d - C\hat{x}]$$

□

La diferencia tras la obtención del estimador restringido y el no restringido en palabras de [51] se expone así “Nosotros podemos concluir que la diferencia entre el vector de coeficientes de mínimos cuadrados entre el restringido y el no restringido es una función lineal del vector $(Cx - d)$, que mide el grado en el cual el último vector de coeficientes no satisface las restricciones”.

Aunque la obtención de la solución es directa, pues depende de valores conocidos, también depende del estimador de mínimos cuadrados ordinarios, precisamente la dependencia genera una deficiencia del estimador restringido dado el no restringido. Por ejemplo, si la matriz de diseño tiene problemas de dependencia lineal entre sus columnas, no sería posible obtener una estimación para el estimador a través de este método, siendo necesario solucionar el problema de la multicolinealidad, por lo tanto, al no ser completamente dependiente de la información extraída por el experimento, como menciona sobre 3-2 en [24] “No debe ser vista como una definición general del estimador de mínimos cuadrados restringidos desde que x_{MCR} no pueda ser definido en términos del estimador de mínimos cuadrados sin restricciones \hat{x} si este último no existe”.

A pesar de este hecho, es uno de los métodos comúnmente utilizados, aunque los procesos necesarios para el cálculo del vector de estimadores son costosos en términos de las inversas que requiere la expresión 3-2, y aún más en un gran volumen de datos, pues se recae en errores acumulados de redondeo que afectan la exactitud, [59] afirma que el método es ineficiente para su propósito, como también lo afirma [23]. Pese a sus desventajas, este método ha sido estudiado a fondo y suele ser utilizado también para estudiar los momentos del estimador, cuyos fundamentos se exponen en la siguiente sección.

3.1.1. Momentos del estimador directo

Los momentos de la esperanza y la varianza del estimador de la ecuación 3-2 se analizan a continuación.

La insesgidez del estimador de mínimos cuadrados restringidos coincide con la esperanza del estimador de mínimos cuadrados ordinarios, es decir, es un estimador insesgado. Pero esta afirmación se mantiene solo en el caso donde las restricciones impuestas al modelo son verdaderas, pues solo así, el segundo término se ve anulado dado que $d - C\hat{x} = 0$, obteniéndose

$$E(x_{MCR}) = E \left[\hat{x} - (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right]^{-1} [d - C\hat{x}] \right]$$

$$E(x_{MCR}) = E \left[\hat{x} - (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right]^{-1} \right] E [d - C\hat{x}]$$

$$E(x_{MCR}) = E(\hat{x}) - E \left[(A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right]^{-1} 0 \right]$$

$$E(x_{MCR}) = E(\hat{x}) = x$$

En caso contrario se añade un sesgo por el incumplimiento de las restricciones, la consecuencia es evidente, las estimaciones obtenidas distan de ser aquellas con las que se pueda garantizar que el estimador es MELI. Ahora analicemos la varianza del estimador 3-2.

El valor del estimador no restringido se reescribe en términos del parámetro x y del residual ε , entonces

$$\begin{aligned}\hat{x} &= (A^T A)^{-1} A^T b \\ \hat{x} &= (A^T A)^{-1} A^T (Ax + \varepsilon) \\ \hat{x} &= (A^T A)^{-1} A^T Ax + (A^T A)^{-1} A^T \varepsilon \\ \hat{x} &= x + (A^T A)^{-1} A^T \varepsilon\end{aligned}\tag{3-6}$$

Esta expresión se reemplaza en la ecuación 3 – 2

$$x_{MCR} = x + (A^T A)^{-1} A^T \varepsilon + (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right]^{-1} [d - C(x + (A^T A)^{-1} A^T \varepsilon)]$$

$$x_{MCR} = x + (A^T A)^{-1} A^T \varepsilon + (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right]^{-1} [d - Cx - C(A^T A)^{-1} A^T \varepsilon]$$

Asumiendo que las restricciones son verdaderas, se anula el término $d - Cx$

$$x_{MCR} = x + (A^T A)^{-1} A^T \varepsilon + (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right]^{-1} [-C(A^T A)^{-1} A^T \varepsilon]$$

$$x_{MCR} = x - (A^T A)^{-1} A^T \varepsilon + (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right]^{-1} [C(A^T A)^{-1} A^T \varepsilon]$$

$$x_{MCR} - x = \left[I - (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right]^{-1} C \right] (A^T A)^{-1} A^T \varepsilon$$

Por simplicidad, se propone la siguiente sustitución

$$M = [I - (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right]^{-1} C]$$

Así, se continua escribiendo el valor de $x_{MCR} - x$ como

$$x_{MCR} - x = M(A^T A)^{-1} A^T \varepsilon$$

Según la expresión obtenida, la estimación de la varianza sobre el estimador de interés, inicia según su definición, es decir,

$$\begin{aligned}V(x_{MCR}) &= E[(x_{MCR} - x)(x_{MCR} - x)'] \\ V(x_{MCR}) &= E [M(A^T A)^{-1} A^T \varepsilon \varepsilon A^T (A^T A)^{-1} M^T]\end{aligned}$$

$$V(x_{MCR}) = M(A^T A)^{-1} A^T \sigma^2 I A^T (A^T A)^{-1} M^T$$

$$V(x_{MCR}) = M(A^T A)^{-1} A^T \sigma^2 A^T (A^T A)^{-1} M^T$$

$$V(x_{MCR}) = \sigma^2 M(A^T A)^{-1} M^T$$

Con más tratamiento matricial, puede llegarse a otra expresión como en [42] que es la forma habitual en la literatura de expresar la varianza del estimador

$$V(x_{MCR}) = \sigma^2 (A^T A)^{-1} - \sigma^2 (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right] R (A^T A)^{-1}$$

Ahora bien, en comparación de las varianzas para el estimado restringido y el no restringido, la diferencia recae en el segundo término, pues

$$V(x) - V(x_{MCR}) = \sigma^2 (A^T A)^{-1} - \sigma^2 (A^T A)^{-1} - \sigma^2 (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right] R (A^T A)^{-1}$$

$$V(x) - V(x_{MCR}) = \sigma^2 (A^T A)^{-1} C^T \left[C (A^T A)^{-1} C^T \right] R (A^T A)^{-1} \geq 0$$

Es decir, el modelo restringido a través de este término, demuestra tener una varianza mejor respecto del modelo de mínimos cuadrados ordinarios, pues es una matriz definida no negativa.

Tras lo analizado en los momentos del estimador, se puede observar, que si las restricciones son verdaderas, el estimador es insesgado, pues coincide su esperanza con el de mínimos cuadrados ordinarios. En cambio, la varianza del estimador al ser una matriz semidefinida no negativa, tiene una varianza menor respecto del estimador no restringido, esta menor varianza tiene impacto en intervalos de confianza más pequeños, y a su vez, estos se ven reflejados en intervalos de confianza para tener predicciones más acertadas, al igual que en otros cálculos basados en errores estándar, como el desarrollo de pruebas de significancia t y F .

Un estimador restringido insesgado y de varianza mínima es lo deseable, sin embargo, en la práctica no siempre es plausible la obtención de un estimador con dichas características, es posible que sea sesgado, pero aún así se garantiza que su varianza es menor al estimador de mínimos cuadrados ordinarios. Entonces, a pesar de que no hay criterios unificados y unánimes sobre la elección del estimador, es posible que se garantice fiabilidad en el estimador restringido con menor varianza a coste de un incremento en el sesgo.

Finalmente se recalca en la desventaja del estimador 3-2, es que depende del estimador de mínimos cuadrados ordinarios, luego las falencias expuestas en para los supuestos de mínimos cuadrados ordinarios recaen para este estimador, por lo tanto, si existe dependencia lineal en las columnas de A , habrán inconvenientes para el cálculo del estimador, y se perdería la validez en los resultados.

3.1.2. Pruebas para las restricciones lineales

La comparación entre un modelo que presente restricciones y el modelo no restringido, se puede solventar a través de pruebas de hipótesis, mediante comparación directa, ya sea a través de la prueba t de Student o F de Fisher. La necesidad de su comparación radica en conocer si al añadir restricciones al modelo, es decir, involucrando el coste, las cualidades del estimador o la información a priori que permite generar las restricciones; este cambio representa una mejora en el ajuste del modelo, o en otras palabras, una mejoría en las estimaciones realizadas, pues en caso contrario, una mejoría que no sea lo suficientemente significativa puede recaer en el rechazo del modelo con las restricciones y preferir resultados aceptables basados en un modelo parsimonioso como lo es mínimos cuadrados ordinarios.

Consideremos un tipo de restricciones de tipo exclusión, que es quizás una de las restricciones comúnmente utilizadas en los análisis de los modelos, es decir, $x_i = 0$, pues de esta misma manera se mide la significancia de una variable en un modelo, que usualmente es puesta a prueba a través de la prueba t de Student. También se pueden probar de manera simultánea $x_i = x_j = x_k = \dots = 0$ o de la forma $x_i + \alpha x_j = 0$ con $\alpha \in \mathfrak{R}$, que son restricciones obtenidas meramente según información adicional, ó información que se denomina extramuestral, así, es como la información a priori se puede justificar y hay validez al realizar la imposición de las restricciones en el modelo por ajustar.

En caso de ser verdaderas las restricciones para este caso particular de una restricción de tipo exclusión, se omite información en el modelo, un ejemplo de utilidad para su uso, es un modelo donde se presenta sobreajuste, allí debe haber evidencias de multicolinealidad entre las variables independientes, a través de una alta correlación entre más de dos variables, entonces no se puede afirmar que éstas sean independientes como se afirma en la notación general del modelo, pues, recordemos que desde un enfoque matricial, los cálculos de las inversas de la matriz de diseño, no tendrían inversa, o harían que la matriz pase a ser mal condicionada, y desde otro enfoque para afirmar la misma idea, una matriz con columnas linealmente dependientes, es decir, una matriz de rango deficiente, indicaría que el espacio columna sobre el cual se busca hacer la proyección del vector respuesta, tiene vectores que no permiten formar una base que genere el espacio, dado que entre los vectores no se presenta ortogonalidad.

En las pruebas de hipótesis el estadístico de prueba bajo la hipótesis nula de una variable sería $H_o : x_i = \beta$. La fórmula general estaría dada por:

$$t = \frac{\hat{x}_i - \beta}{ee(\hat{x}_i)}$$

Para el caso de restricciones con exclusión, se plantea la hipótesis nula de una variable como sería $H_o : x_i = 0$, y el estadístico de prueba se determina basado en

$$t = \frac{\hat{x}_i}{ee(\hat{x}_i)}$$

Siendo \hat{x}_i la variable a juzgar y $ee(\hat{x}_i)$ el error estándar de la variable. Dicho estadístico de prueba, es comparado con el percentil $1 - \alpha$ de una distribución $t_{\alpha, n-k}$, siendo α el nivel de significancia escogido, a una o dos colas, y $n - k$ los grados de libertad, siendo este valor la diferencia entre el número de variables y el número de restricciones, que para el caso de una exclusión se tiene $n - 1$.

Además de esta prueba, una comparación directa pero evaluada con mayor generalidad en el modelo, permite probar el valor o la significancia de varias variables en el modelo a través de la prueba F de Fisher. En el enfoque de la prueba t pueden surgir complicaciones cuando se presenta más de una variable, pues la prueba t pierde unanimidad sobre la decisión a elegir debido a las posibles decisiones que puede tener sobre cada variable [57].

Recordemos que el objetivo del ajuste es la minimización de los residuales del modelo, es decir, de la suma de cuadrados de los residuales (SCR), debido a que buscamos comparar el modelo restringido del no restringido, usemos respectivamente la notación así SCR_R y SCR_{NR} .

Esta prueba consiste en una comparación directa de la suma de cuadrados de ambos modelos. Es necesario considerar estos valores, además de los grados de libertad de los modelos, pues recordemos que el valor de un grado de libertad está directamente relacionado con la cantidad de restricciones que se le impone al modelo de estudio.

Definición 4 (Prueba F para MCR). *El estadístico de contraste para probar la significancia de las restricciones está definido por*

$$F = \frac{(SCR_{NR} - SCR_R)/q}{SCR_{NR}/n - k - 1}$$

Siendo además de los valores SCR_R y SCR_{NR} , n es el número de observaciones, k el número de variables independientes y q el número de restricciones impuestas.

Es decir, el valor del estadístico F, es el cambio relativo que tiene la suma de cuadrados de los residuales al pasar del modelo no restringido al restringido [57].

Siguiendo con el análisis para el caso de las restricciones de exclusión, al no hacer la inclusión de las variables en el modelo, el valor de la suma de cuadrados de los residuales es evidentemente menor que en el modelo que no tiene restricciones, entonces la diferencia entre ambas cantidades será positiva, y también así el valor de estadístico F *idem*.

Así las cosas, evidente se hace el hecho de que para comparar los modelos, se deben tener las estimaciones de ambos modelos. Por otra parte, como se evidenció en el caso particular estudiado, de manera general se mantiene la premisa de que el valor de SCR_R bajo ninguna circunstancia puede ser inferior al valor de SCR_{NR} , pues, debido a que se realizan acotaciones mediante las restricciones al subespacio de soluciones que debe satisfacer el vector de coeficientes, en el peor caso, estas restricciones coinciden con el modelo no restringido, en conclusión, el estadístico es siempre positivo.

Otra versión en *idem*, para el cálculo del mismo estadístico se basa en los valores R-cuadrado¹ de cada modelo, formulándose

$$F = \frac{(R_{NR}^2 - R_R^2) / q}{(1 - R_{NR}^2) / (n - k)}$$

A continuación se realiza un análisis de la prueba de manera general. Se busca probar la hipótesis $H_0 : Jx = l$, comúnmente $l = 0$ y $J \in \mathbb{R}^{q \times p}$ es la matriz del contraste a probar, siguiendo el planteamiento expuesto por [55] es una prueba de hipótesis relevante sobre la compensación del sesgo y la varianza del estimador restringido. Se puede obtener que el valor de F está dado por la forma cuadrática

$$F = \frac{[Jx - l]^T [J(A^T A)^{-1} J^T]^{-1} [Jx - l]}{q\sigma^2}$$

Sin embargo, se evidencia que la hipótesis nula aún no considera la imposición de las restricciones $Cx = d$, por lo tanto, un factor de corrección relacionado con el grado de cumplimiento de las restricciones debe ser añadido. Tras este hecho, la hipótesis nula ha sido alterada de manera implícita a la siguiente forma

$$\bar{H}_0 : J(x - W(Cx - d)) = m$$

Con $W = (A^T A)^{-1} C^T [C (A^T A)^{-1} C^T]^{-1}$. Entonces se define al valor para F del estimador restringido así

$$F_R = \frac{[Jx - W(Cx - d) - l]^T [J(A^T A)^{-1} - W(A^T A)^{-1} J^T]^{-1} [Jx - W(Cx - d) - l]}{q\sigma^2}$$

Si las restricciones se cumplen $Cx - d = 0$, el factor de corrección se anula, y así coincide el valor del estadístico F restringido con el no restringido, en caso de no ser verdadera esta afirmación, se presenta un caso adicional donde aún pueden coincidir los valores de ambas estimaciones según [55], esto es para el caso $JW = 0$, o equivalentemente $J(A^T A)^{-1} C^T = 0$, es decir, que exista ortogonalidad entre las matrices J y W .

Como caso particular suele usarse una distribución central, esto es, bajo la hipótesis nula la referencia a probar es el valor cero, en caso no ser así, se usa una distribución no central, añadiéndose un parámetro de no centralidad λ ; al ser falsa la hipótesis nula -pues el valor referencia es diferente de cero-, ésta deberá ser rechazada evidentemente, no hacerlo implica una decisión incorrecta, que es cometer un error tipo II, por lo tanto el uso de una distribución no central es útil el cálculo de la potencia estadística [33].

¹En la comparación de los valores R_{NR}^2 y R_R^2 , la fórmula es aplicable si la variable respuesta es la misma en ambos casos (eso incluye si es una variable transformada), de lo contrario no son directamente comparables.

Téngase en cuenta que la distribución $F(\nu_1, \nu_2, \lambda)$, ahora considera 3 parámetros dado que se considera como no central, siendo los dos primeros los grados de libertad que tienen ambas distribuciones provenientes de la distribución χ^2 Chi cuadrado ² y el tercero es el parámetro de no centralidad.

En ambos casos expuestos de la prueba, es decir, para F y F_R , éste parámetro de no centralidad tiene una forma cuadrática en su representación [55], cuanto mayor sea este valor -el parámetro de no centralidad-, para el caso de la distribución F se observa que sus valores tienden hacia la derecha, cambiando la forma de distribución de la función, entonces se encuentran más valores a la derecha del nivel de significancia escogido α , aumentando la probabilidad de rechazar la hipótesis nula cuando ésta es verdadera, que es la probabilidad de cometer un error tipo I, pero en contraparte aumenta la potencia de la prueba, la cual es evitar cometer un error tipo II, [55] menciona este mismo hecho explicando que en tal caso H_0 pasa a ser .

Los valores del parámetro de no centralidad λ por su estructura son formas cuadráticas, de manera geométrica se evidencia que la proyección del estimador x_{MCR} se da sobre el subespacio generado por H_0 que corresponde a Jx , alternamente se produce una proyección sobre \bar{H}_0 correspondiente a $J(x - W(Cx - d))$.

La proximidad de estos subespacios, tiene como consecuencia una proximidad entre las proyecciones, así, es válido que consideremos una medida de similitud entre ambos casos a partir de una matriz de covarianzas Σ , dicha matriz tendrá la forma $(C(A^T A)^{-1} C^T)$ para λ y $(C[(A^T A)^{-1} - WC(A^T A)^{-1}] C^T)^{-1}$ para λ_R , cuyo subíndice hace referencia al caso restringido [55]. Se concluye que estas expresiones del parámetro de no centralidad expresan el cuadrado de las distancias de dicho parámetro sobre H_0 y \bar{H}_0 , pues por su forma representan la fórmula de una distancia de Mahalanobis, que por definición considera la similitud entre las variables a medir.

Otro enfoque para la comparación de los modelos se establece en [52] así, “Si uno está dispuesto a aceptar algún sesgo a cambio de una reducción de la varianza, incluso si la restricción no es cierta, aún podría preferir los estimadores restringidos. Ésta noción conduce a una tercera forma de ver las restricciones en un marco de regresión lineal.”

A partir de la estimación de la varianza y del cuadrado del sesgo se obtiene el error cuadrático medio, *idem.* propone la comparación entre ambos modelos según este criterio mediante pruebas de hipótesis, bajo la hipótesis nula se expresa mayor eficiencia del estimador restringido.

²En caso de ser ambas distribuciones χ^2 centrales, también lo será la distribución F , si una de ellas es no central, se obtiene una distribución no central F , y si ambas distribuciones son no centrales, se obtiene una distribución doblemente no central con parámetros $F(\nu_1, \nu_2, \lambda_1, \lambda_2)$, véase [33].

$$H_0 : \lambda \leq 1/2$$

$$H_1 : \lambda > 1/2$$

Se acepta la hipótesis para el caso $W < W_T$, siendo W el estadístico de prueba

$$W = \frac{SSR(\beta) - SSR(b)}{SSR(b)/(m - n)} \quad (3-7)$$

donde $SSR(\beta)$ corresponde a la suma de cuadrados de los residuos en el modelo restringido y $SSR(b)$ representa estadístico en el modelo sin restricciones. Este valor se compara con los valores críticos tabulados para W_T , figura **3-1**, obtenidos a través del nivel de significancia α y de los grados de libertad $m - n$ (expresado en *idem.* como $T - K$). En caso contrario, la no aceptación de la hipótesis se da en el caso $W \geq W_T$. Se asume que $\varepsilon_{Ax-b} \sim iid N(0, \sigma^2)$.

También se expresa la tabla para el cálculo de la potencia de la prueba, la cual puede ser calculada según la tabla que le procede en el artículo, a través del valor de α y λ .

Mediante estos dos enfoques se puede estimar la mejoría entre los modelos a comparar, sin embargo como indica el autor: “Al poner este artículo en perspectiva, asumimos que la prueba MSE no se utilizará de manera mecánica. En cualquier tipo de marco de regresión paso a paso existe la tentación de incluir un grupo de variables para dejar que la máquina las clasifique en función de un criterio.” Por tal motivo, se evita la mecanicidad del proceso al asumir el cumplimiento aproximado de los supuestos y la obtención de un modelo bien planteado o especificado. Además como lo menciona el autor, esta perspectiva es de utilidad cuando adquiere más importancia para el investigador las consecuencias de imponer restricciones en el espacio del parámetro³ en el modelo que la verdad o falsedad de la restricción, pues esta característica es la diferencia para el enfoque de las pruebas recién expuestas.

3.1.3. Algoritmos para el desarrollo de MCR

Una vez se afirma que las restricciones son verdaderas o que se ofrecen mejorías en las cualidades del estimador como la varianza mínima, discutida anteriormente, los métodos de solución varían, según el uso de descomposiciones matriciales, métodos directos u otros basados en métodos numéricos. En esta sección se discutirán métodos matriciales comúnmente utilizados de nivel teórico y de nivel práctico.

Lawson & Hanson en su libro “Solving Least Squares Problem” publicado en 1974, ofrecieron 3 métodos para la solución de mínimos cuadrados restringidos con restricciones de igualdad, cabe resaltar que también discuten y desarrollan otros algoritmos para casos donde se imponen restricciones de tipo desigualdad.

³Todos aquellos valores posibles que puede tomar el parámetro configuran un subconjunto Ω , denominado como espacio del parámetro.

| $T - K$ \ α | .05 | .10 | .25 | .50 |
|--------------------|--------|-------|-------|------|
| 1 | 340.34 | 86.15 | 12.65 | 2.29 |
| 2 | 37.38 | 17.50 | 5.53 | 1.56 |
| 3 | 19.93 | 11.16 | 4.33 | 1.38 |
| 4 | 14.96 | 9.04 | 3.88 | 1.31 |
| 5 | 12.66 | 8.06 | 3.62 | 1.26 |
| 6 | 11.36 | 7.44 | 3.47 | 1.23 |
| 7 | 10.57 | 7.05 | 3.37 | 1.21 |
| 8 | 9.96 | 6.79 | 3.30 | 1.20 |
| 9 | 9.58 | 6.57 | 3.24 | 1.19 |
| 10 | 9.25 | 6.41 | 3.20 | 1.18 |
| 11 | 9.04 | 6.30 | 3.16 | 1.17 |
| 12 | 8.83 | 6.20 | 3.13 | 1.16 |
| 13 | 8.68 | 6.13 | 3.11 | 1.16 |
| 14 | 8.54 | 6.05 | 3.09 | 1.15 |
| 15 | 8.41 | 5.98 | 3.07 | 1.15 |
| 16 | 8.31 | 5.93 | 3.05 | 1.15 |
| 17 | 8.23 | 5.88 | 3.04 | 1.15 |
| 18 | 8.18 | 5.84 | 3.03 | 1.15 |
| 19 | 8.10 | 5.82 | 3.02 | 1.14 |
| 20 | 8.05 | 5.78 | 3.01 | 1.14 |
| 21 | 7.98 | 5.75 | 3.00 | 1.14 |
| 22 | 7.96 | 5.71 | 3.00 | 1.14 |
| 23 | 7.91 | 5.69 | 2.99 | 1.14 |
| 24 | 7.83 | 5.68 | 2.98 | 1.14 |
| 25 | 7.82 | 5.66 | 2.98 | 1.13 |
| 30 | 7.65 | 5.60 | 2.95 | 1.13 |
| 40 | 7.52 | 5.49 | 2.92 | 1.12 |
| 60 | 7.33 | 5.40 | 2.89 | 1.12 |
| 120 | 7.14 | 5.31 | 2.86 | 1.11 |

Figura 3-1.: Valores críticos de la prueba del error cuadrático medio para restricciones lineales.

Método del espacio nulo

Consideremos nuevamente la función a minimizar $\|Ax - b\|$ sujeto a las restricciones $\|Cx - d\|$.

En este método a partir de vectores base, obtendremos un conjunto ortogonal para que, al ser base, se genere el espacio nulo de la matriz de restricciones C , de esta manera se garantiza una única solución para el problema. Es una de las soluciones ampliamente utilizadas, la ejecución de los algoritmos ha cambiado manteniendo la idea fundamental, al menos 3 versiones pueden distinguirse en [11]. Del cual daremos a conocer uno de ellos, el cual fue propuesto por [35].

Las restricciones siguen $Cx = d$, C se puede descomponer a través de una descomposición ortogonal, por ejemplo la factorización QR según

$$C = QR$$

Dado que el rango de $C \in \mathbb{R}^{m \times n}$ es $\rho(C) = p$ con $p \leq m$, se puede realizar una partición por bloques

$$C = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}$$

La matriz Q_1 contiene las primeras p columnas de C que son ortogonales, mientras que Q_2 posee los $n - p$ vectores que nos permiten formar una base ortogonal al espacio columna de la matriz C , y una vez generada esta base, se realiza la minimización para un valor x_2 de este espacio nulo, pues x acorde a C es particionada como $x = (x_1 \ x_2)$.

Dado que se busca minimizar la norma para el valor de x , por las propiedades ortogonales⁴ tenemos que el valor óptimo es equivalente al de Qx , es decir $\hat{x} = Q(x_1 \ x_2)$, este valor solo puede ser óptimo para el caso donde C sea de rango completo, así, $R_2 = 0$ y x_2 sería un vector de valores arbitrario. En caso de $R_2 \neq 0$ se debe minimizar la función para el valor de x_2 y la solución quedaría expresada en términos de

$$\hat{x} = \min \|x_1 + Q_2 x_2\| \quad (3-8)$$

Dado lo anterior, téngase en cuenta que las restricciones $Cx = d$, quedarían reexpresadas como $Cx_1 = d$, además en el valor de x_1 se va a involucrar el valor de Q_1 , a través de la factorización QR,

$$x_1 = C^{-1}d$$

Véase que el valor de $Ax - b$ con el resultado de 3-8 es:

$$Ax - b = Ax_1 + AQ_2 x_2 - b$$

⁴Véase la demostración en el Apéndice A, Definición 8.

Una vez obtenida una expresión en términos de matrices conocidas, el propósito se mantiene en la minimización del valor x_2

$$\min \|Ax - b\| = \min \|Ax_1 + AQ_2x_2 - b\|$$

Entonces la solución de x_2 sería

$$x_2 = (AQ_2)^{-1}(B - Ax_1)$$

Finalmente, la solución estaría dada por el reemplazo de x_2 en 3-8.

Por otra parte, y como es considerado en la literatura de éste método u otros, el problema de mínimos cuadrados restringidos considera la existencia de la solución y de su unicidad a través de la ecuación 3-9, ésta ecuación dice que el rango de la matriz de restricciones es de rango completo, la condición impone una solución que sea consistente, pero si además de ésta encontramos que el rango de la matriz ampliada A y B es completo, es decir n, la solución es única.

$$\rho(B) = k, \text{ y además } \rho \begin{bmatrix} A \\ B \end{bmatrix} = n \quad (3-9)$$

Que el rango de las matriz de diseño y de restricciones sea completo, tiene como consecuencia directa que no poseen columnas linealmente dependientes, entonces, el núcleo es el vector nulo,

$$N(A) \cap N(B) = 0$$

Para este caso, la solución a través de este método puede definirse como una matriz C por bloques, cumpliría con ser de dos bloques debido al rango obtenido para la matriz C -esto es Q_1 y Q_2 -, pues aunque es de rango completo, se puede considerar como consistente pero sobredeterminada ($m \geq n$), entonces el primer bloque cumple con ser de rango n , y el segundo bloque $n - p$ así

$$T = \begin{pmatrix} C \\ A \end{pmatrix} Q_C = \begin{pmatrix} R_C^T & 0 \\ AQ_1 & AQ_2 \end{pmatrix}$$

Para que la matriz T sea triangular, lo debe ser el componente $t_{22} = AQ_2$. Aunque AQ_2 es una forma para expresar una matriz triangular, Q es una matriz ortogonal para C , razón por la cual la matriz ortogonal y triangular para A son Q_A y R_A . Al descomponer AQ_2 se obtiene la expresión,

$$Q_A(AQ_2) = \begin{pmatrix} R_A \\ 0 \end{pmatrix}$$

Y ahora, R_A es una matriz triangular superior y no singular. Finalmente, la solución en este caso, descrita por el mismo modelo 3-8 tendría los siguientes componentes

$$c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = Q_A^T(b - Ax_1)$$

$$R_A x_2 = c_1$$

Históricamente se le atribuye buena estabilidad numérica al método del espacio nulo, y el conteo operacional es bajo comparado con la estructura y fundamentos que pueden tener otros métodos [8], pero puede ser un motivo engañoso, porque si la cantidad de datos en las matrices aumenta considerablemente, pierde la eficiencia en las características del resultado.

Método de eliminación directa

En éste método se busca derivar un problema sin restricciones que equivalga al problema original -con restricciones-, para ser solucionado.

La matriz B de rango k , permite descomponer la matriz de bloques basada en A y C conforme dicho rango así

$$\begin{pmatrix} C \\ A \end{pmatrix} = \begin{pmatrix} C_1 & C_2 \\ A_1 & A_2 \end{pmatrix}$$

Siendo $C_1 \in \mathbb{R}^{m_1 \times m_1}$, $A_1 \in \mathbb{R}^{m_2 \times m_1}$, $C_2 \in \mathbb{R}^{m_1 \times n-m_1}$ y $A_2 \in \mathbb{R}^{m_2 \times n-m_1}$. Con esta misma idea de particiones, para x se realiza el mismo proceso, $x^T = \begin{pmatrix} x_1 & x_2 \end{pmatrix}$ de tamaños, m_1 y $n - m_1$ respectivamente.

Las restricciones $Cx = d$, ahora son

$$Cx = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = d$$

$$C_1 x_1 = d - C_2 x_2$$

$$x_1 = C_1^{-1}(d - C_2 x_2) \quad (3-10)$$

En la matriz de restricciones que buscamos minimizar x_2 , $\min_{x_2} \|Ax - b\|$, la norma se reexpresa a través de sustitución para el valor de x_1 como

$$\|A_1(C_1^{-1}(d - C_2 x_2)) + A_2 x_2 - b\|$$

Agrupando términos para x_2

$$\|(A_2 - A_1 C_1^{-1} C_2) x_2 - b(A_1 C_1^{-1} d)\| \quad (3-11)$$

En la literatura se encuentra la notación $\hat{A} = (A_2 - A_1 C_1^{-1} C_2)$ y $\hat{b} = (A_1 C_1^{-1} d)$, para expresar de manera explícita que del problema original se ha derivado una minimización sin

restricciones expresada como $\|\hat{A}x - \hat{b}\|$, sin embargo, no se utilizara dicha notación en el desarrollo del método.

Ahora, véase que en la expresión 3-11 el valor de x_2 es

$$x_2 = (A_2 - A_1 C_1^{-1} C_2)^{-1} (A_1 C_1^{-1} d)$$

Finalmente este valor de x_2 se reemplaza en 3-10 para obtener el vector de coeficientes resultante.

Además de esta formulación, basada en [35], podemos encontrar modificaciones, por ejemplo en [8] o [19] se presenta una versión considerando una matriz de permutaciones Π .

Para este caso, se inicia realizando una descomposición a B de la forma

$$QB\Pi = \begin{pmatrix} R_1 & R_2 \\ 0 & 0 \end{pmatrix}$$

Con R_1 y R_2 con k filas y columnas k y $n-k$ respectivamente.

Esta partición desemboca en particiones del mismo orden, es decir, según el rango de B , que es k , para los vectores x , A y d . Cabe resaltar que para el vector d si se supone que la solución es única, -véase 3-1-, entonces en la partición $d^T = (d_1 \ d_2)$, se cumple que $d_2 = 0$.

Luego, la notación \bar{x} , \bar{A} y \bar{d} indica que se ha aplicado la operación de permutación $x\Pi$, $A\Pi$ y $d\Pi$.

Las restricciones $Cx = d$ tras la descomposición matricial pasan a ser

$$R_1 \bar{x}_1 + R_2 \bar{x}_2 = \bar{d}$$

de donde

$$\bar{x}_1 = R_1^{-1} (\bar{d} - R_2 \bar{x}_2)$$

Por otra parte, si en $\|\bar{A}\bar{x} - b\|$ se tiene $\bar{A}_1 \bar{x}_1 + \bar{A}_2 \bar{x}_2 - d$, y se realiza la sustitución del valor de \bar{x}_1 , tenemos

$$[\bar{C}_2 - \bar{C}_1 R_1^{-1} R_2] \bar{x}_2 - [\bar{b} - \bar{C}_1 R_1^{-1} \bar{d}] \quad (3-12)$$

En esta expresión se observa que se ha unificado en una sola expresión los datos y las restricciones, y es aquí donde se muestra la versión explícita del nuevo problema sin restricciones con la forma $\min_{x_2} \|\check{C}x_2 - \check{d}\|$, siendo \check{C} y \check{d} según los términos asociados en 3-12.

En pro de facilitar los cálculos y las propiedades de las estimaciones se realiza una descomposición QR para \check{C}

$$Q_C \check{C} = \begin{pmatrix} R_C \\ 0 \end{pmatrix}$$

Un vector g , es definido como

$$Q_C \check{d} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$

Finalmente la solución queda determinada por el sistema

$$\begin{pmatrix} R_1 & R_2 \\ 0 & Q_C \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} = \begin{pmatrix} \bar{d} \\ g_1 \end{pmatrix}$$

El valor del problema original queda determinado a través de $\Pi \bar{x} = \hat{x}$.

Análisis realizados para el método de eliminación directa muestra un comportamiento similar en términos de estabilidad numérica comparado con el método del espacio nulo.

Otros métodos pueden ser aplicados para la solución de mínimos cuadrados restringidos con restricciones de igualdad, sin embargo presentan más restricciones que los expuestos hasta ahora.

El método de ponderación, depende de un factor μ para obtener soluciones adecuadas, este valor que actúa como un ponderador, busca igualar la magnitud de la norma de $\|Cx - d\|$ a través de $\|(Ax)\mu - b\|$, se espera que al encontrar igualar o ser lo más parecido posible la magnitud de la norma, se encuentra una solución del problema a través de MCO. Pero diversos autores adoptan diversos valores de μ para hallar solución, algunos un valor de $\mu^{-1/3}$ como en [4], o según [53]

$$\mu = \sqrt{\frac{10^{-4}}{\gamma} \eta}$$

Siendo η el valor de la máquina de precisión relativa y γ el valor de la norma infinita para la matriz aumentada A y C .

Las disparidades encontradas en la literatura para la elección del valor de μ , tendría implicaciones a considerar, una de ellas es que la matriz puede volverse mal condicionada, según [4] el valor de μ repercute en la convergencia y la exactitud que puede tener la solución, así, *idem* afirma que no cubre un amplio espectro de problemas, o en otras palabras, éste método aunque puede ser fácilmente implementado al igual que el método del espacio nulo y de eliminación directa, sin embargo, no se garantiza que la solución hallada represente una buena estimación calculada.

En [5], se exponen métodos iterativos para la solución de MCR como gradiente conjugado o el método SOR, sin embargo, algunos inconvenientes se presentan cuando interviene la

factorización de Choleski en dichos métodos, pues se restringe su uso para matrices que no son simétricas y/o definidas no negativas.

A continuación adoptaremos un enfoque diferente, basado en la factorización QR de las matrices, indicando con mayor rigor para este método, la importancia de realizar descomposiciones ortogonales en los fundamentos de mínimos cuadrados restringidos.

4. Factorización QR generalizada

Los encantos de esta ciencia sublime, las matemáticas, sólo se le revelan a aquellos que tienen el valor de profundizar en ella.

Carl Friedrich Gauss.

Algunos misterios siempre escapan a la mente humana. Para convencernos de ello, sólo hay que echar un vistazo a las tablas de los números primos, y ver que no reina ni orden, ni reglas.

Evariste Galois.

Métodos comúnmente usados para el desarrollo de mínimos cuadrados, al igual que para mínimos cuadrados restringidos están basados en cálculos desarrollados a través de inversas matriciales o g-inversas, no obstante, aunque ofrecen soluciones directas, las propiedades de los resultados pueden verse afectadas, si se maneja una gran cantidad de datos, así mismo, se refleja en la ejecución de los algoritmos un mayor costo computacional, que a su vez, puede afectar directamente en la exactitud de los resultados, considerando por ejemplo acumulaciones en los errores de redondeo.

Como se evidencia en [23], la factorización QR desempeña un rol protagónico cuando las exigencias para la solución de un problema de mínimos cuadrados demandan el uso de operadores ortogonales, los cuales garantizan características deseables de un resultado, interviniendo en la estabilidad numérica, el costo computacional y la exactitud de los resultados. Además, se puede evidenciar en la literatura, que muchos de los métodos expuestos previamente, incluyen como subrutinas la implementación de la factorización QR, evidenciando mejorías en los algoritmos, pues el uso de esta descomposición hereda sus propiedades al método principal, consecuencia de esto es que sea un proceso usado para resolver problemas de mínimos cuadrados y en particular también para mínimos cuadrados restringidos con restricciones de igualdad.

El uso de descomposiciones matriciales en distintos algoritmos ha sido ampliamente utilizado en los campos de aplicación, con el propósito de facilitar el cálculo e implementación computacional.

En 1986 Hammarling introdujo en el artículo “The numerical solution of the general Gauss-Markov linear model”, el concepto de factorización QR generalizada, conocida como GQR por sus siglas en inglés, es una factorización de tipo QR, que permite transformar de manera simultánea dos matrices A y B a una forma triangular. Estudios posteriores fueron realizados en 1990 por Paige al publicar el artículo “Some aspects of generalized QR factorization” y por Anderson, Bai y Dongarra en 1992 a través de “Generalized QR factorization and its applications”. Un tratamiento para realizar de manera simultánea la factorización de una mayor cantidad de matrices fue propuesto por [14].

Factorización QR generalizada

Dadas dos matrices $A \in \mathbb{R}^{m \times n}$ y $B \in \mathbb{R}^{p \times n}$, la factorización GQR permite descomponer simultáneamente las matrices a una forma triangular.

Proposición 4.0.1. *Sea $A \in \mathbb{R}^{m \times n}$ y $B \in \mathbb{R}^{m \times p}$, entonces*

1. **Factorización GQR:** *Si $m \geq n$, existen dos matrices ortogonales Q y V tales que cumplen $Q^T A = R$ y $Q^T B V = S$, siendo R, S matrices trapezoidales superiores*

1.

¹Una matriz trapezoidal aplica para matrices que no son cuadradas $A \in \mathbb{R}^{m \times n}$; trapezoidal superior es

2. **Factorización GRQ:** Si $m < n$, existen dos matrices ortogonales $Q_{m \times m}$ y $U_{n \times n}$ tales que $Q^T A U = R$ y $Q^T B = S$.

Demostración. La demostración es constructiva, se prueba el caso 1, de manera análoga se puede probar 2.

La matriz A es definida a través de la factorización QR, de tal manera, la matriz triangular correspondiente R se puede expresar como

$$R = Q^T A$$

Al ser $A \in \mathbb{R}^{m \times n}$ una matriz sobredeterminada, se realiza una reexpresión a través de particiones matriciales con la siguiente forma

$$R = Q^T A = \begin{pmatrix} R_1 \\ 0 \end{pmatrix} \begin{matrix} n \times n \\ m - n \times n \end{matrix}$$

De la misma manera, una matriz $B \in \mathbb{R}^{m \times p}$ puede ser reducida a una forma triangular y particionada a través de

$$[Q^T B] V = S$$

Ésta expresión se obtiene al premultiplicar el valor de Q^T a B para realizarle a este término conjunto la factorización RQ, -véase sus fundamentos en el Apéndice A y el código para su ejecución en el lenguaje R en el Apéndice B-, entonces la matriz triangular S toma las siguientes formas

$$S = \begin{cases} \begin{bmatrix} 0 & S_1 \end{bmatrix} & \text{si } m \leq p \\ \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} & \text{si } m > p \end{cases}$$

Siendo en el primer caso $S_1 \in \mathbb{R}^{m \times m}$ y el vector $\mathbf{0} \in \mathbb{R}^{m \times p - m}$. En cambio para el segundo caso de S , se tiene que $S_1 \in \mathbb{R}^{m-p \times p}$ y $S_2 \in \mathbb{R}^{p \times p}$.

□

Ejemplo 1. Sean A y B dos matrices definidas como

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -3 & 2 & 1 \\ 2 & 0 & -1 \\ 3 & -1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ -3 & 2 & -2 & 1 & 2 \\ 2 & 3 & 4 & -2 & -1 \\ 1 & 3 & -2 & 2 & 1 \end{bmatrix}.$$

aquella en la cual los elementos a_{ii} y $a_{ij} \neq 0$ para $i > j$, en caso contrario se dice matriz trapezoidal inferior. Ésta definición no discrimina el caso $m > n$ ó $m < n$.

La factorización GQR para A y B esta dada a través de las siguientes matrices

$$Q = \begin{bmatrix} -0,2085 & -0,8792 & 0,1562 & -0,3989 \\ 0,6255 & -0,4147 & 0,1465 & 0,6444 \\ -0,4170 & -0,2322 & -0,7665 & 0,4296 \\ -0,6255 & 0,0332 & 0,6054 & 0,4910 \end{bmatrix}$$

$$R = \begin{bmatrix} -4,7958 & 1,4596 & -0,8341 \\ 0 & -2,6210 & -2,7537 \\ 0 & 0 & 2,5926 \\ 0 & 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} 0,3375 & -0,0791 & -0,2690 & -0,6363 & 0,6345 \\ -0,8926 & -0,2044 & 0,1635 & -0,2770 & 0,2407 \\ -0,0534 & -0,5118 & -0,5793 & -0,2833 & -0,5651 \\ -0,1584 & -0,1280 & -0,6087 & 0,6280 & 0,4401 \\ -0,2478 & 0,8208 & -0,4413 & -0,2091 & -0,1626 \end{bmatrix}$$

$$S = \begin{bmatrix} 0 & -3,431 & 2,8692 & -1,8585 & 0,1388 \\ 0 & 0 & 7,0240 & 2,1937 & 0,1571 \\ 0 & 0 & 0 & -5,9566 & 1,0776 \\ 0 & 0 & 0 & 0 & 3,9630 \end{bmatrix}.$$

Otra factorización GQR esta dada a través de Q y R como se mostraron anteriormente, y las matrices V y S como

$$V = \begin{bmatrix} -0,8926 & -0,0534 & -0,1585 & -0,2478 & 0,3375 \\ -0,2044 & -0,5118 & -0,1280 & 0,8208 & -0,0791 \\ 0,1635 & -0,5794 & -0,6087 & -0,4414 & -0,2689 \\ -0,2771 & -0,2833 & 0,6280 & -0,2091 & -0,6363 \\ 0,2407 & -0,5651 & 0,4401 & -0,1626 & 0,6345 \end{bmatrix}$$

$$S = \begin{bmatrix} 3,4311 & 2,8692 & 1,8585 & 0,1388 & 0 \\ 0 & 7,0240 & -2,1937 & 0,1571 & 0 \\ 0 & 0 & 5,9566 & 1,0776 & 0 \\ 0 & 0 & 0 & 3,9630 & 0 \end{bmatrix}.$$

Tras lo propuesto en 4.0.1, la formación de una matriz triangular para las matrices permite no trabajar directamente con sus inversas A^{-1} ó B^{-1} en caso de ser requerido. El trabajo de resolver un sistema en donde se encuentre una matriz triangular presente, se da por medio de sustitución hacia atrás y, en caso de requerirse una inversa de una matriz triangular, el número de operaciones requeridas disminuye respecto de la matriz original, pues se obtendrá otra matriz triangular con la misma forma inicial (superior o inferior), cuyos elementos inversos en la diagonal principal serán los recíprocos de los valores originales; de allí las

ventajas de la ejecución del método y la estabilidad numérica de los resultados, disminuyendo respecto de otros métodos el error por redondeo [3].

La factorización GQR al estar basada en descomposiciones ortogonales, hereda las propiedades directas de la factorización QR en cuanto a la estabilidad numérica de las observaciones, que a su vez, provienen del método empleado para la ortogonalización, como el proceso de Gram-Schmidt, el proceso de Gram-Schmidt modificado, reflexiones de Householder o rotaciones de Givens, véase Apéndice A.

La factorización GQR además de permitir la solución a través de una factorización directa, la descomposición realizada permite que en el enfoque computacional no sea un proceso costoso ni dispensioso en la cantidad de operaciones, y por último, el factor triangular brinda información sobre el condicionamiento de la matriz, estará mal condicionada si son cercanos a cero los valores de la diagonal principal, este condicionamiento recae en la sensibilidad de los resultados obtenidos, pues, un pequeño cambio en los valores de entrada puede producir un gran cambio en los resultados de salida.

Por las razones expuestas anteriormente, el campo de aplicación de este método ha sido útil en problemas de mínimos cuadrados restringidos, con restricciones de igualdad, o incluso en restricciones de tipo no lineal [12], además de ser usado para mínimos cuadrados ponderados, mínimos cuadrados generalizados y en mínimos cuadrados sobre cuaterniones cuya fundamentación se expondrá posteriormente para éste último caso.

De esta manera se expone la solución del problema de mínimos cuadrados restringidos con restricciones de igualdad a través de la factorización QR generalizada, y finalmente, la aplicación de dicha factorización para el caso de problemas de mínimos cuadrados restringidos con restricciones de igualdad en cuaterniones.

4.1. Factorización GQR para mínimos cuadrados restringidos

Al tener como característica principal la descomposición simultánea de matrices, es un método que se adapta a las características del problema de mínimos cuadrados restringidos, sin necesidad y a diferencia de otros métodos, de derivar un problema sin restricciones u otros procesos que no conservan la estructura esencial del modelo. De tal manera, la solución de un problema con restricciones es resuelto con mayor eficiencia en términos de computación y de brindar información sobre el condicionamiento de las matrices con base en la factorización QR generalizada.

4.1.1. Problemas de mínimos cuadrados restringidos con restricciones de igualdad

Como se evidenció en §3, la inclusión de información a priori puede derivar como caso particular el problema de mínimos cuadrados restringidos a través de igualdades, conocido y abreviado por sus siglas en inglés como el problema LSE, en este, se busca un vector $x \in \mathbb{R}^n$ que satisfaga la ecuación (4-1).

$$\min_x \|Ax - b\|, \text{ sujeto a } Cx = d, \quad (4-1)$$

donde $A \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{p \times n}$ y con los vectores $b \in \mathbb{R}^m$ y $d \in \mathbb{R}^p$. Sea el caso más frecuente cuando $m \geq n$. El problema LSE puede ser reestructurado de la forma 4-2.

$$\min_{Cx=d} \|Ax - b\|. \quad (4-2)$$

El problema de mínimos cuadrados restringidos puede ser resuelto a través de la factorización QR cf. [58], en este trabajo utilizamos un método más eficiente basado en la descomposición presentada en la proposición 4.0.1. En ese orden de ideas, como se definió, en la factorización GRQ se tienen dos matrices ortogonales Q y U tales que $A^T = QRU$ y $C^T = QS$, con $S \in \mathbb{R}^{p \times n}$ y $R \in \mathbb{R}^{m \times n}$ son matrices triangulares superiores, luego

$$\min_{Cx=d} \|Ax - b\| = \min_{Cx=d} \|[QRU]^T x - b\|.$$

A través de la sustitución $y = Q^T x$, esta ecuación puede ser reescrita como

$$\min_{[QS]^T [Qy]=d} \|[U^T R^T Q^T] x - b\|.$$

Debido a que U es una matriz ortogonal (unitaria para el caso complejo), la norma euclideana no se ve afectada por transformaciones ortogonales, véase Apéndice A, entonces tenemos

$$\min_{[S^T Q^T][Qy]=d} \|U(U^T R^T Q^T x - b)\|.$$

$$\min_{S^T y=d} \|R^T Q^T x - Ub\|.$$

$$\min_{S^T y=d} \|R^T y - Ub\| \quad (4-3)$$

Para el caso donde la matriz A es sobredeterminada, es decir, posee más ecuaciones que variables, $m \geq n$, la matriz S es $S = \begin{bmatrix} S_1 \\ 0 \end{bmatrix}$ y, $S^T y = [S_1^T \ 0] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, entonces

$$S_1^T y_1 = d. \quad (4-4)$$

Asumiendo que $P = Ub$ y que la matriz R tenga forma triangular así

$$R = \begin{array}{ccc|c} R_{11} & R_{12} & 0 & p \\ 0 & R_{22} & 0 & n-p \\ \hline p & n-p & m-n & \end{array}$$

Aunque R también puede ser considerada de la forma

$$R = \begin{array}{ccc|c} 0 & R_{11} & R_{12} & p \\ 0 & 0 & R_{22} & n-p \\ \hline m-n & p & n-p & \end{array}$$

Ahora, de 4-3 se tiene que,

$$\min_{S^T y=d} \|R^T y - P\| = \min_{S^T y=d} \left\| \begin{bmatrix} 0 & 0 \\ R_{11}^T & 0 \\ R_{12}^T & R_{22}^T \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} - \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \right\|.$$

Al calcular el valor de y_1 en 4-4, el problema puede ser truncado -debido a que y_1 se calcula en 4-4- a

$$\begin{aligned} \min_{y_2} \|R_{12}^T y_1 + R_{22}^T y_2 - p_2\| \\ \min_{y_2} \|R_{22}^T y_2 - (p_2 - R_{12}^T y_1)\|. \end{aligned}$$

Como R_{22} es una matriz con valores $a_{ii} \neq 0$, es por tanto una matriz no singular, se justifica que

$$y_2 = R_{22}^{-T} (p_2 - R_{12}^T y_1) \quad (4-5)$$

Dado el valor de y_1 en 4-4 y de y_2 en 4-5, el valor de x queda determinado por la solución

$$x = Qy = Q_1 y_1 + Q_2 y_2 \quad (4-6)$$

Se observa de manera explícita, que en los cálculos requeridos, la inversa necesaria R_{22}^{-T} ó $[R_{22}^T]^{-1}$, es una matriz triangular superior, y dada la facilidad lograda comparada con otros métodos, véase la solución estimada a través de los multiplicadores de Lagrange §3-2 se obtienen características de un resultado con mayor estabilidad numérica, exactitud.

Ejemplo 2. Resolver el problema

$$\min_{x_1+x_2=1} \left\| \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 7 \\ 1 \\ 3 \end{bmatrix} \right\|.$$

Sea $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$, $b = \begin{bmatrix} 7 \\ 1 \\ 3 \end{bmatrix}$, $C = [1 \ 1]$, $d = [1]$, por la factorización GQR obtenemos

$$Q = \begin{bmatrix} -0,7071 & -0,7071 \\ -0,7071 & 0,7071 \end{bmatrix},$$

$$R = \begin{bmatrix} 4 & 8,5732 & 0 \\ 0 & -1,2247 & 0 \end{bmatrix}$$

$$U = \begin{bmatrix} 0,7071 & 0 & -0,7071 \\ -0,5773 & -0,5773 & -0,5773 \\ 0,04082 & -0,8165 & 0,4082 \end{bmatrix},$$

$$S = \begin{bmatrix} -1,4142 \\ 0 \end{bmatrix}.$$

Por lo tanto, la solución dada por 4-6 es $x_1 = 0,33333$ y $x_2 = 0,66667$, que claramente demuestra el cumplimiento de la restricción de clausura. El problema desarrollado anteriormente fue presentado en [59], y fue resuelto a través de distintos métodos.

4.1.2. Aplicación de la factorización GQR en cuaterniones

El estudio del problema de mínimos cuadrados en cuaterniones, conocido como problema QLS, y los diferentes caminos para resolver este problema ha ganado interés [31], [32]. En este trabajo se presenta una manera de resolver el problema de mínimos cuadrados restringidos con restricciones de igualdad en cuaterniones a través de la factorización QR generalizada. Previamente a la fundamentación se expondrán conceptos básicos sobre los números complejos, para luego introducir el concepto de cuaternión.

Gracias a la labor de Thomas Graves, en introducirle fundamentos de números complejos a Sir Edwart Rowan Hamilton en 1829, pudo realizar extensos estudios en esta rama, la visión que adquirió, era contemplar la posibilidad de engranar el concepto matemático del número complejo con una interpretación geométrica, para luego permitirse pensar sobre la generalización de los números complejos a más dimensiones.

Fue en 1843, tras estudios y publicaciones realizadas en la década, como se puede ver con mas detalle en [39], y tras años de trabajos y análisis fallidos por parte de Hamilton al buscar la generalización de los números, que yendo a una reunión de la Royal Irish Academy, y al pasar por el puente Broome tuvo la epifanía sobre la solución a su problema, y con angustia de que pudiese suceder algún infortunio, con su navaja escribió en una roca la solución, ésta desapareció, y posteriormente se erigió una placa con la ecuación. Hamilton escribió en una carta a su hijo Archibald [10]:

“Mañana será el decimoquinto cumpleaños de los cuaterniones. Surgieron a la vida, o a la luz, ya crecidos, el 16 de octubre de 1843, cuando me encontraba caminando con la Sra. Hamilton hacia Dublín, y llegamos al Puente de Broughman. Es decir, entonces y ahí, cerré el circuito galvánico del pensamiento y las chispas que cayeron fueron las ecuaciones fundamentales entre i, j, k ; exactamente como las he usado desde entonces. Saqué, en ese momento, una libreta de bolsillo, que todavía existe, e hice una anotación, sobre la cual, en ese mismo preciso momento, sentí que posiblemente sería valioso el extender mi labor por al menos los diez (o podían ser quince) años por venir. Es justo decir que esto sucedía porque sentí, en ese momento, que un problema había sido resuelto, un deseo intelectual aliviado, deseo que me había perseguido por lo menos los quince años anteriores. No pude resistir el impulso de coger mi navaja y grabar en una piedra del Puente Brougham la fórmula fundamental con los símbolos i, j, k :

$$i^2 = j^2 = k^2 = ijk = -1$$

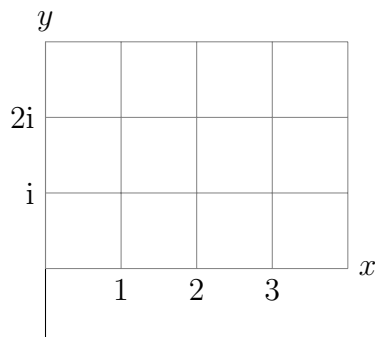
que contenían la solución del Problema, que desde entonces sobrevive como inscripción.”
p.637

Así, una nueva rama surgió de las manos de Hamilton, él consideraba que el espacio tiempo podía ser representado a través de estos números hipercomplejos debido a que viven en \mathbb{R}^4 , 3 dimensiones representan el espacio y la restante el tiempo. Hamilton se dedicó a trabajar enteramente sobre sus fundamentos y aplicaciones, fue así como en 1853 publicó “Lecciones sobre cuaterniones” y en 1866, como obra póstuma tras un año fue publicado “Elementos de los cuaterniones”, siendo sus obras más representativas acerca de estos números, de igual manera, realizó diversos aportes en campos de la física como óptica, dinámica y astronomía.

La cadena de generalización continuó, pues de los números reales están los complejos, luego los cuaterniones y se desarrolló mayor complejidad tras su legado, los octoniones, representando números de un algebra de dimensión 8, que gracias a los cuaterniones formaron nuevas ramas y estudios sobre el no cumplimiento de las propiedades básicas de ciertos operadores; si en el caso de los cuaterniones no se cumple la conmutatividad, para el caso de los octoniones no se cumple las propiedades conmutativa y asociativa. Mayores detalles de la historia del surgimiento de estos números pueden ser encontrados en [39], [10].

Ahora desde un enfoque conceptual, analicemos los números complejos y como se generalizan a cuaterniones.

Los números complejos, fueron definidos en palabras de Hamilton como un par de números reales $x = (a, b)$ que cumplen ciertas operaciones algebraicas, por otra parte, Euler utiliza la notación $a + bi$ donde $i^2 = -1$, siendo ambas equivalentes, a corresponde a la parte real del número complejo, y b corresponde a la parte imaginaria, éstas se denotan por $Re(x) = a$ y $Im(x) = b$. Sea la propiedad modulativa para la adición el valor $(0, 0)$ y para la multiplicación $(1, 0)$.

Figura 4-1.: Representación del plano de Argand para números complejos.

Intervienen estos números con dos operaciones conformando un cuerpo, siendo la adición y multiplicación respectivamente

$$(a, b) + (c, d) = (a + c, b + d)$$

$$(a, b) * (c, d) = (ac - bd, ad + bc)$$

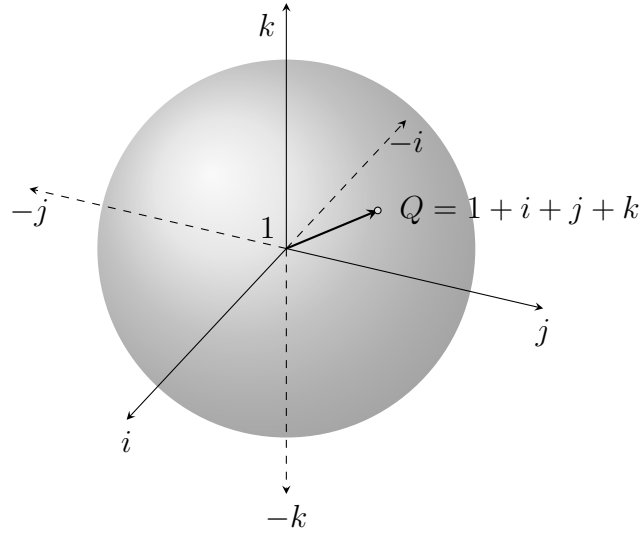
A través del plano de Argand -figura 4-1- se puede realizar una representación gráfica de los números complejos. En dicho plano, los números reales se ven representados por un eje, mientras que los números complejos generan el plano siendo i la unidad imaginaria.

Gauss indicó que las dificultades del análisis gráfico para los números complejos realizados por Caspar Wessel y por Sir. William Rowan Hamilton consistían en que utilizaban en su representación gráfica segmentos de recta determinado por las componentes de a para las abscisas y de b para las ordenadas; entonces él matemático alemán indicó que un punto de vista idóneo era considerar los números complejos solamente como puntos en el plano, además el consideró y fue aceptado en el mundo académico no llamarlos números imaginarios sino números complejos.

Las operaciones de adición y sustracción permanecen análogas al tratamiento gráfico que se realiza para los vectores en \mathbb{R}^2 , por otra parte, la multiplicación de dos números complejos queda representado por una rotación además de la dilatación o contracción según sea el caso de los vectores, y basados en los planteamientos de Euler, una manera alterna de ver este efecto es en términos de la norma y la ángulo del vector.

Se puede encontrar como representación matricial de un número complejo la siguiente expresión para $z = a + bi = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

Un cuaternión generaliza a más dimensiones un número complejo, esto es posible al añadir tres unidades imaginarias i , j y k .

Figura 4-2.: Representación gráfica para un cuaternión Q en \mathbb{R}^3 .

| | | | | |
|---|---|----|----|----|
| | 1 | i | j | k |
| 1 | 1 | i | j | k |
| i | i | -1 | k | j |
| j | j | -k | -1 | i |
| k | k | j | -i | -1 |

Tabla 4-1.: Multiplicación de cuaterniones

Definición 5 (Cuaternión). Sean a, b, c y $d \in \mathbb{R}^1$, además considere tres unidades imaginarias i, j y k . Un cuaternión Q es un número representado por la forma

$$Q = a + bi + cj + dk \quad (4-7)$$

Véase la representación gráfica de un cuaternión en la figura 4-2; además considere que la relación de las unidades imaginarias se da a partir de satisfacer las siguientes propiedades:

$$j = k = -ji \quad (4-8)$$

$$i = j = -ik \quad (4-9)$$

$$k = i = -kj \quad (4-10)$$

$$i^2 = j^2 = k^2 = ijk = -1 \quad (4-11)$$

La tabla de multiplicación para los cuaterniones permite aplicar con sencillez el resultado del producto entre las unidades imaginarias como se evidencia en la tabla 4-1.

Veamos las propiedades de un cuaternión escuetamente. Se cumple la propiedad de adición de cuaterniones para p y q así

$$\begin{aligned} p + q &= (p_0 + p_1i + p_2j + p_3k) + (q_0 + q_1i + q_2j + q_3k) \\ &= (p_0 + q_0) + (p_1 + q_1)i + (p_2 + q_2)j + (p_3 + q_3)k \end{aligned}$$

La multiplicación está dada al aplicar la propiedad distributiva entre todos los términos y aplicar las condiciones de Hamilton presentadas en las Ecuaciones (4-8)-(4-11)

$$\begin{aligned} p * q &= (p_0 + p_1i + p_2j + p_3k) * (q_0 + q_1i + q_2j + q_3k) \\ &= (p_0q_0 - p_1q_1 - p_2q_2 - p_3q_3) + (p_0q_1 + p_1q_0 + p_2q_3 - p_3q_2)i \\ &\quad + (p_0q_2 - p_1q_3 + p_2q_0 + p_3q_1)j + (p_0q_3 + p_1q_2 + p_2q_1 + p_3q_0)k \end{aligned}$$

La multiplicación se puede expresar a través de un vector 4×1 según [27]

$$p * q = (p_0, p_1, p_2, p_3) * (q_0, q_1, q_2, q_3) = \begin{bmatrix} p_0q_0 - p_1q_1 - p_2q_2 - p_3q_3 \\ p_1q_0 + p_0q_1 + p_2q_3 - p_3q_2 \\ p_2q_0 + p_0q_2 + p_3q_1 - p_1q_3 \\ p_3q_0 + p_0q_3 + p_1q_2 - p_2q_1 \end{bmatrix}$$

El primer componente del vector representa el producto interno entre dos cuaterniones

$$p \cdot q = p_0q_0 + p_1q_1 + p_2q_2 + p_3q_3$$

Siguiendo con otras propiedades que cumplen los cuaterniones se tiene la existencia de un módulo aditivo, esto es $p + (-p) = 0$, y un modulo multiplicativo $pp^{-1} = p^{-1}p = 1$.

En cuanto a la asociatividad de la adición y multiplicación para los cuaterniones se cumple que

$$\begin{aligned} p + (q + r) &= (p + q) + r \\ (pq)r &= p(qr) \end{aligned}$$

La conmutatividad se satisface para la adición, $p + q = q + p$, sin embargo no se cumple esta ley en la multiplicación $pq \neq qp$, véase por ejemplo la tabla de multiplicación de los cuaterniones, así, cumplen propiedades que lo permiten conformar un cuerpo no conmutativo.

Como propiedades adicionales se define el conjugado de un cuaternión por

$$\bar{p} = p_0 - p_1i - p_2j - p_3k$$

El inverso de p es

$$p^{-1} = \frac{\bar{p}}{|p|^2}$$

La norma o valor absoluto de p es

$$|p| = \sqrt{p_0^2 + p_1^2 + p_3^2 + p_4^2}$$

Una vez definidas las propiedades básicas, y conociendo su representación gráfica, puede ser definido un cuaternión a través de matrices, siguiendo a [31],[32], tenemos que un cuaternión $q = a + bi + cj + dk$, está representado a través de una matriz con valores reales de orden 2

$$\begin{bmatrix} a + bi & c + di \\ -c + di & a - bi \end{bmatrix}$$

O, aunque más extensa, pero con mayor facilidad para su manejo, en una matriz de orden 4

$$\begin{bmatrix} a & -b & d & -c \\ b & a & -c & -d \\ -d & c & a & -b \\ c & d & b & a \end{bmatrix}. \quad (4-12)$$

Por ejemplo, $q = -1 + j$ está representado por una matriz $\begin{bmatrix} -1 & -0 & 0 & -1 \\ 0 & -1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}$.

Ejemplo 3. Resolver el problema

$$\min_x \left\| \begin{bmatrix} -1 + j & -k \\ i & 1 + j \end{bmatrix} x - \begin{bmatrix} -1 \\ i \end{bmatrix} \right\|$$

En [32] se presenta un método para el problema de mínimos cuadrados sobre cuaterniones, en este trabajo a través de la factorización GQR se resuelve de manera equivalente $\min_x \|Ax - b\|$.

Dada la representación matricial expresada en 4-12, es facil comprobar que

$$A = \begin{bmatrix} -1 & -0 & 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

La solución del problema de mínimos cuadrados es

$$X = \begin{bmatrix} \frac{2}{3} & 0 & 0 & -\frac{1}{3} \\ 0 & \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 0 & -\frac{1}{3} & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{3} \\ 0 & 0 & \frac{1}{3} & 0 \end{bmatrix}.$$

O, realizando el proceso inverso en 4-12, se obtiene la solución $\begin{bmatrix} \frac{2}{3} + \frac{1}{3}j \\ \frac{1}{3}i \end{bmatrix}$.

Del mismo modo que se presenta la solución al problema de mínimos cuadrados sobre cuaterniones, conocido por sus siglas en inglés como QLS, también se puede aplicar el mismo proceso, para el problema de mínimos cuadrados restringidos con restricciones de igualdad sobre cuaterniones.

Los cuaterniones tienen múltiples aplicaciones en diferentes campos, es muy útil en rotaciones espaciales para la aplicación en robótica, al ser una alternativa frente al uso de ángulos de Euler, otras aplicaciones son expuestas en [27]. En [60] es usado para reconocimiento facial, pero también tiene otras aplicaciones en geometría, mecánica electromagnética, relatividad, modelado en naves espaciales.

5. Aplicaciones

Es la naturaleza lo que proporciona los primeros elementos de las cosas y la enseñanza lo que completa su desarrollo.

Retórica a Herenio, Lib. III. Autor Desconocido.

5.0.1. Simulación

Se presenta mediante la simulación a través del software R para la comparación de cuatro métodos a través de dos criterios en tres etapas.

Los cuatro métodos son: basado en multiplicadores de Lagrange §3.1, método del espacio nulo y eliminación directa §3.1.3 y el método basado en la factorización QR generalizada, expuesto en el capítulo 3. Los dos criterios para la comparación son la norma euclidiana del residual $Ax - b$ y la norma euclidiana del residual $Cx - d$, -para la última etapa se evalúa a través de la norma Frobenius-. Las tres etapas bajo los cuales se ponen a prueba los algoritmos son: para la primera etapa en aplicaciones de MCR con restricciones de igualdad con matrices sobredeterminadas, pues comúnmente este tipo de matrices son encontradas en la práctica, la segunda etapa es añadiendo colinealidad a la matriz de restricción y/o de diseño; y la última etapa es para la aplicación de MCR con restricciones de igualdad en cuaterniones.

Etapa 1: Matrices sobredeterminadas

Dado que en la práctica se suelen encontrar diversas aplicaciones en estadística multivariada, diseño de experimentos, bioestadística, actuarial, aplicaciones físicas, etc. donde la matriz de diseño A tiene más ecuaciones que variables. Se ponen a prueba los 4 métodos desarrollados en este trabajo variando el tamaño de A según los siguientes tres casos.

Caso 1: $m = 10$

Siendo $A \in R^{10 \times 5}$ una matriz sobredeterminada y $C \in R^{4 \times 5}$ una matriz subdeterminada, la simulación de 500 matrices para el primer caso se observa en la tabla **5-1**.

En la tabla se evalúan los dos criterios para los cuatro métodos desarrollados considerando 10 ecuaciones para la matriz de diseño. Los residuales en la tabla expresan la cantidad de problemas en los que la norma euclidiana fue inferior que los demás métodos.

| Comparación para $m = 10$ | | | | |
|---------------------------|----------|------------------|-----------|-----|
| Criterio | Lagrange | Dir. Elimination | Nullspace | GQR |
| $Ax - b$ | 240 | 240 | 240 | 260 |
| $Cx - d$ | 226 | 226 | 226 | 274 |

Tabla 5-1.: Comparación de 4 métodos de solución

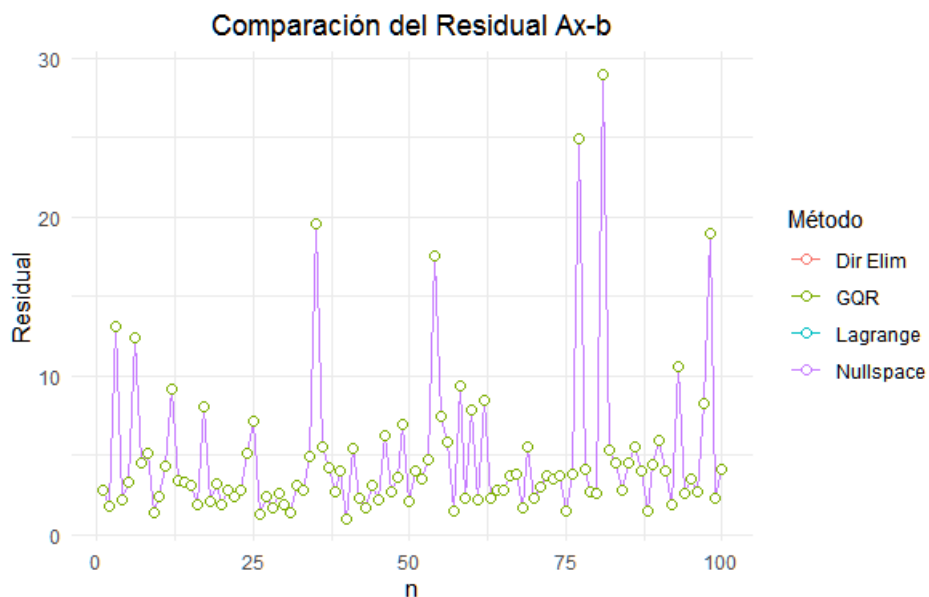


Figura 5-1.: Residual $Ax - b$ para 4 métodos de solución

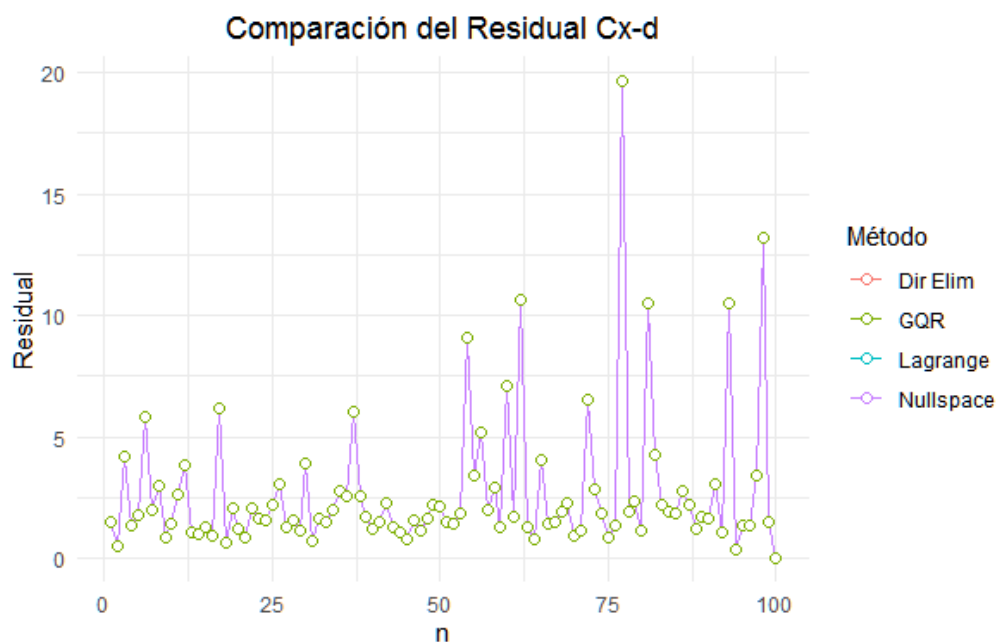


Figura 5-2.: Residual $Cx - d$ para 4 métodos de solución

El método basado en multiplicadores de Lagrange, Eliminación Directa y el método del Espacio Nulo, presentan resultados idénticos para los residuales $Ax - b$ y $Cx - d$, por lo tanto,

no hay criterio de comparación entre estos tres métodos. Por consiguiente, la comparación a nivel de resultados en los residuales es de estos tres métodos, denotados desde ahora como el grupo 1 frente al rendimiento del método GQR.

En la gráfica **5-1** se observan 100 problemas distintos, sin diferencias visualmente significativas para el residual $Ax - b$, además la tabla **5-1** muestra que entre el grupo 1 y el método GQR, la norma del residual para GQR fue inferior en los 500 problemas distintos para el 52% de los casos, indicando que para $m = 10$, no hay diferencias significativas para los cuatro métodos.

De manera análoga para el residual $Cx - d$, como se muestra en la gráfica **5-2**, no se aprecian diferencias ostensibles entre los métodos, del total de problemas, un 54.8% muestra un inferior valor para el residual de la restricción en el método GQR.

Caso 2: $m = 50$

Considere las matrices $A \in \mathbb{R}^{50 \times 25}$ y $C \in \mathbb{R}^{20 \times 25}$, la tabla **5-2** resume los indicadores para comparar los cuatro métodos a través de la simulación de 500 problemas.

| Comparación para $m = 50$ | | | | |
|---------------------------|----------|------------------|-----------|-----|
| Criterio | Lagrange | Dir. Elimination | Nullspace | GQR |
| $Ax - b$ | 255 | 255 | 255 | 245 |
| $Cx - d$ | 235 | 235 | 235 | 265 |

Tabla 5-2.: Comparación de 4 métodos de solución

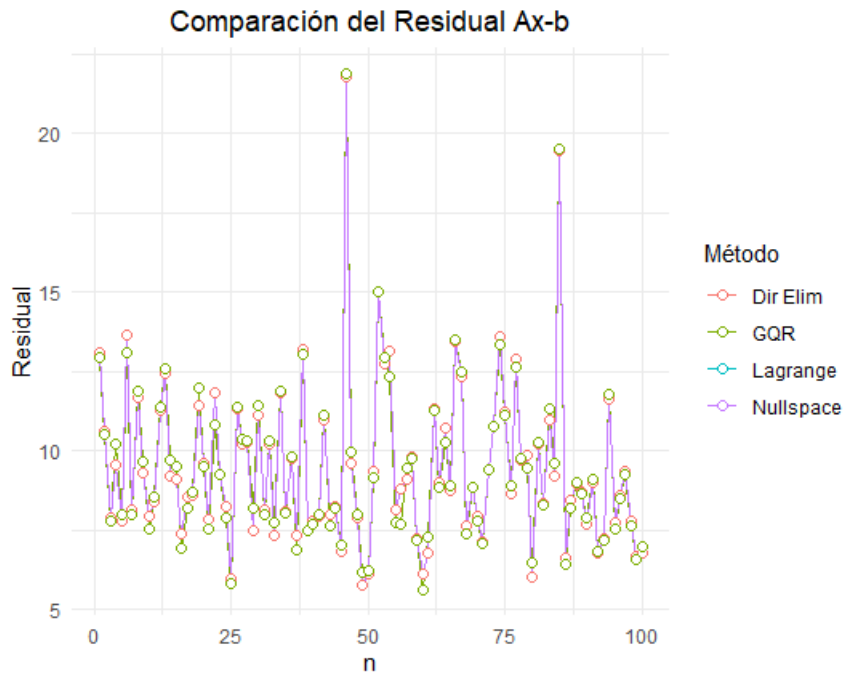


Figura 5-3.: Residual $Ax - b$ para 4 métodos de solución

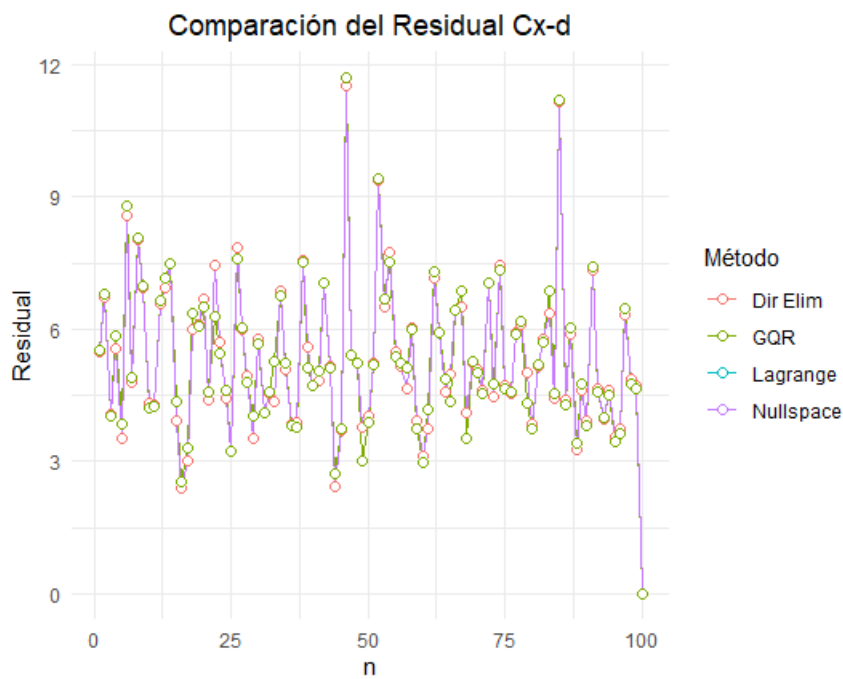


Figura 5-4.: Residual $Cx - d$ para 4 métodos de solución

En la gráfica **5-3** no se observan diferencias significativas para el residual $Ax - b$, y la tabla **5-2** muestra que entre el grupo 1 y el método GQR, la norma euclidiana del residual para GQR fue inferior entre los 500 problemas distintos para el 49 % de estos, evidencia de que a pesar del aumento de ecuaciones, se mantienen las similitudes entre los cuatro métodos. De manera análoga para el residual $Cx - d$, como se muestra en la gráfica **5-4**, no se aprecian diferencias ostensibles entre los métodos; del total de problemas un 53 % muestra un valor inferior para el residual de la restricción a favor del método GQR.

Caso 3: $m = 100$

Finalmente para la primera etapa, a través de las matrices $A \in \mathbb{R}^{100 \times 50}$ y $C \in \mathbb{R}^{25 \times 50}$, en la tabla **5-3** se observan los resultados obtenidos de los dos criterios de comparación para la su aplicación en 500 problemas.

| Comparación para $m = 100$ | | | | |
|----------------------------|----------|------------------|-----------|-----|
| Criterio | Lagrange | Dir. Elimination | Nullspace | GQR |
| $Ax - b$ | 231 | 231 | 231 | 269 |
| $Cx - d$ | 224 | 224 | 224 | 276 |

Tabla 5-3.: Comparación de 4 métodos de solución

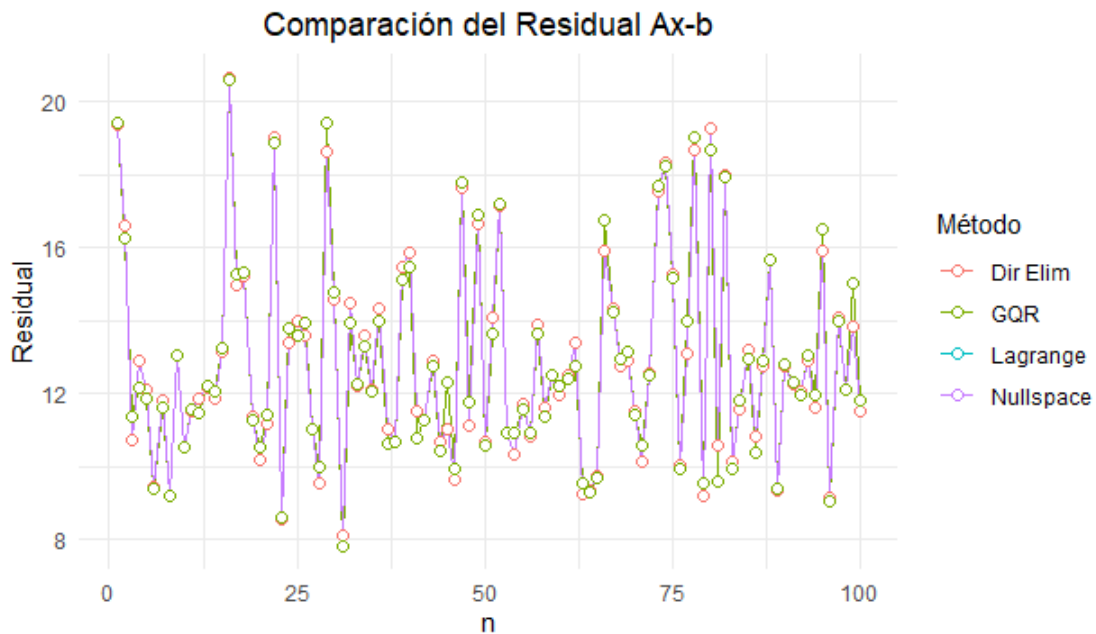


Figura 5-5.: Residual $Ax - b$ para 4 métodos de solución

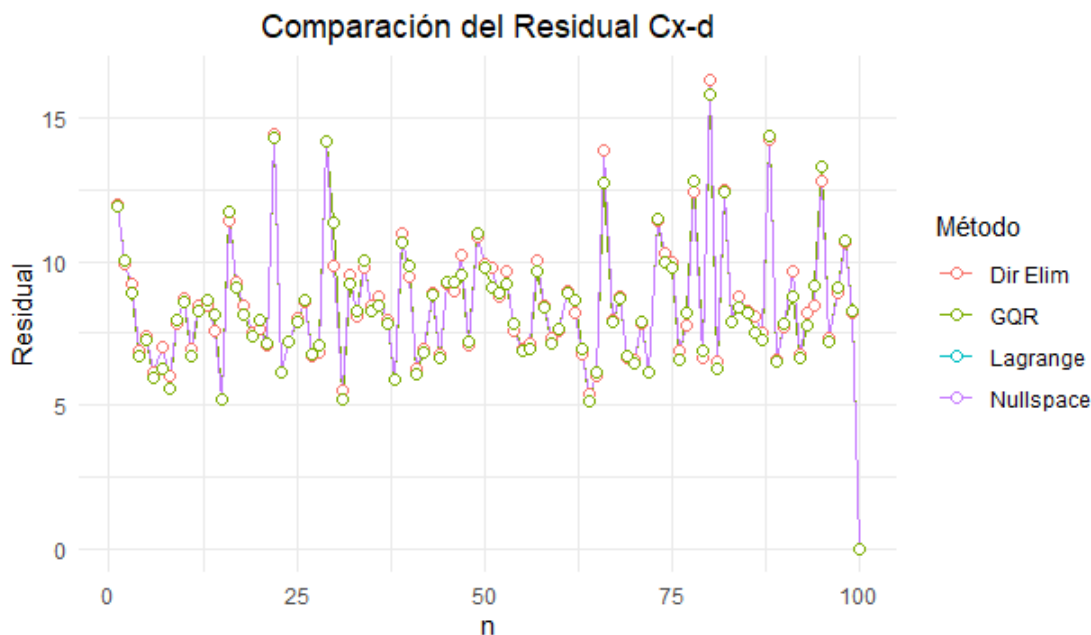


Figura 5-6.: Residual $Cx - d$ para 4 métodos de solución

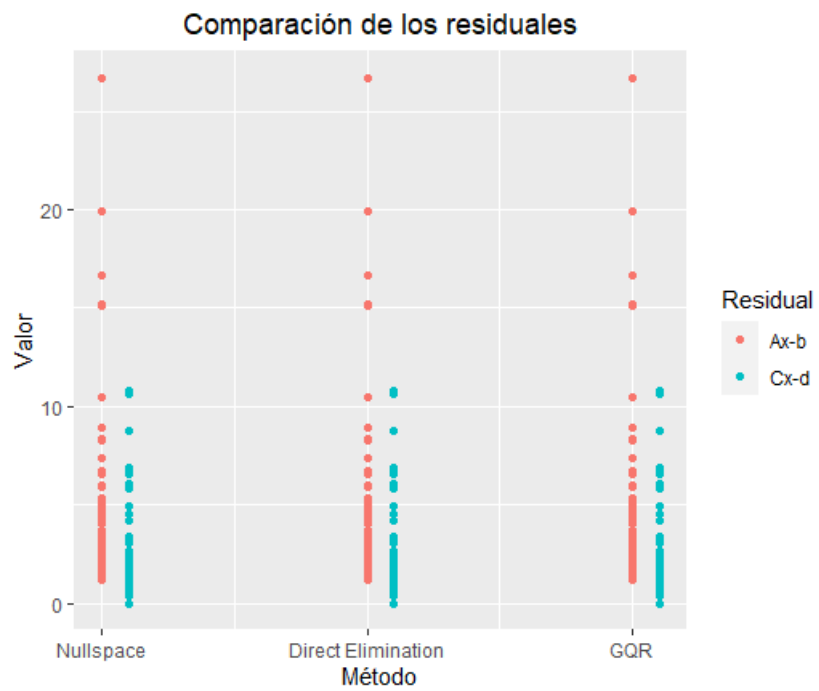
En la gráfica 5-5 no se observan diferencias significativas, y la tabla 5-3 muestra que entre el grupo 1 y el método GQR, la norma euclidiana del residual $Ax - b$ para GQR fue inferior entre los 500 problemas distintos para el 53.8% de los casos, evidencia de que a pesar del aumento del doble de ecuaciones y variables, se mantienen las similitudes entre los cuatro métodos en la comparación del grupo 1 y el método GQR.

Análogamente, analizando el residual $Cx - d$, según la gráfica 5-6, no se aprecian altas discrepancias en el desempeño de los tres métodos, del total de problemas un 55.2% muestra un inferior valor para el residual de la restricción en el método GQR.

Etapa 2: Multicolinealidad

La correlación entre variables afecta las estimaciones como se enunció en §2.3, modelos sobreajustados o modelos en los que existe y no se prueba la multicolinealidad, pierden la validez de las estimaciones, para esta etapa, se trabaja particularmente con la colinealidad, es decir, la presencia de 2 variables correlacionadas. Se analiza la presencia de multicolinealidad para A , para C y para A y C al tiempo.

Dada una matriz $A \in \mathbb{R}^{m \times n}$, con rango k , tal que $k < n$, y a través de 100 iteraciones para los tres métodos, con variaciones en los tamaños de las matrices de diseño para $m = 10$ y $m = 100$, se obtienen las comparaciones de los residuales para $Ax - b$ y $Cx - d$ como sigue.

Subetapa 2.1: Colinealidad para $m=10$ Figura 5-7.: Residuales para la colinealidad en A con $m = 10$

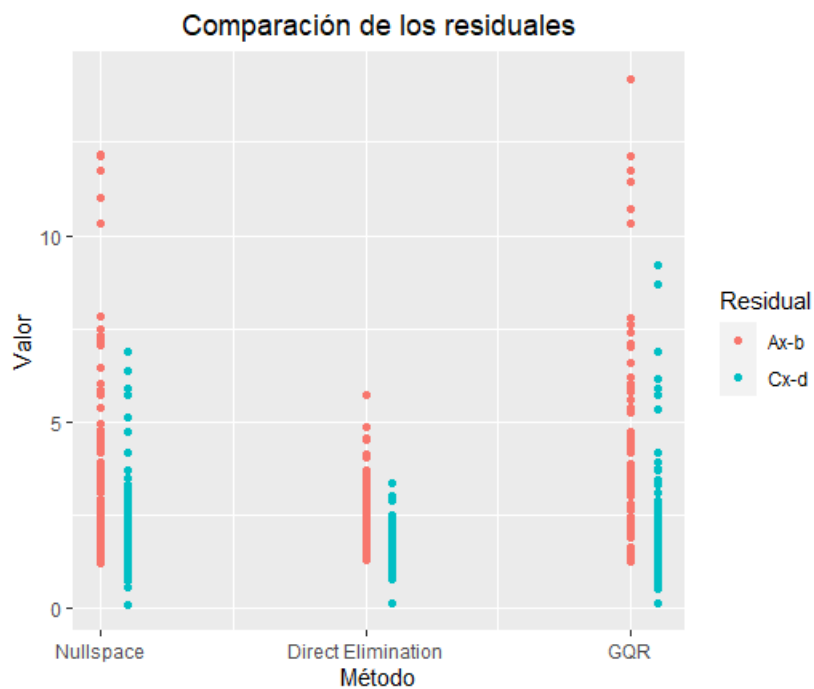


Figura 5-8.: Residuales para la colinealidad en A y C con $m = 10$

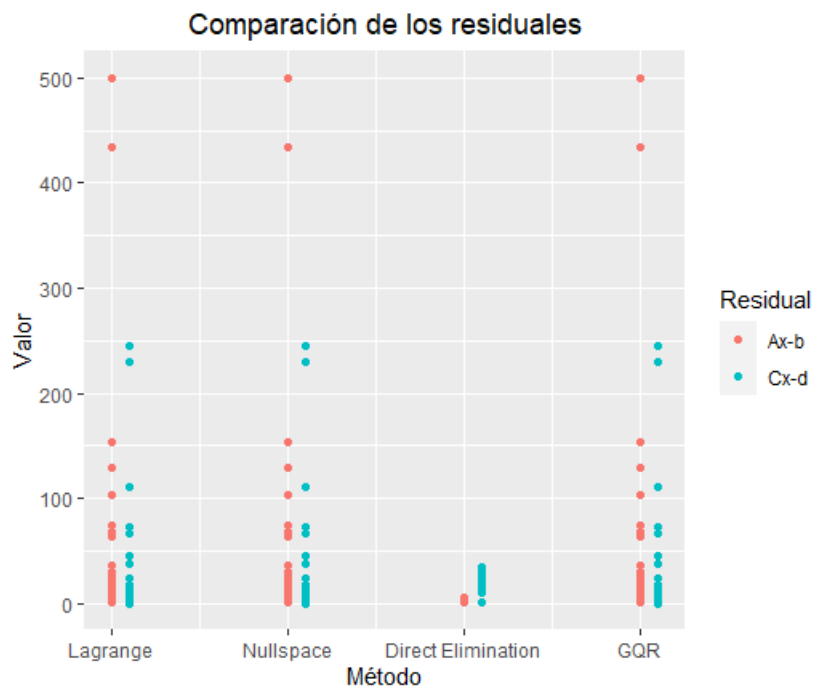
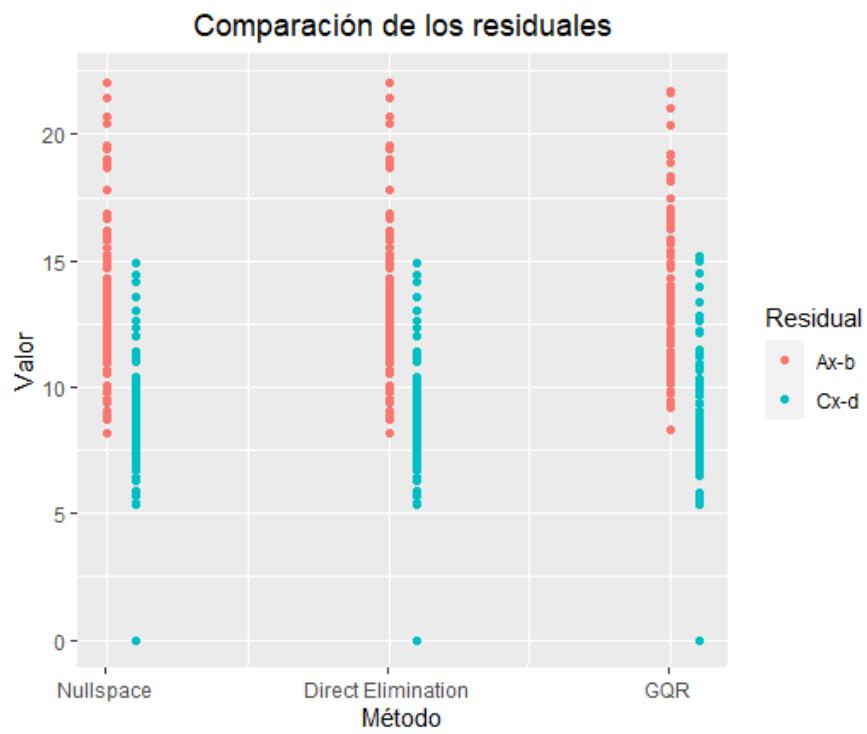


Figura 5-9.: Residuales para la colinealidad en C con $m = 10$

Subetapa 2.2: Colinealidad para $m=100$ Figura 5-10.: Residuales para la colinealidad en A con $m = 100$

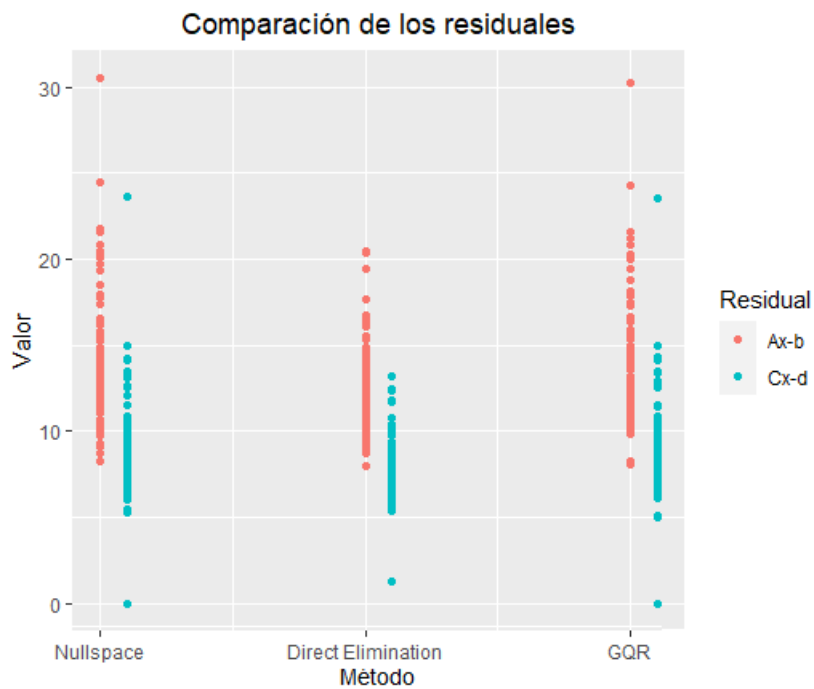


Figura 5-11.: Residuales para la colinealidad en A y C con $m = 100$

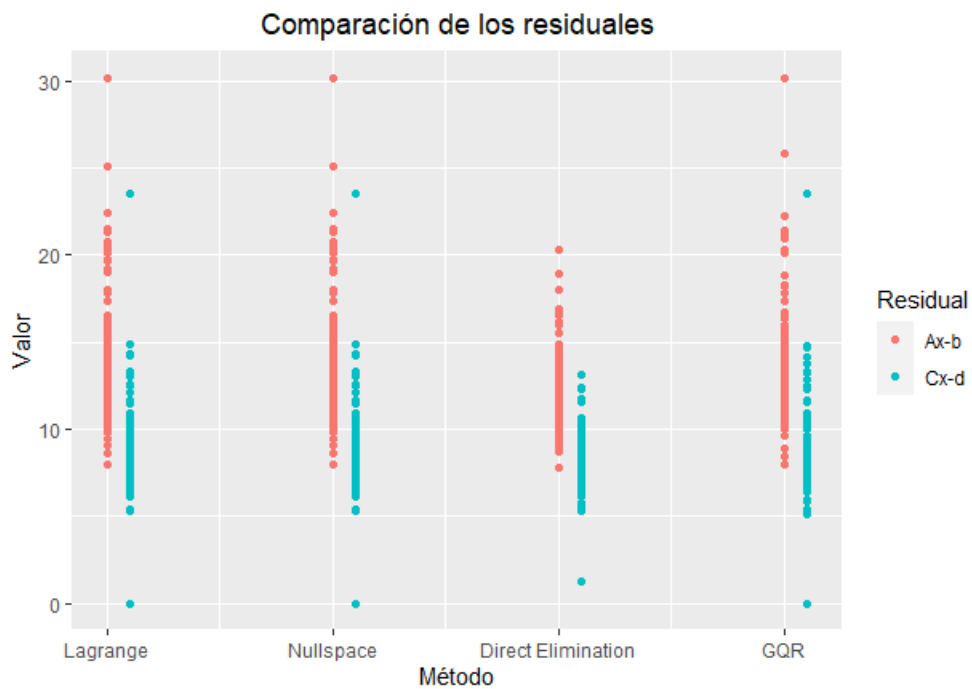


Figura 5-12.: Residuales para la colinealidad en C con $m = 100$

El método basado en multiplicadores de Lagrange no es útil cuando se presenta colinealidad en la matriz de diseño, tal como se mencionó en §3.1, y cuando solo se presenta colinealidad en la matriz de restricciones se mantiene la estabilidad en sus resultados tal como para el método del espacio nulo y GQR.

Las diferencias que se obtuvieron en la simulación de los 100 problemas se dió para la colinealidad en la matriz de restricciones C a favor del método de eliminación directa, y para las matrices con colinealidad simultánea en A y C cuando aumenta el tamaño de la muestra en $m = 100$ el valor de ambos residuales disminuye ligeramente respecto de los demás métodos. Por otra parte, no se aprecian diferencias significativas entre el método del espacio nulo y GQR para el tamaño de las matrices, para el tipo de colinealidad o de los residuales $Ax - b$, ni $Cx - d$.

Etapa 3: Cuaterniones

La dinámica del algoritmo para su aplicación en cuaterniones, se evidencia a través de valores simulados a partir de números aleatorios distribuidos normalmente. Son puestas a prueba 100 matrices aleatorias tales que $A \in \mathbb{C}^{3 \times 2}$ y $C \in \mathbb{C}^{1 \times 2}$, obteniéndose solución para los cuatro métodos: Lagrange, Espacio Nulo, Eliminación Directa y GQR.

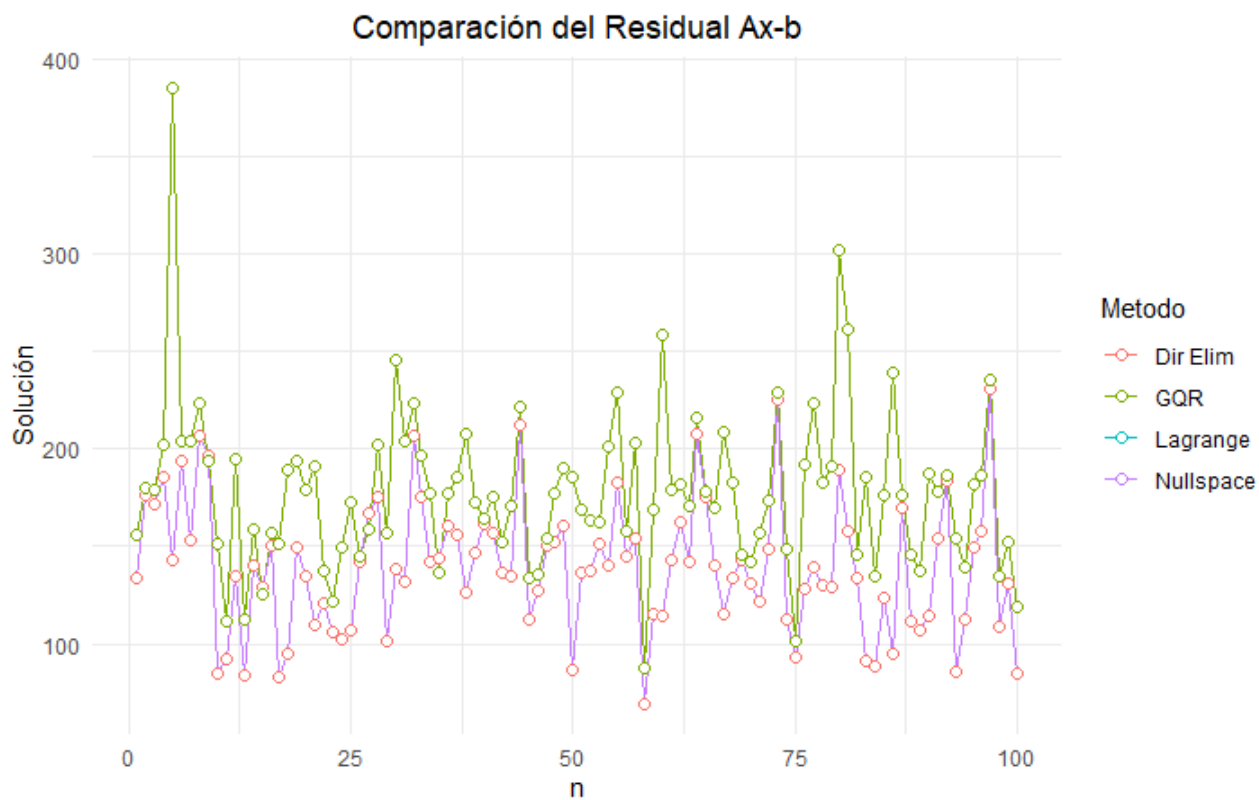


Figura 5-13.: Residuales para los métodos de solución para cuaterniones

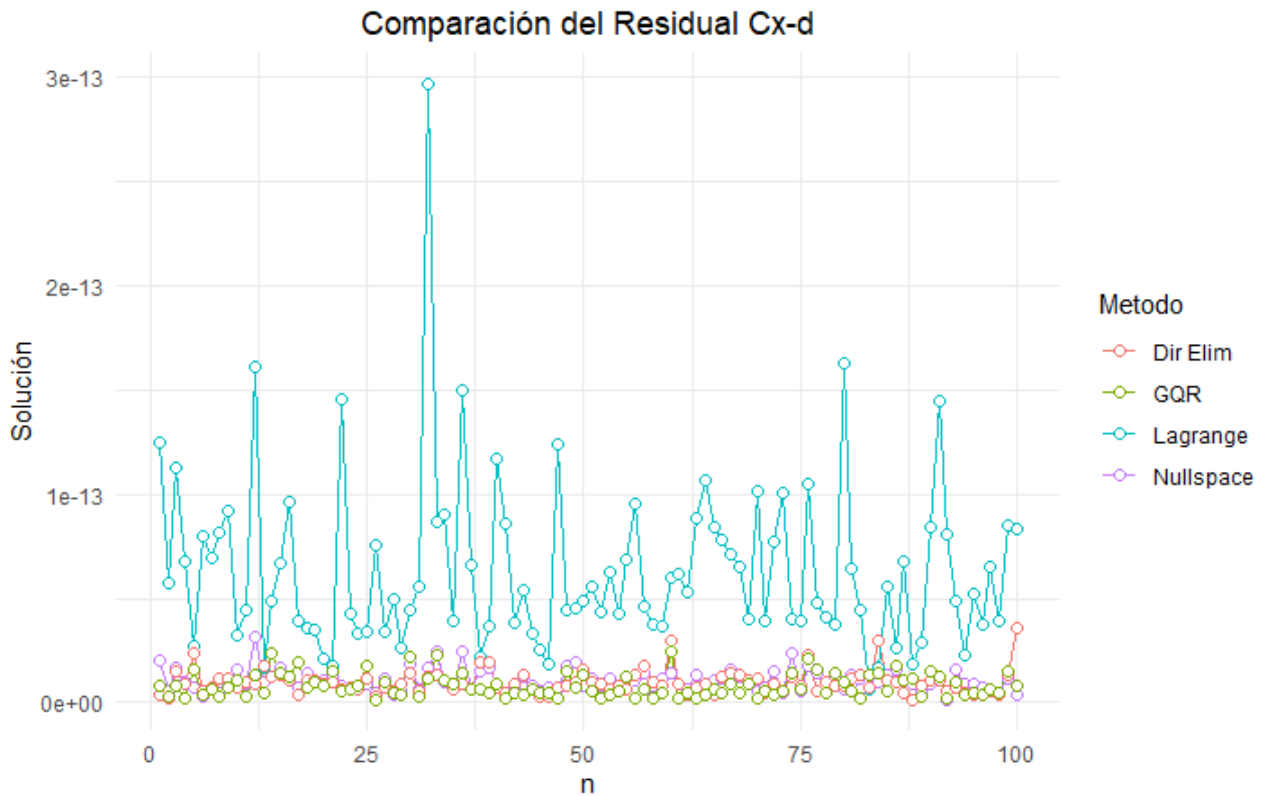


Figura 5-14.: Residuales para los métodos de solución para cuaterniones

Debido a que los resultados de solución son matrices complejas, la comparación de los métodos es evaluada a través de la norma Frobenius, pues por definición es apropiada para el uso de matrices.

De las 100 matrices simuladas, la norma Frobenius para los residuales $Ax - b$ fueron en un 94% menores para el grupo 1 contra el método GQR; para el residual $Cx - d$ fueron menores en un 74% de los problemas evaluados a favor del método GQR. Aunque el porcentaje es evidencia de diferencias significativas en los residuales a través de la norma Frobenius y que son compensadas entre ambos criterios, se observa que salvo casos excepcionales el rendimiento del método GQR presenta altas similitudes entre los cien problemas. El método basado en multiplicadores de Lagrange presenta diferencias en el valor del residual $Cx - d$ respecto de los demás métodos en la totalidad de los problemas simulados.

6. Conclusiones y recomendaciones

6.1. Conclusiones

El estimador de mínimos cuadrados restringidos basado en los multiplicadores de Lagrange, garantiza la existencia de un estimador insesgado y eficiente si las restricciones impuestas en el modelo son verdaderas respecto del estimador no restringido. Además, si las restricciones no son verdaderas, se obtiene un estimador que a pesar de ser sesgado permanece siendo más eficiente que el estimador sin las restricciones. La comparación entre ambos modelos ha sido establecida mediante la metodología de la prueba F y la prueba de hipótesis del error cuadrático medio.

La resolución de problemas de mínimos cuadrados restringidos de manera sistemática ha sido expuesta en el Apéndice B, desarrollando códigos de programación sin antecedentes previos de ninguno de los siguientes algoritmos en el software R a través del paquete *LSE* que se encuentra en el repositorio CRAN y GitHub:

- Factorización RQ de una matriz.
- Factorización QR generalizada de una matriz.
- Desarrollo de mínimos cuadrados restringidos a través de la factorización QR generalizada.
- Desarrollo de mínimos cuadrados restringidos a través de multiplicadores de Lagrange.
- Desarrollo de mínimos cuadrados restringidos a través del método del espacio nulo.
- Desarrollo de mínimos cuadrados restringidos a través del método de eliminación directa.
- Transformaciones matriciales que permiten pasar del cuerpo de los cuaterniones a los reales y viceversa.

La comparación de métodos permite demostrar la validez de los resultados del método GQR, las estimaciones obtenidas han sido evaluadas a través de una simulación entre los cuatro métodos de solución desarrollados. En términos de residuales no hay diferencias superiores al 56 % entre las estimaciones obtenidas del método GQR frente a los demás métodos para el

caso de matrices sobredeterminadas ni para el caso de colinealidad en la matriz de diseño A , de la matriz de restricciones C , y en presencia de colinealidad en ambas matrices simultáneamente. En lo que respecta a la aplicación de cuaterniones, el comportamiento del algoritmo GQR muestra ser útil, no así el método de multiplicadores de Lagrange en la medición de la norma Frobenius para el residual de la restricción $Cx - d$ en comparación con los métodos restantes.

6.2. Recomendaciones

Se recomienda el estudio para la obtención de un estimador que no se base en multiplicadores de Lagrange, dada la ineficiencia que muestra el método en algunas situaciones, como en las estimaciones de la varianza del estimador, la cual recae en otros procesos como cálculo de errores estándar, que influye en intervalos de confianza y predicciones. Además, es un método muy limitante y genera poca confiabilidad a pesar de la facilidad para su implementación. Aunque permite el desarrollo de los momentos para el estimador restringido, la falta de versatilidad para los factores expuestos en la simulación pone en duda que las estimaciones sean válidas.

El método de eliminación directa y el método del espacio nulo presentaron un buen rendimiento en términos de estabilidad numérica y exactitud de los resultados para todos los factores considerados en la simulación.

Evaluar el comportamiento de otros factores que permitan analizar el modelo de mínimos cuadrados restringidos, en caso de ser apropiado los factores pueden incluir, y sin limitarse a:

- El uso de otras factorizaciones matriciales como LU, Choleski, descomposición de valores singulares (DVS).
- El uso de otros métodos numéricos como gradiente descendiente, SOR, etc.; y métodos basados en técnicas de regularización.
- En contextos de sensibilidad de resultados, como el número de condición, el error hacia adelante o hacia atrás, etc.
- En matrices con pivoteo, actualizando filas o columnas, matrices densas o dispersas, etc.

Bibliografía

- [1] ABDI, Hervé [u. a.]: The method of least squares. En: *Encyclopedia of measurement and statistics* 1 (2007), p. 530–532
- [2] ALDRICH, John [u. a.]: Fisher and Regression. En: *Statistical Science* 20 (2005), Nr. 4, p. 401–417
- [3] ANDERSON, E ; BAI, Zhaojun ; DONGARRA, J: Generalized qr factorization and its applications. En: *Linear Algebra and its Applications* 162 (1992), p. 243–271
- [4] BARLOW, Jesse L.: Error analysis and implementation aspects of deferred correction for equality constrained least squares problems. En: *SIAM journal on numerical analysis* 25 (1988), Nr. 6, p. 1340–1358
- [5] BARLOW, Jesse L. ; NICHOLS, NK ; PLEMMONS, Robert J.: Iterative methods for equality-constrained least squares problems. En: *SIAM journal on scientific and statistical computing* 9 (1988), Nr. 5, p. 892–906
- [6] BELL, Eric T.: *Men of Mathematics: The Lives and Achievements of the Great Mathematicians from Zeno to Poincaré*. USA : Simon and Schuster, 2014
- [7] BERKSON, Joseph: Estimation by least squares and by maximum likelihood. (2020), p. 1–12
- [8] BJÖRCK, Åke: *Numerical Methods for Least Squares Problems*. Suecia : SIAM, 1996
- [9] BOYD, Stephen ; VANDENBERGHE, Lieven: *Introduction to applied linear algebra*. Cambridge university press, 2018
- [10] BURTON, D: *The history of Mathematics. 7 Edition*. USA : MCGraw Hill, 2011
- [11] COX, Anthony J. ; HIGHAM, Nicholas J.: Accuracy and stability of the null space method for equality constrained least square problem. En: *BIT Numerical Mathematics* 39 (1999), Nr. 1, p. 34–50
- [12] COX, Anthony J. ; HIGHAM, Nicholas J.: Backward error bounds for Constrained Least Squares Problems. En: *BIT Numerical Mathematics* 39 (1999), Nr. 2, p. 210–227

-
- [13] DARLINGTON, Richard B. ; HAYES, Andrew F.: *Regression Analysis and Linear Models*. 2017
- [14] DE MOOR, Bart ; VAN DOOREN, Paul: Generalizations of the singular value and QR decompositions. En: *SIAM journal on matrix analysis and applications* 13 (1992), Nr. 4, p. 993–1014
- [15] DRAPER, Norman R. ; SMITH, Harry: *Applied Regression Analysis. 3 Edition*. Vol. 326. John Wiley & Sons, 1998
- [16] EDNEY, Matthew H. ; PEDLEY, Mary S.: *The History of Cartography. Vol. 4.* Vol. 4. University of Chicago Press, 2020
- [17] EL-DERENI, M ; RASHWAN, N: Solving Multocollinearity problem. 6 No. 12 (2011), p. 585–600
- [18] ESTRADA, Hernán ; RUIZ, Jorge M. ; TRIANA, Juan G.: El origen del metro y la confianza en las matemáticas. En: *Matemáticas: Enseñanza Universitaria* 19 (2011), Nr. 1, p. 89–101
- [19] GALLIGANI, Emanuele ; ZANNI, Luca: Error analysis of elimination methods for equality constrained quadratic programming problems. En: *Mathematical Research* 89 (1996), p. 107–112
- [20] GALTON, Francis: Regression towards mediocrity in hereditary stature. En: *The Journal of the Anthropological Institute of Great Britain and Ireland* 15 (1886), p. 246–263
- [21] GENTLE, James E.: *Matrix Algebra. Theory, Computations, and Applications in Statistics*. Springer, 2017. – 978–0 p.
- [22] GILONI, A ; PADBERG, M: Alternative methods of linear regression. En: *Mathematical and Computer Modelling* 35 (2002), Nr. 3-4, p. 361–374
- [23] GOLUB, Gene H. ; VAN LOAN, Charles F.: *Matrix Computations. 4 Edition*. Vol. 3. JHU press, 2013
- [24] GREENE, William H. ; SEAKS, Terry G.: The restricted least squares estimator: a pedagogical note. En: *The Review of Economics and Statistics* (1991), p. 563–567
- [25] GUERRERO, Sara C.: Una metodología para el tratamiento de la multicolinealidad a través del escalamiento multidimensional. En: *Ciencia en Desarrollo* 8 (2017), Nr. 2, p. 9–24
- [26] GUJARATI, D ; PORTER, D: *Econometría, 5 Edición*. México D.F. : McGraw Hill, 2009
- [27] HANSON, Andrew J.: *Visualizing quaternions*. Morgan Kaufmann Publisher, 2005

-
- [28] HARTER, W L.: The method of least squares and some alternatives. Part I. En: *International Statistical Review/Revue Internationale de Statistique* (1974), p. 147–174
- [29] HIGHAM, Nicholas J.: *Accuracy and Stability of Numerical Algorithms*. SIAM, 2002
- [30] HILL, R C. ; GRIFFITHS, William E. ; LIM, Guay C.: *Principles of Econometrics, 5 Edition*. John Wiley & Sons, 2018
- [31] JIANG, Tongsong ; WEI, Musheng: Equality Constrained Least Squares Problem over Quaternion Field. En: *Applied mathematics letters* 16 (2003), Nr. 6, p. 883–888
- [32] JIANG, Tongsong ; ZHAO, Jianli ; WEI, Musheng: A new technique of quaternion equality constrained least squares problem. En: *Journal of computational and applied mathematics* 216 (2008), Nr. 2, p. 509–513
- [33] JOHNSON, Norman L. ; KOTZ, Samuel ; BALAKRISHNAN, Narayanaswamy: *Continuous Univariate Distributions*. Vol. 289. John wiley & sons, 1995
- [34] KOENKER, Roger: Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics. En: *Journal of Econometrics* 95 (2000), Nr. 2, p. 347–374
- [35] LAWSON, Charles L. ; HANSON, Richard J.: *Solving Least Squares Problems*. SIAM, 1995
- [36] LEHMANN, Erich L. ; CASELLA, George: *Theory of Point Estimation. Second Edition*. Springer Science & Business Media, 2006
- [37] LUENBERGER, David G. ; YE, Yinyu [u. a.]: *Linear and nonlinear programming*. Vol. 2. Springer, 1984
- [38] MERRIMAN, Mansfield: On the history of the method of least squares. En: *The Analyst* 4 (1877), Nr. 2, p. 33–36
- [39] MUÑOZ, José Manuel S.: Historias de matemáticas Hamilton y el descubrimiento de los cuaterniones. En: *Pensamiento matemático* (2011), Nr. 1, p. 7
- [40] NIEVERGELT, Yves: A tutorial history of least squares with applications to astronomy and geodesy. En: *Journal of Computational and Applied Mathematics* 121 (2000), Nr. 1-2, p. 37–72
- [41] NÖEL, G ; TILLEUIL, P: D’où sort la méthode des moindres carrés. No. 151 (2005), p. 17–44
- [42] RADHAKRISHNA RAO, C ; TOUTENBURG, H ; HEUMANN, C: *Linear Models and Generalizations Least Squares and Alternatives*. Berlin, Germany: Springer, 2008

- [43] RAMBAUX, N ; WILLIAMS, JG: The Moon's physical librations and determination of their free modes. En: *Celestial Mechanics and Dynamical Astronomy* 109 (2011), Nr. 1, p. 85–100
- [44] SALKUYEH, Davod K.: A preconditioner for the normal equations. En: *Journal of applied mathematics & informatics* 28 (2010), Nr. 3.4, p. 687–696
- [45] STANTON, Jeffrey M.: Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors. En: *Journal of Statistics Education* 9 (2001), Nr. 3
- [46] STIGLER, Stephen M.: Gauss and the invention of least squares. En: *the Annals of Statistics* (1981), p. 465–474
- [47] STIGLER, Stephen M.: *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press, 1986
- [48] STIGLER, Stephen M.: Regression towards the mean, historically considered. En: *Statistical methods in medical research* 6 (1997), Nr. 2, p. 103–114
- [49] STILLMAN, Drake: *Dialogue concerning the Two Chief World Systems -Ptolomaic & Copernican-*. London : University of california Press, Berkeley and Los Angeles, California, 1967
- [50] THEIL, Henri: *Economic forecasts and policy*. (1961)
- [51] THEIL, Henri: *Principles of Econometrics*. USA : Wiley, 1971
- [52] TORO-VIZCARRONDO, Carlos ; WALLACE, T D.: A test of the mean square error criterion for restrictions in linear regression. En: *Journal of the American Statistical Association* 63 (1968), Nr. 322, p. 558–572
- [53] VAN BENTHEM, Mark H. ; KEENAN, Michael R. ; HAALAND, David M.: Application of equality constraints on variables during alternating least squares procedures. En: *Journal of Chemometrics: A Journal of the Chemometrics Society* 16 (2002), Nr. 12, p. 613–622
- [54] VOSS, H ; HOGBEN, L: *Handbook of Linear Algebra*. (2014)
- [55] WALKER, Esteban ; O'BRIEN, Ralph G.: Using Restricted Least Squares to Delineate the Effects of Misspecification in Linear Models. En: *Journal of the Royal Statistical Society: Series D (The Statistician)* 41 (1992), Nr. 4, p. 467–476
- [56] WATKINS, David S.: *Fundamentals of Matrix Computation*. Vol. 64. John Wiley & Sons, 2004

-
- [57] WOOLDRIDGE, Jeffrey M.: *Introducción a la econometría*. Editorial Paraninfo, 2006
- [58] ZEB, Salman ; YOUSAF, Muhammad: Updating QR factorization procedure for solution of linear least squares problem with equality constraints. En: *Journal of inequalities and applications* 2017 (2017), Nr. 1, p. 1–17
- [59] ZHDANOV, Aleksandr I. ; GOGOLEVA, Sofya Y.: Solving Least Squares Problems With Equality Constraints Based On Augmented Regularized Normal Equations. En: *Appl. Math. E-Notes* 15 (2015), p. 218–224
- [60] ZOU, Cuiming ; KOU, Kit I. ; DONG, Li ; ZHENG, Xianwei ; TANG, Yuan Y.: From grayscale to color: Quaternion linear regression for color face recognition. En: *IEEE Access* 7 (2019), p. 154131–154140

A. Ortogonalidad

A.1. Propiedades ortogonales

El uso de vectores ortogonales dada su caracterización es útil para mejorar la estabilidad numérica de los resultados, pues permite preservar si así se desea, las columnas o filas más significativas, es decir, aquellas que aportan más información, entonces se reducen las dimensiones originales.

Definición 6 (Vector ortogonal). *Dos vectores x_i y x_j son ortogonales si se cumple que $x_i^T x_j = 0$, este hecho es representado también como $x_i \perp x_j$. De manera más general, se puede pensar la ortogonalidad entre dos vectores como $a^H b = 0$, para el caso donde $a, b \in \mathbb{C}$, pues ésta es la definición de producto interno en el conjunto de los números complejos.*

Definición 7 (Norma de un vector ortogonal). *Sea un vector x transformado por una matriz ortogonal Q así $\|Qx\|$. Se puede probar que $\|Qx\| = \|x\|$.*

Demostración. Sea $\|\cdot\|$ la norma euclídea, por definición

$$\|x\| = \sqrt{\langle x, x \rangle}$$

Tomando cuadrados se tiene $\|x\|^2 = \langle x, x \rangle$, entonces para el caso en $\|Qx\|$ a través de dicha igualdad, obtenemos

$$\begin{aligned} \|Qx\|^2 &= \langle Qx, Qx \rangle \\ \langle Qx, Qx \rangle &= (Qx)^T (Qx) = x^T Q^T Qx = x^T Ix = x^T x \end{aligned}$$

Y finalmente tomando raíz cuadrada se obtiene

$$\sqrt{\langle x, x \rangle} = \|x\|$$

□

De tal manera se afirma que una transformación ortogonal preserva longitudes, las transformaciones constan de reflexiones, rotaciones y traslaciones.

También se puede observar que el ángulo entre dos vectores es el mismo a través de transformaciones ortogonales, recuerde que

$$\langle x, y \rangle = \|x\| \|y\| \cos(\theta)$$

entonces

$$\arccos \frac{(Qx, Qy)}{\|Qx\|_2 \|Qy\|_2} = \arccos \frac{(x, y)}{\|x\|_2 \|y\|_2}$$

Definición 8 (Vector ortonormal). *Cuando la longitud del vector es equivalente a la longitud de un vector unitario, calculado a través de $x_i/\|x_i\|_2$, y además el vector es ortogonal, se dice que se obtiene un vector ortonormal.*

Definición 9 (Independencia lineal en vectores ortonormales). *Un conjunto de vectores ortonormales son linealmente independientes.*

Demostración. Si la colección $x_1, x_2, x_3, \dots, X_i$ se supone que es un conjunto ortonormal, además se supone que $\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_i x_i$, entonces

$$\begin{aligned} 0 &= x_i (\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_i x_i) \\ 0 &= \alpha_1 (x_i^T x_1) + \alpha_2 (x_i^T x_2) + \dots + \alpha_i (x_i^T x_i) + \dots + \alpha_i (x_i^T x_i) \end{aligned}$$

Para el componente x_i dado que es ortonormal, entonces $x_i x_i = 1$, pero así se incumpliría el supuesto planteado inicialmente, por lo tanto, solo se mantiene la igualdad si todos los coeficientes $\alpha_i = 0$ dado que los vectores sean ortonormales, véase [9]. \square

Definición 10 (Subespacio ortogonal). *Dos subespacios son mutuamente ortogonales cuando $x^T y = 0$, si $x \in S_i$ y $y \in S_j$, con $i \neq j$. Un vector que es ortogonal a todos los vectores de un subespacio se dice que es un vector normal al subespacio.*

Definición 11 (Vector ortonormal). *El complemento ortogonal de un subespacio A , es una relación simétrica entre dos subespacios, donde dicho complemento, denotado por A^\perp es el conjunto de todos los vectores ortogonales a A .*

Definición 12 (Matriz ortogonal). *Una matriz A se dice ortogonal si $AA^T = A^T A = I$, pues la matriz está conformada por filas o columnas ortonormales. Para el caso $A \in \mathbb{C}$, se denomina matriz unitaria si $AA^H = A^H A = I$, siendo A^H la matriz transpuesta conjugada. Por lo tanto se concluye para el caso real y complejo respectivamente que $A^{-1} = A^T$ y $A^{-1} = A^H$.*

Definición 13 (Dimensión de una matriz ortogonal). *El espacio generado por las columnas de la matriz A , cumple que si $A \in \mathbb{R}^{m \times n}$, y el rango de la matriz es $\text{rango}(A) = k(\min\{m, n\})$, entonces $\text{rango}(A) = \text{nulidad}(A)^\perp$*

A.2. Proyección

Definición 14 (Proyección). *La proyección de un vector x sobre un vector y , donde $x \neq 0$ es*

$$P_y x = \frac{\langle x, y \rangle}{\|y\|^2} y$$

Definición 15 (Proyección Ortogonal). *Si y es la proyección, y el vector resultante de la diferencia entre la proyección y el vector proyectado es un vector ortogonal al subespacio donde pertenece y , entonces $(y - x)y = 0$, ésta es la base sobre la cual se maneja la minimización de este vector, llamado residual, para el ajuste del método de mínimos cuadrados, y se denomina proyección ortogonal.*

En términos de proyección recordemos que se obtiene ortogonalidad entre los vectores, si $P_y x = 0$. Ahora, para una proyección como sucede en mínimos cuadrados, sea el subespacio $V \in \mathbb{R}^n$, al proyectar un vector x sobre V , se obtiene el vector z , si esta proyección es ortogonal, entonces es la menor distancia, y se asevera que $|x - z| < |x - w|, \forall w \in V : w \neq z$, y también que $P_z(x - z) = 0$.

En palabras, se tiene que el vector $(x - z)$ es el vector residual, del cual se obtendrá el que tenga la menor norma, a través del vector z , que es la proyección del vector respuesta sobre el espacio generado por las columnas de A , denotado usualmente por $C(A)$, así es como z es una combinación lineal de las columnas de A que garantiza minimizar el vector residual.

De manera matricial, se puede obtener el operador de proyección a través de una matriz $A \in \mathbb{C}$, con las siguientes características de A :

- A es una matriz simétrica, i.e., $A^T = A$.
- A es una matriz idempotente, i.e., $A^2 = A$.
- $\|P_y x\| \leq \|x\|$
- A es una proyección ortogonal si $Nulidad(A) = (Imagen(A))^\perp$.
- Si B es una matriz ortonormal de rango completo k , entonces $BB^H = A$, la matriz A es una proyección ortogonal.
- Dados distintos vectores que generen el mismo subespacio, la matriz de proyección es única.

En el contexto de la proyección para la aplicación en mínimos cuadrados se tiene que del modelo $Ax - b$, donde b es el vector a ser ajustado, se realiza una proyección ortogonal de b sobre el espacio generado por las columnas de A , que es el espacio columna $C(A)$, ésta proyección es el vector Ax . Si A es una matriz de rango completo, entonces la proyección es $P_{C(A)} = A(A^T A)^{-1} A^T$. Por último, el vector residual ε puede ser expresado en términos de $\varepsilon = (I - P_{C(A)})b$.

A.3. Métodos de ortogonalización

Con el objetivo de realizar una descomposición matricial, para facilitar los procesos numéricos, se aplica una transformación ortogonal a una matriz para que las filas ó las columnas sean linealmente independientes, dadas las razones expuestas en §2.3 para justificar que el eje central es la factorización QR, se presentan tres métodos comúnmente utilizados.

Si bien, para los tres casos se define la factorización QR, la cual consiste en ortogonalizar las columnas de una matriz, empezando por la primera hasta la última; se define la factorización RQ de una matriz como el proceso de ortogonalizar las filas de una matriz iniciando con la última hasta terminar en la primera fila, para lo cual, los tres métodos son válidos a través de los siguientes pasos.

Dada una matriz $A \in \mathbb{R}^{m \times n}$, la factorización RQ comprende la siguiente secuencia

- Sea \hat{A} la matriz A a la cual se le ha invertido el orden a las filas.
- Se calcula la factorización QR de $\hat{A}^T = \hat{Q}\hat{R}$.
- Se obtiene la matriz ortogonal Q a través de invertir nuevamente las filas de la matriz \hat{Q} .
- El factor triangular \hat{R} es una matriz triangular inferior, al permutar las filas y posteriormente las columnas se obtiene R como triangular superior.

En caso de ser $A \in \mathbb{R}^{m \times m}$, se obtiene la permutación de filas a través de una matriz

$$P = \begin{bmatrix} 0 & \dots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \dots & 0 \end{bmatrix}$$

Así, PA indica la permutación de las filas y AP la permutación de las columnas. En §B se expone el código asociado a la factorización RQ para una matriz $A \in \mathbb{R}^{m \times n}$.

A.3.1. Proceso de ortogonalización de Gram-Schmidt

Sean x_1, x_2, \dots, x_m vectores linealmente independientes, se busca general el mismo espacio a través de $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m$ vectores ortogonales. El proceso es realizado a través del uso de proyección ortogonal como sigue.

Los vectores y denotarán los nuevos vectores ortogonales, para el primer caso tenemos:

$$y_1 = x_1$$

El segundo vector y_2 , es la diferencia -la cual es ortogonal-, del vector x_2 con la proyección ortogonal de x_2 sobre y_1

$$y_2 = x_2 - \frac{\langle x_2, y_1 \rangle}{\langle y_1, y_1 \rangle} y_1$$

Para el tercer caso se acumulan las diferencias de x_3 sobre proyecciones ortogonales de los vectores anteriores

$$y_3 = x_3 - \frac{\langle x_3, y_1 \rangle}{\langle y_1, y_1 \rangle} y_1 - y_2 = x_3 - \frac{\langle x_3, y_1 \rangle}{\langle y_1, y_1 \rangle} y_1 - \frac{\langle x_2, y_2 \rangle}{\langle y_2, y_2 \rangle} y_2$$

En general, el cálculo está determinado por

$$y_m = x_m - \sum_{i=1}^{m-1} \frac{\langle x_m, y_i \rangle}{\langle y_i, y_i \rangle} y_i$$

Finalmente, la base ortogonal está formada por y_1, y_2, \dots, y_m . La base obtenida puede hacerse ortonormal en caso de ser requerido a través de

$$\frac{y_1}{\|y_1\|}, \frac{y_2}{\|y_2\|}, \dots, \frac{y_m}{\|y_m\|}$$

En caso de terminar el algoritmo antes de obtener los m vectores, el algoritmo ofrece claras evidencias de dependencia lineal entre el vector en el cual se haya obtenido $y_i = 0$.

Puede verse en Bjorck, una versión modificada, que busca mejorar la estabilidad numérica del método expuesto. A continuación aplicaremos otras técnicas basadas en los conceptos de reflexión y rotación

A.3.2. Rotaciones de Givens

Además de que una transformación ortogonal preserva longitudes, también se demostró que preserva ángulos, si se cumplen estas dos condiciones se denomina transformación isométrica.

El operador matricial para una rotación en sentido horario está definido por

$$Q = \begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) \\ \text{sen}(\theta) & \cos(\theta) \end{bmatrix}$$

Es trivial que esta matriz es ortogonal.

Las rotaciones permiten la obtención de ceros, para este caso en \mathbb{R}^2 , son ceros en la segunda coordenada, a través del sistema $Q^T x = z$, que de manera expandida es

$$\begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) \\ \text{sen}(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} z \\ 0 \end{bmatrix}$$

El proceso es secuencial, para cada columna, se giran los vectores y se forman ceros bajo el pivote, de esta manera, se forma la matriz triangular para el cálculo de R .

Suponga la matriz en \mathbb{R}^2 , definida por

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Inicialmente, el proceso es el cálculo de la rotación para la primer columna de A a través del sistema $Q^T x = z$, es decir,

$$\begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) \\ \text{sen}(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} = \begin{bmatrix} z_{11} \\ 0 \end{bmatrix}$$

Donde $z_1 = \sqrt{a_{11}^2 + a_{21}^2}$. Así, se obtiene la primer columna de la matriz triangular superior R , la segunda columna es

$$\begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) \\ \text{sen}(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} = \begin{bmatrix} z_{12} \\ z_{22} \end{bmatrix}$$

Entonces se cumple que $A = QR$, donde Q es la matriz de rotación obtenida del producto de las matrices utilizadas en el proceso, y R es

$$R = \begin{bmatrix} z_{11} & z_{12} \\ 0 & z_{22} \end{bmatrix}$$

De manera general, este proceso se obtiene de una rotación en los componentes i, j del sistema $Q^T x = z$, que al expandirse es

$$\begin{bmatrix}
 1 & \dots & z_{1i} & \dots & z_{ki} & \dots & z_{ji} & \dots & z_{1n} \\
 \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
 0 & \dots & \cos(\theta) & \dots & \dots & \dots & \sin(\theta) & \dots & z_{in} \\
 \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
 0 & \dots & \dots & \dots & 1 & \dots & \dots & \dots & z_{kn} \\
 \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
 0 & \dots & -\sin(\theta) & \dots & \dots & \dots & \cos(\theta) & \dots & z_{jn} \\
 \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
 0 & \dots & 0 & \dots & 0 & \dots & 0 & \dots & 1
 \end{bmatrix}
 \begin{bmatrix}
 x_{1i} \\
 \vdots \\
 x_{1i} \\
 \vdots \\
 x_{1k} \\
 \vdots \\
 x_{1j} \\
 \vdots \\
 x_{in}
 \end{bmatrix}
 =
 \begin{bmatrix}
 z_{1i} \\
 \vdots \\
 z_{1i} \\
 \vdots \\
 z_{1k} \\
 \vdots \\
 z_{1j} \\
 \vdots \\
 z_{in}
 \end{bmatrix}$$

A.3.3. Reflexiones de Householder

La reflexión también es representada a través de una matriz, pero, en principio observe que para realizar una reflexión a través de una recta, sea ésta la recta x , se puede obtener un vector ortonormal y , que conforman una base de \mathbb{R}^2 , de tal manera que, la reflexión se define por

$$-\alpha y + \beta x$$

Entonces la transformación lineal realiza el proceso

$$Q(\alpha y + \beta x) = -\alpha y + \beta x$$

Sea la matriz de proyección para y así, $P = yy^T$, y $Px = 0$ si $\langle x, y \rangle = 0$. La matriz de reflexión se define a partir de $Q = I - 2P$. Pues así podemos obtener que $Qy = y - 2Py = y - 2y = -y$, además $Qx = x - 2Px = x - 2(0) = x$.

Por lo tanto el operador reflector está definido por la matriz

$$Q = I - \beta yy^T, \text{ donde } \beta = \frac{2}{y^T y}$$

O, equivalentemente, donde \hat{y} , es el vector normalizado

$$Q = I - \beta \hat{y} \hat{y}^T$$

Siendo y un vector ortogonal al plano de reflexión. Por otra parte, la matriz Q es ortogonal $QQ^T = I$, simétrica $Q = Q^T$, idempotente $Q^2 = Q$ y es una involución $Q = Q^{-1}$.

Al igual que las rotaciones de Givens, el proceso de obtención de Q , es secuencial, entonces para n columnas, se tiene que $Q = Q_1 Q_2 \dots, Q_n$.

El plano de reflexión se obtiene a partir de una columna arbitraria de la matriz original, a través de la expresión

$$y = x - \|x\|e_1, \text{ con } e_1 \text{ como el vector canónico de la misma dimensión que } x.$$

Tras la obtención del valor de y , se calcula la matriz de reflexión Q . El proceso entonces es, aplicar $Q_1 A$, en la matriz resultante se observará ceros bajo el pivote de la columna escogida. Se repite el proceso utilizando la submatriz del pivote utilizado previamente, es decir, tomando una nueva columna -de una dimensión menor que la matriz original- para calcular una nueva matriz Q_2 , y realizar el producto $Q_1 Q_2 = Q$. Al finalizar con las n columnas de A , se define la matriz ortogonal final como $Q = Q_1 Q_2 \dots Q_n$ y la matriz triangular superior como $R = Q^T A$.

B. Códigos de programación

A continuación se exponen los códigos de programación para el software R, se requiere de los paquetes *pracma* y *MASS*.

B.1. Método basado en multiplicadores de Lagrange

```
Lagrange = function(A,C,b,d){  
  
  x_norm = solve(t(A)%*%A) %*% t(A)%*%b  
  x_norm + solve(t(A)%*%A) %*% t(C) %*%  
    solve( C %*% solve(t(A)%*%A) %*% t(C) ) %*% (d-C %*% x_norm)  
}
```

B.2. Método del espacio nulo

```
Nullspace = function(A,C,b,d){  
  
  stopifnot(is.numeric(A) & is.numeric(C) & is.numeric(b) & is.numeric(d))  
  
  RC = householder(t(C))$R  
  QC = householder(t(C))$Q  
  
  p = Rank(C) #Rank es del paquete pracma  
  
  Q1C = QC[,1:p]  
  Q2C = QC[, (p+1):ncol(QC)]  
  
  x1 = ginv(C) %*% d  
  
  y2 = ginv(A %*% Q2C) %*% (b-A %*% x1)
```

```
x = x1+Q2C %*% y2
```

```
x
```

```
}
```

B.3. Método de eliminación directa

```
DirElimination = function(A,C,b,d){

  stopifnot(is.numeric(A) & is.numeric(C) & is.numeric(b) & is.numeric(d))

  p = Rank(C) #Rank del paquete pracma

  C_1 = C[1:p,1:p]
  C_2 = C[1:p,(p+1):ncol(C)]

  A_1 = A[,1:p]
  A_2 = A[, (p+1):ncol(A)]

  b1 = b[1:p,]
  b2 = b[(p+1):nrow(b),]

  E = A_2 - A_1 %*% ginv(C_1) %*% C_2
  J = b - A_1 %*% ginv(C_1) %*% d

  x2 = ginv(E)%*%J
  x1 = ginv(C_1) %*% (d-C_2%*%x2)

  rbind(x1,x2)

}
```

B.4. Algoritmo para la factorización RQ de una matriz

```
RQ = function(y){

  stopifnot(is.numeric(y))
```

```

Permutar_fila = function(x){
  A = matrix(0,nrow(x),ncol(x))
  for (i in 1:nrow(x)) {
    A[nrow(x)+1-i,] = x[i,]
  }
  A
}

Permutar_columna = function(x){
  A = matrix(0,nrow(x),ncol(x))
  for (i in 1:ncol(x)) {
    A[,ncol(x)+1-i] = x[,i]
  }
  A
}

# Paso 1: Revertir las filas de la matriz
yy = Permutar_fila(y)

# Paso 2: Descomponer la matriz
Q1 = householder(t(yy))$Q
R1 = householder(t(yy))$R

# Paso 3: Permutar Q
J = Permutar_fila(t(Q1))

# Paso 4: Permutar filas y luego columnas de R
# para ser una matriz triangular superior
K_1 = Permutar_fila(t(R1))
K = Permutar_columna(K_1)

# Resultado
list("Q"=J,"R"=K)
}

```

B.5. Algoritmo GQR de una matriz

```
GQR = function(x,y){
```

```
if(nrow(x)!=nrow(y)){
  stop("Las matrices x y y deben tener el mismo número de filas.")
}

stopifnot(is.numeric(y),is.numeric(y))

Q = householder(y)$Q
S = householder(y)$R

W = RQ( t(Q) %*% x )$Q
R = RQ( t(Q) %*% x )$R

list("Sean Q,W matrices ortogonales y R,S matrices triangulares superiores,
# se cumple que  $A^T=QR$  y  $B^T=QSW$ .",
      "Q"=Q,
      "R"=R,
      "W"=W,
      "S"=S)
}
```


B.6. Algoritmo de LSE a través de GQR

```

LSE = function(A,C,b,d){

  stopifnot(is.numeric(A) & is.numeric(C) & is.numeric(b) & is.numeric(d))

  if(nrow(t(A))!=nrow(t(C))){
    stop("Las matrices A^T y C^T deben tener el mismo número de filas.")
  }

  m = ncol(t(A))
  n = nrow(t(A)) # o equivalentemente n = nrow(t(C))
  p = ncol(t(C))

  RT = t(GQR(t(A),t(C))$R)
  R11T = RT[m-n+1:p,p:p]
  R12T = RT[m-n+p+1:p,p:p]
  R22T = RT[m-n+p+1:p,n-p:n-p]

  ST = t(GQR(t(A),t(C))$S)
  S11T = ST[1:p,1:p]

  P = GQR(t(A),t(C))$W**b
  P1 = P[m-n,1]
  P2 = P[m-n+1,1]
  P3 = P[m-n+p+1,1]

  y1 = ginv(S11T)**d
  y2 = ginv(R22T)**(P3-R12T**y1)

  y = rbind(y1,y2)
  x = GQR(t(A),t(C))$Q**y
  x
}

```

B.7. Algoritmo de LSE a través de GQR para cuaterniones

La aplicación del algoritmo para la solución de mínimos cuadrados a través de cuaterniones permanece invariante en los códigos de los métodos previamente expuestos, pues como se indicó en 4-12, al transformar un cuaternión a una matriz con valores reales de orden 4, enseguida se ejecutan los algoritmos propuestos. De igual manera se desarrolla el código opuesto para transformar las matrices con valores reales a matrices cuyas columnas representan los ejes '1', 'i', 'j' y 'k'.

```

Cuaternion = function(a,b,c,d){
Q = matrix(c(a,-b,d,-c,b,a,-c,-d,-d,
            c,a,-b,c,d,b,a),4,4,byrow = T)
Q
}

```

```

Anticuaternion = function(x){

t2 = ncol(x)/4
t1 = nrow(x)/4
a=0
b=0
c=0
d=0
aq = matrix(0,t1,4*t2)

for (l in 1:t2) {

for (i in 1:t1) {
a[i] = x[4*i-3,4*l-3]
a
}
for (j in 1:t1) {
b[j] = x[(4*j-2),4*l-3]
b
}
for (j in 1:t1) {
c[j] = x[(4*j-1),4*l-2]
c
}
}
}

```

```
}
for (j in 1:t1) {
  d[j] = x[(4*j),4*1-2]
  d
}

for (j in 1:t1) {
  aq[j,(4*1-3):(4*1)] = cbind(a[j],b[j],c[j],d[j])
}

}

columnas = c("1","i","j","k")
colnames(aq) = c(rep(columnas,t2))

list("Filas de la matriz compleja"=t1, "Columnas de la matriz compleja"=t2, aq)

}
```