

Aplicación del algoritmo de transformación de Householder en mínimos cuadrados para un análisis de la pobreza en el Perú

Ayrton Fabio Coronado Huamán ¹, Guillermo Joel Borjas Córdova ², Israel Danilo Blas Salas ³, Tomás García Sifuentes ⁴

Facultad de Ciencias 1, Universidad Nacional de Ingeniería 1, e-mail:

Facultad de Ciencias 2, Universidad Nacional de Ingeniería 2, e-mail:

Palabras Claves:

Keywords:

1. Introducción

En este proyecto nos dedicaremos a construir un algoritmo eficiente, en términos de tiempo de ejecución, del método matemático "Transformación de Householder", el cual nos permite obtener un tipo de descomposición de matrices llamado "Descomposición QR". Esto nos servirá en la resolución de un sistema de ecuaciones proveniente del ajuste de una curva utilizando la técnica de mínimos cuadrados donde intervienen una variable dependiente y una independiente, y en la cual la relación entre ellas se aproxima por medio de una línea recta, todo esto con la finalidad de poder aplicarse a una situación particular en el campo de la economía, más concretamente, para hacer un análisis de la pobreza en el Perú y de esta manera hacer predicciones o pronósticos que nos ayudarán a tener un mejor entendimiento de este problema.

Con el fin de llevar a cabo este proyecto, se emplearán los datos tomados de una muestra real,

extraídos de un informe hecha por el INEI de pobreza, pobreza extrema y el coeficiente de Gini (indicador de la desigualdad económica en una población) en los años 2010 - 2013 de los principales departamentos del Perú.

Muchos autores que han hecho estudios sobre modelos de regresión, entre los que se pueden citar a: Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2001), Devore, J. L. (2005), Evans, M., & Rosenthal, J. S. (2005), Freund, J. E., & Simon, G. A. (1994), Levin, R. I., & Rubin, D. S. (2004) y Miller, I. (2000); coinciden en que siempre que se analizan datos observados o recopilados para llegar a una función o ecuación matemática que describa la relación entre las variables por medio de una regresión, se deben enfrentar tres problemas:

1. Decidir qué clase de curva muestran los puntos y por tanto qué clase de ecuación se debe usar.
2. Encontrar la ecuación particular que mejor se ajuste a los datos.
3. Demostrar que la ecuación particular encontrada cumple con ciertos aspectos referentes a los méritos de ésta para hacer pronósticos.

Para decidir qué clase de función podría ajustarse a la curva, debe hacerse una gráfica de dispersión de los datos observados. Si en dicha gráfica se aprecia que los puntos se distribuyen alrededor de una recta, se procede a realizar un análisis de regresión lineal.

2. Conceptos Previos

2.1. SISTEMAS LINEALES DE ECUACIONES SOBREDETERMINADOS

El problema lineal de mínimos cuadrados se acerca a la optimización si se encuentra un vector \mathbf{x} que

en algún sentido da la mejor (aunque no la perfecta) aproximación entre \mathbf{Ax} y \mathbf{b} .

Considérese un sistema de m ecuaciones con n incógnitas escrito en la forma

$$\mathbf{Ax} = \mathbf{b} \quad (\text{II.1})$$

Aquí, \mathbf{A} es de $m \times n$, \mathbf{x} es de $n \times 1$ y \mathbf{b} es de $m \times 1$. Supondremos que el rango de \mathbf{A} es n , de lo que resulta que $m \geq n$. Si $m > n$, se dice que el sistema lineal de ecuaciones es sobredeterminado. Los sistemas sobredeterminados son generalmente no compatibles.

2.2. EL VECTOR RESIDUAL Y EL PROBLEMA DE LOS MÍNIMOS CUADRADOS

Por lo general, el sistema (II.1) no tiene solución debido a que \mathbf{b} no pertenece al subespacio de \mathbf{R}^m de dimensión n , generado por las columnas de \mathbf{A} . Es frecuente en tales casos que se requiera encontrar un \mathbf{x} tal que \mathbf{b} esté cerca de \mathbf{Ax} , es decir que minimice una norma del vector residual $\mathbf{r} = \mathbf{b} - \mathbf{Ax}$. El primer problema será determinar el sentido de cercanía, es decir la norma que mide la diferencia entre \mathbf{b} y \mathbf{Ax} . La interpretación y la naturaleza de la solución de este problema varía dependiendo de la norma que se use. La norma más comúnmente usada es la norma euclidiana, la norma l_2 . Entonces, la solución en mínimos cuadrados de (II.1) es el vector \mathbf{x} que hace de $\|\mathbf{r}\|_2 = \|\mathbf{b} - \mathbf{Ax}\|_2$ un mínimo

$$\min_{\mathbf{x} \in \mathbf{R}^n} \|\mathbf{b} - \mathbf{Ax}\|_2^2, \quad (\text{II.2})$$

Dado que las normas no pueden ser negativas, minimizar una norma es equivalente a minimizar su cuadrado; hablaremos entonces libremente de normas cuadradas y no. Según lo que se ha supuesto acerca del rango de \mathbf{A} , la solución \mathbf{x} será única. La formulación de los mínimos cuadrados es popular sobre todo por el hecho de que resolver un problema

de mínimos cuadrados lineal es más fácil que minimizar otras normas de \mathbf{r} . Veremos que el problema de mínimos cuadrados se puede resolver transformando la matriz \mathbf{A} y resolviendo un sistema compatible a ella relacionado. Al contrario, minimizar la norma l_1 o la norma l_∞ es equivalente a un problema de programación lineal no diferenciable, cuya solución necesita una técnica de iteración.

Es útil pensar en el problema de los mínimos cuadrados en términos de los subespacios definidos por la matriz \mathbf{A} . El subespacio rango de \mathbf{A} , $\text{rango}(\mathbf{A})$, consiste en todos los vectores m -dimensionales que son combinaciones lineales de las columnas de \mathbf{A} y el subespacio complementario, el subespacio nulo de \mathbf{A}^t , $\text{nulo}(\mathbf{A})$, contiene todos los vectores m -dimensionales que son \mathbf{z} -ortogonales a las columnas de \mathbf{A} , es decir tales que $\mathbf{A}^t \mathbf{z} = \mathbf{0}$. Si r denota el rango de la matriz \mathbf{A} , entonces r es la dimensión del subespacio rango de \mathbf{A} y $(m - r)$ es la dimensión del subespacio nulo de \mathbf{A}^t . Dada una matriz no nula \mathbf{A} , cualquier vector m -dimensional \mathbf{c} se puede expresar como la suma de un vector \mathbf{c}_R en el rango de \mathbf{A} , $\mathbf{c}_R \in \text{rango}(\mathbf{A})$, y de un vector \mathbf{c}_N en el espacio nulo de \mathbf{A}^t , $\mathbf{c}_N \in \text{nulo}(\mathbf{A}^t)$, es decir

$$\mathbf{c} = \mathbf{c}_R + \mathbf{c}_N. \quad (\text{II.3})$$

Los vectores \mathbf{c}_R y \mathbf{c}_N son únicos y satisfacen

$$\mathbf{c}_R = \mathbf{Ac}_A, \mathbf{A}^t \mathbf{c}_N = \mathbf{0}, \mathbf{c}_R^t \cdot \mathbf{c}_N = 0, \quad (\text{II.4})$$

donde \mathbf{c}_A se refiere a un vector n -dimensional cualquiera tal que $\mathbf{c}_R = \mathbf{Ac}_A$. La unicidad de \mathbf{c}_R y \mathbf{c}_N implica que las componentes en el espacio rango y en el espacio nulo de vectores iguales tienen que ser iguales. En efecto,

$$\mathbf{c} = \mathbf{d} \text{ si y solo si } \mathbf{c}_R = \mathbf{d}_R \text{ y } \mathbf{c}_N = \mathbf{d}_N. \quad (\text{II.5})$$

Aunque \mathbf{c}_R es único, el vector \mathbf{c}_A es único solo si las columnas de \mathbf{A} son linealmente independientes. Debido al hecho de que la norma euclídea está definida en términos del producto escalar, las relaciones

(II.3) y (II.4) tienen una consecuencia muy importante:

$$\|c\|_2^2 = c^t \cdot c = \|c_R\|_2^2 + \|c_N\|_2^2, \quad (\text{II.6})$$

de manera que la representación de un vector en términos de sus componentes en el espacio rango y nulo separa el cuadrado de la norma euclídea en dos partes independientes. Esta propiedad no vale para otras normas.

Estas observaciones son relevantes para el problema de los mínimos cuadrados dado que permiten determinar la más pequeña posible norma euclídea del vector residual $r = b - Ax$ para todos vectores x . Dado que ambos b y el vector residual r son vectores m -dimensionales, ambos pueden escribirse en la representación (II.3):

$$b = b_R + b_N \text{ con } b_R = Ab_A, \quad (\text{II.7})$$

$$r = r_R + r_N \text{ con } r_R = Ar_A$$

Combinando la definición de r como $b - Ax$ con estas relaciones, se obtiene

$$r = r_R + r_N = b - Ax = b_R + b_N - Ax = b_R - Ax + b_N \quad (\text{II.8})$$

Un punto obvio pero crucial en esta expresión es que el vector Ax está enteramente en el rango de A . Cuando se resta Ax del vector b para crear el residual, sigue de (II.5) que la componente en el espacio rango de $b - Ax$ tiene que ser $b_R - Ax$. En contraste, el restar Ax no puede eliminar nada a la componente de b en el espacio nulo, de manera que la componente en el espacio nulo de $b - Ax$ es igual a b_N . Entonces, las componentes en el espacio rango y en el nulo del vector residual satisfacen

$$r_R = b_R - Ax \text{ y } r_N = b_N. \quad (\text{II.9})$$

Como ya dicho, el problema de mínimos cuadrados consiste en minimizar la norma l_2 del vector residual. De (II.6) y (II.9) sigue que la norma euclídea del vector residual r para cualquier vector x satisface

$$\|r\|_2^2 = \|b_R - Ax\|_2^2 + \|b_N\|_2^2 \geq \|b_N\|_2^2. \quad (\text{II.10})$$

Dado que b_N se mantiene en el residual, se encontrará el mínimo de $\|b - Ax\|_2$ cuando la norma de la componente en el subespacio rango, $\|b_R - Ax\|_2$, sea la más pequeña posible.

Además, por definición b_R está en el rango de A , luego tiene que existir un vector x tal que $Ax = b_R$. Para este vector particular x , se elimina la componente de b en el espacio rango cuando se le resta Ax , que significa que $b - Ax = b_N$ y que la norma euclídea del vector residual es igual a su cota inferior $\|b_N\|_2$.

Escoger x de esta manera, $Ax = b_R$, no solo minimiza la norma del vector residual, sino que además fuerza al residual a estar enteramente en el espacio nulo de A^t . De esta manera se llega a dos caracterizaciones equivalentes de la solución optimal del problema de mínimos cuadrados:

$$x \text{ minimiza } \|b - Ax\|_2 \iff A^t(b - Ax) = 0, \quad (\text{II.11})$$

$$x \text{ minimiza } \|b - Ax\|_2 \iff \begin{cases} Ax = b_R, \\ b - Ax = b_N. \end{cases} \quad (\text{II.12})$$

Un problema de mínimos cuadrados se dice compatible si su vector residual optimal es cero, en este caso $b_N = 0$ y $b = b_R$; si éste no es el caso se dice incompatible. La ortogonalidad del vector residual respecto de las filas de la matriz A corresponde al principio geométrico familiar de que la distancia más corta de un punto a un plano es la longitud de la perpendicular que los une. Aquí el punto es el vector b y el plano es el subespacio rango de A .

La unicidad del vector b_N implica que el vector residual optimal para el problema de mínimos cuadrados es único. El vector b_R es también único, y el sistema $Ax = b_R$ es compatible por definición. Sin embargo, el vector x que satisface $Ax = b_R$ es único si y solo si las columnas de la matriz A son linealmente independientes. Dado que cualquier vector x que

satisface $\mathbf{Ax} = \mathbf{b}_R$ es una solución de mínimos cuadrados, sigue una importante conclusión: la solución del problema de mínimos cuadrados es única si y solo si la matriz \mathbf{A} es de rango n .

Un aspecto crucial de la relación (II.12) es que revela cómo cambios en el vector \mathbf{b} afectan a la solución \mathbf{x} y al problema de mínimos cuadrados $\min \|\mathbf{b} - \mathbf{Ax}\|_2^2$. Para sistemas de ecuaciones lineales no singulares y compatibles, cambios en \mathbf{b} implicaban encontrar una solución \mathbf{x} diferente. Esta propiedad no se verifica para el problema de los mínimos cuadrados, cuyo análisis es más complicado.

Se supone por simplicidad que \mathbf{A} tiene rango lleno, entonces \mathbf{x} es la única solución del problema de mínimos cuadrados $\min \|\mathbf{b} - \mathbf{Ax}\|_2^2$ y \mathbf{r} es el vector residual optimal, $\mathbf{r} = \mathbf{b} - \mathbf{Ax}$. Si \mathbf{b} viene perturbado por un vector que está enteramente en el subespacio rango de \mathbf{A} , es decir $\tilde{\mathbf{b}} = \mathbf{b} + \delta\mathbf{b}_R$, donde $\delta\mathbf{b}_R = \mathbf{A}\delta\mathbf{x}$ para un único n -vector $\delta\mathbf{x}$, entonces

$$\tilde{\mathbf{b}}_R = \mathbf{b}_R + \delta\mathbf{b}_R = \mathbf{b}_R + \mathbf{A}\delta\mathbf{x} \text{ y } \tilde{\mathbf{b}}_N = \mathbf{b}_N. \quad (\text{II.13})$$

Sigue entonces de (II.12) que la solución de mínimos cuadrados $\tilde{\mathbf{x}}$ y el residual $\tilde{\mathbf{r}}$ correspondientes a $\tilde{\mathbf{b}}$ satisfacen

$$\tilde{\mathbf{x}} = \mathbf{x} + \delta\mathbf{x} \text{ y } \tilde{\mathbf{r}} = \mathbf{r}, \quad (\text{II.14})$$

mostrando que cambios en \mathbf{b} en el espacio rango modifican el vector solución \mathbf{x} pero no el vector residual \mathbf{r} . Por el contrario, si \mathbf{b} viene perturbado por un vector que está enteramente en el subespacio nulo de \mathbf{A}^t , es decir $\mathbf{b} = \mathbf{b} + \mathbf{z}$, donde $\mathbf{A}^t\mathbf{z} = \mathbf{0}$, entonces

$$\mathbf{x} = \mathbf{x} \text{ y } \mathbf{r} = \mathbf{r} + \mathbf{z}. \quad (\text{II.15})$$

mostrando que cambios en \mathbf{b} en el espacio nulo modifican el vector residual \mathbf{r} , mientras que el vector solución \mathbf{x} permanece inalterado.

2.3. LAS ECUACIONES NORMALES

La propiedad más significativa del vector residual optimal del problema de mínimos cuadrados es que está enteramente en el espacio nulo de la matriz \mathbf{A}^t . En términos algebraicos, una solución \mathbf{x} satisface

$$\mathbf{A}^t(\mathbf{b} - \mathbf{Ax}) = \mathbf{0}, \quad (\text{II.16})$$

que es equivalente a

$$\mathbf{A}^t\mathbf{Ax} = \mathbf{A}^t\mathbf{b}. \quad (\text{II.17})$$

Este sistema de ecuaciones se conoce como el de las ecuaciones normales.

La matriz simétrica $\mathbf{A}^t\mathbf{A}$, conocida como la matriz de la ecuación normal, es semidefinida positiva para una matriz cualquiera \mathbf{A} , y es definida positiva si y solo si las columnas de \mathbf{A} son linealmente independientes. Las ecuaciones normales son siempre compatibles, es decir existe siempre una solución de (II.17) aunque $\mathbf{A}^t\mathbf{A}$ sea singular. Cualquier solución \mathbf{x} tiene que satisfacer las ecuaciones normales, y cualquier vector \mathbf{x} que satisface las ecuaciones normales tiene que ser una solución de mínimos cuadrados.

Las ecuaciones normales son de gran importancia práctica por el hecho de que ofrecen una manera directa para calcular la solución de mínimos cuadrados. Si $\mathbf{A}^t\mathbf{A}$ es definida positiva, es decir si \mathbf{A} tiene rango lleno, las ecuaciones normales (II.17) tienen solución única, que se puede encontrar usando el algoritmo de factorización de Cholesky. Si $\mathbf{A}^t\mathbf{A}$ es singular, se puede hallar una solución de las ecuaciones normales usando una versión de la factorización de Cholesky que incluya un intercambio simétrico. Entonces, el problema de mínimos cuadrados se puede resolver siguiendo los siguientes pasos:

- (I) Formar la matriz de la ecuación normal $\mathbf{A}^t\mathbf{A}$ y el vector $\mathbf{A}^t\mathbf{b}$;
- (II) Calcular la factorización de Cholesky $\mathbf{A}^t\mathbf{A} = \mathbf{L}^t\mathbf{L}$, con \mathbf{L} triangular superior;

- (III) Resolver los dos sistemas $L^t y = A^t b$ y $Lx = y$. El vector x es la solución deseada.

Desde el punto de vista computacional, resolver el problema de mínimos cuadrados con las ecuaciones normales es eficiente: para formar $A^t A$ y $A^t b$ se necesitan aproximadamente $m \cdot n^2/2$ operaciones, calcular la factorización de Cholesky usa del orden de $n^3/6$ operaciones y resolver los dos sistemas triangulares implica aproximadamente n^2 operaciones. Desafortunadamente, el uso de las ecuaciones normales presenta el problema de la estabilidad numérica, dado que el número de condición de la matriz $A^t A$ es el cuadrado del número de condición de A . Consecuentemente, la matriz de la ecuación normal está seriamente mal-condicionada si la matriz A misma está ligeramente mal-condicionada.

El mal condicionamiento de la matriz de la ecuación normal asociada puede conducir no solo a una aproximación no precisa de la solución calculada de las ecuaciones normales, sino también a una pérdida de información cuando el rango numérico de A es marginal.

2.4. APLICACIONES

Al igual que en el problema del análisis de pobreza, a menudo coleccionamos datos e intentamos encontrar una relación funcional entre las variables. Si los datos son $n + 1$ puntos del plano, es posible encontrar un polinomio de grado n o inferior que pasa por todos los puntos. Este polinomio se llama polinomio de interpolación. Dado que los datos tienen en general error experimental, no hay razón para pedir que las funciones pasen por todos los puntos. De hecho, polinomios de grado inferior que no pasan por los puntos de manera exacta, dan una mejor descripción de la relación entre variables. Por ejemplo, si la relación entre variables es actualmente lineal y los datos tienen un pequeño error, sería desastroso usar un polinomio de interpolación.

Dada una tabla de datos

x	x_1	x_2	\dots	x_m
y	y_1	y_2	\dots	x_m

(II.18)

deseamos encontrar una función lineal

$$y = c_0 + c_1 x \quad (\text{II.19})$$

que mejor aproxima los datos en el sentido de mínimos cuadrados. Si se pide que

$$y_i = c_0 + c_1 x_i \text{ para } i = 1, 2, \dots, m \quad (\text{II.20})$$

obtenemos un sistema de m ecuaciones en dos incógnitas

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \quad (\text{II.21})$$

La función lineal cuyos coeficientes son la solución de mínimos cuadrados de (II.21) viene a ser la aproximación de mínimos cuadrados a los datos con una función lineal.

Si los datos no aparecen en relación lineal, se podría usar un polinomio de grado mayor. Es decir, para encontrar los coeficientes c_0, c_1, \dots, c_n de la mejor aproximación por mínimos cuadrados a los datos

x	x_1	x_2	\dots	x_m
y	y_1	y_2	\dots	x_m

(II.18)

con un polinomio de grado n , tenemos que encontrar la solución de mínimos cuadrados al sistema

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^n \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \quad (\text{II.22})$$

2.5. LAS TRANSFORMACIONES DE HOUSEHOLDER

Debido a las posibles dificultades numéricas del uso de las ecuaciones normales, los modernos métodos de mínimos cuadrados se han desarrollado basándose en las transformaciones ortogonales, que preservan las distancias euclídeas y no empeoran las condiciones de la matriz A . La idea es transformar un problema de mínimos cuadrados de manera que sea fácil de resolver, reduciendo la matriz A a la forma que revela el rango. El término forma triangular que revela el rango una matriz genérica $m \times n$ de rango r correspondiente a la matriz $\tilde{T} = \begin{pmatrix} T \\ 0 \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ 0 & 0 \end{pmatrix}$, con T_{11} matriz $r \times r$ no singular triangular superior y T_{12} matriz $r \times (n - r)$. Cuando \tilde{T} tiene rango lleno de fila entonces no aparecen ambos bloques de ceros; si T' es de rango lleno de columna, la matriz T_{12} y el bloque de ceros derecho no aparecerán. Mas en general, el rango de una matriz $m \times n$, \tilde{F} , es r si \tilde{F} es de la forma $\tilde{F} = \begin{pmatrix} F \\ 0 \end{pmatrix}$, con F matriz $r \times n$ y las filas de F son linealmente independientes.

Dado que ya no estamos resolviendo ecuaciones, el conjunto de transformaciones que se puede aplicar a A sin cambiar la solución está restringido. Las transformaciones ortogonales son obvias en este contexto dado que estas no alteran la norma l_2 . Sea A una matriz no nula $m \times n$, con rango igual a r . Supóngase que se pueda encontrar una matriz $m \times m$ ortogonal Q que produzca la forma que revela el rango

$$QA = \begin{pmatrix} F \\ 0 \end{pmatrix}, \quad (\text{III.1})$$

donde F es una matriz $r \times n$ y tiene las filas linealmente independientes; el bloque de ceros no aparece si $r = m$. La forma (III.1) nos permite resolver el problema de mínimos cuadrados usando las matrices Q y F . Sea d el vector transformado $Q^t b$, descom-

puesto en

$$d = Q^t b = \begin{pmatrix} d_r \\ d_{r-m} \end{pmatrix}, \quad (\text{III.1})$$

Usando esta definición y la estructura mostrada en (III.1), se puede escribir el vector residual transformado como

$$Q^t(b - Ax) = \begin{pmatrix} d_r \\ d_{r-m} \end{pmatrix} - \begin{pmatrix} Fx \\ 0 \end{pmatrix} = \begin{pmatrix} d_r - Fx \\ d_{m-r} \end{pmatrix}. \quad (\text{III.3})$$

Dado que la matriz Q^t es ortogonal, su aplicación al vector residual no altera la norma euclídea, y

$$\|b - Ax\|_2^2 = \|Q^t(b - Ax)\|_2^2. \quad (\text{III.4})$$

Combinando esta relación con la forma especial del vector residual transformado, se concluye que

$$\|b - Ax\|_2^2 = \|Q^t(b - Ax)\|_2^2 = \|d_r - Fx\|_2^2 + \|d_{m-r}\|_2^2 \geq \|d_{m-r}\|_2^2. \quad (\text{III.5})$$

que significa que $\|b - Ax\|_2$ no puede ser menor que $\|d_{m-r}\|_2$ para cualquier vector x . El valor más pequeño posible para $\|b - Ax\|_2^2$ es la cota inferior. Igualdades con la cota inferior se obtienen si y sólo si el vector x satisface

$$Fx = d_r. \quad (\text{III.6})$$

Dado que el rango de la matriz F es igual a r , el sistema $Fx = d_r$ tiene que ser compatible. Cuando $Fx = d_r$, el vector residual transformado satisface $Q^t(b - Ax) = \begin{pmatrix} 0 \\ d_{r-m} \end{pmatrix}$ y $\|b - Ax\|_2 = \|d_{m-r}\|_2$. Esto demuestra que cualquier vector x que satisfaga $Fx = d_r$ será una solución de mínimos cuadrados. Suponiendo que los sistemas en los cuales aparecen las matrices F son fáciles de resolver, se sigue que el problema de mínimos cuadrados se puede resolver encontrando una matriz ortogonal Q_t que nos dé la forma (III.1).

Antes de presentar la factorización QR para resolver el problema de mínimos cuadrados, es necesario introducir las transformaciones de Householder.

La técnica más popular para construir una matriz ortogonal que reduzca la matriz \mathbf{A} en forma triangular usa una clase especial de matrices que son simultáneamente simétricas, elementales y ortogonales. Para cualquier vector no nulo \mathbf{u} , la correspondiente transformación de Householder (o matriz de Householder, o reflector de Householder) es una matriz de la forma

$$\mathbf{H} = \mathbf{H}(\mathbf{u}) = \mathbf{I} - \frac{2\mathbf{u}\mathbf{u}^t}{\|\mathbf{u}\|_2^2} = \mathbf{I} - \frac{\mathbf{u}\mathbf{u}^t}{\beta}, \text{ con} \\ \beta = \frac{\|\mathbf{u}\|_2^2}{2}, \quad (\text{III.7})$$

donde el vector \mathbf{u} es el vector de Householder

Teorema(III.1) Si \mathbf{H} es la matriz definida en (III.7), entonces

- 1) $\mathbf{H} = \mathbf{H}^t$,
- 2) $\mathbf{H} = \mathbf{H}^{-1}$,

que es lo mismo que decir que la matriz \mathbf{H} es simétrica y ortogonal.

Las matrices de Householder son simétricas y ortogonales, y dependen sólo de la dirección del vector \mathbf{u} . En el contexto de las reducciones a matrices triangulares, las matrices de Householder poseen dos propiedades cruciales:

- Para cualquier par de vectores distintos de igual norma $\|\mathbf{a}\|_2 = \|\mathbf{b}\|_2$, existe una transformación de Householder que transforma el uno en el otro,

$$\mathbf{H}\mathbf{a} = (\mathbf{I} - \frac{\mathbf{u}\mathbf{u}^t}{\beta})\mathbf{a} = \mathbf{b}, \quad (\text{III.8})$$

con $\|\mathbf{a}\|_2 = \|\mathbf{b}\|_2$. Esto implica que el vector \mathbf{u} tiene que satisfacer la condición

$$-\frac{\mathbf{u}^t\mathbf{a}}{\beta}\mathbf{u} = \mathbf{b} - \mathbf{a}, \quad (\text{III.9})$$

es decir, \mathbf{u} es un múltiplo de $\mathbf{b} - \mathbf{a}$; - cualquier vector \mathbf{c} transformado por una matriz de Householder posee una forma especial:

$$\mathbf{H}\mathbf{c} = (\mathbf{I} - \frac{\mathbf{u}^t\mathbf{c}}{\beta})\mathbf{c} = \mathbf{c} - \frac{\mathbf{u}^t\mathbf{c}}{\beta}\mathbf{u}, \quad (\text{III.10})$$

de manera que $\mathbf{H}\mathbf{c}$ es la diferencia entre el vector original \mathbf{c} y un múltiplo especial del vector de Householder \mathbf{u} . Claramente el vector \mathbf{c} no varía si $\mathbf{u}^t\mathbf{c} = 0$. Además, calcular el producto $\mathbf{H}\mathbf{c}$ no necesita los elementos explícitos de \mathbf{H} , sino sólo el vector \mathbf{u} y el escalar β .

2.6. LA FACTORIZACIÓN QR

Para una matriz genérica \mathbf{A} , $n \times n$, no singular, las propiedades que se acababan de describir permiten construir una sucesión de $n - 1$ matrices de Householder tales que

$$\mathbf{H}_{n-1} \dots \mathbf{H}_2 \mathbf{H}_1 \mathbf{A} = \mathbf{R}, \quad (\text{III.11})$$

donde \mathbf{R} es una matriz $n \times n$ no singular y triangular superior.

El primer paso de este proceso es construir una matriz de Householder \mathbf{H}_1 que transforma \mathbf{a}_1 (la primera columna de \mathbf{A}) en un múltiplo de \mathbf{e}_1 , es decir, se desean crear ceros en las componentes 2 hasta n del vector \mathbf{a}_1 . La norma euclídea se conserva bajo transformaciones ortogonales, de manera que

$$\mathbf{H}_1\mathbf{a}_1 = (\mathbf{I} - \frac{\mathbf{u}_1\mathbf{u}_1^t}{\beta_1})\mathbf{a}_1 = \pm\|\mathbf{a}_1\|_2\mathbf{e}_1 = \begin{pmatrix} r_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad (\text{III.12})$$

donde $|r_{11}| = \|\mathbf{a}_1\|_2$. De la expresión (III.9) sabemos que el vector \mathbf{u}_1 tiene que ser un múltiplo de $\|\mathbf{a}_1\|_2\mathbf{e}_1 - \mathbf{a}_1$, y dado que \mathbf{H}_1 depende sólo de la dirección de \mathbf{u}_1 , podemos elegir el vector \mathbf{u}_1 como

$$\mathbf{u}_1 = \begin{pmatrix} \mathbf{a}_{11} - r_{11} \\ \mathbf{a}_{21} \\ \vdots \\ \mathbf{a}_{n1} \end{pmatrix}. \quad (\text{III.13})$$

Al definir \mathbf{u}_1 , el signo de r_{11} se puede elegir positivo o negativo (excepto cuando \mathbf{a}_1 es ya un múltiplo de \mathbf{e}_1), y para evitar el problema de la cancelación

de términos parecidos, usualmente se escoge el signo opuesto al signo de a_{11} , de manera que

$$\text{signo}(r_{11}) = -\text{signo}(a_{11}). \quad (\text{III.14})$$

Después de la primera aplicación de las transformaciones de Householder, la primera columna de la matriz parcialmente reducida $A^{(2)} = H_1 A$ es un múltiplo de e_1 , y los demás elementos han sido alterados

$$A^{(2)} = H_1 A = \begin{pmatrix} r_{11} & a_{12}^{(2)} & \cdots & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix}. \quad (\text{III.15})$$

En muy importante notar que a diferencia de la eliminación Gaussiana, la primera fila de la matriz A viene modificada por efecto de la transformación de Householder H .

Al construir la segunda transformación de Householder, el objetivo principal es reducir la segunda columna a la forma correcta, sin alterar la primera fila y la primera columna de la matriz parcialmente reducida. Debido a la propiedad (III.10), se puede obtener este resultado definiendo el segundo vector de Householder u_2 con la primera componente nula. Habiendo escogido así el vector u_2 , la aplicación de la matriz de Householder H_2 a un vector genérico no cambia la primera componente, y la aplicación a un múltiplo de e_1 deja el vector entero como está.

Si A es una matriz no singular, se pueden efectuar $n-1$ pasos de reducción de Householder, para obtener $H_{n-1} \dots H_2 H_1 A = R$, con R una matriz triangular superior no singular. Si se denota con Q^t la matriz ortogonal $n \times n$

$$Q^t = H_{n-1} \dots H_2 H_1 \implies Q = H_1 H_2 \dots H_{n-1}. \quad (\text{III.16})$$

Cualquiera de las dos formas

$$Q^t A = R \quad \text{o} \quad A = QR. \quad (\text{III.17})$$

se conoce como la factorización QR de la matriz A . Una vez conocida la factorización QR de A , ecuación (III.17), la solución al sistema $Ax = b$ se obtiene resolviendo el sistema triangular superior $Rx = Q^t b$.

En general, es necesario un intercambio de columnas para asegurar que el rango está plenamente revelado. Sin el intercambio de columnas la reducción de Householder terminaría inmediatamente con una matriz no nula cuya primera columna es cero. Si las demás columnas son también ceros, la reducción termina. De otra manera, existe por lo menos una columna no nula, la columna pivote, que se puede elegir como candidata para la siguiente reducción. Como en la eliminación gaussiana, una estrategia de pivoteo pide que se escoja la columna pivote como la primera columna de norma máxima (otra posibilidad es escoger la columna “menos reducida”).

En general, si A es $m \times n$ de rango r , se necesitarán r permutaciones P_k y r matrices de Householder H_k , $k = 1, \dots, r$. Después de estos r pasos, la configuración final será

$$H_r \dots H_1 A P_1 \dots P_r = \tilde{R}, \quad (\text{III.18})$$

donde

$$\tilde{R} = \begin{pmatrix} R \\ 0 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} \\ 0 & 0 \end{pmatrix}, \quad (\text{III.19})$$

es triangular superior que nos revela el rango, y R_{11} es una matriz $r \times r$ no singular triangular superior. Con una correcta estrategia de pivoteo, dependiente del problema original, y aritmética exacta, este procedimiento de Householder terminará después de r pasos, cuando la matriz restante se transformará en cero. Combinando los intercambios de columnas en una sola matriz de permutación P y las transformaciones de Householder en una sola matriz ortogonal Q^t , se obtiene

$$Q^t A P = \tilde{R}, \quad (\text{III.20})$$

o equivalentemente,

$$\mathbf{A}\mathbf{P} = \mathbf{Q}\tilde{\mathbf{R}}, \quad (\text{III.21})$$

y

$$\mathbf{Q}^t \mathbf{A} = \tilde{\mathbf{R}} \mathbf{P}^t = \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix} \mathbf{P}^t = \begin{pmatrix} \mathbf{R} \mathbf{P}^t \\ \mathbf{0} \end{pmatrix}. \quad (\text{III.22})$$

Las filas de la matriz $\mathbf{R} \mathbf{P}^t$ son linealmente independientes, de manera que la matriz transformada $\mathbf{Q}^t \mathbf{A}$ tiene la forma deseada (III.1), con $\mathbf{F} = \mathbf{R} \mathbf{P}^t$. El problema de mínimos cuadrados se puede entonces resolver con el siguiente algoritmo:

- (I) Calcular la factorización QR de la matriz \mathbf{A} , usando la reducción de Householder (que determina el rango r , y las matrices \mathbf{P} , \mathbf{Q} y \mathbf{R});
- (II) Formar el vector $\mathbf{d} = \mathbf{Q}^t \mathbf{b}$, y denotar con \mathbf{d}_r las primeras r componentes de \mathbf{d} ;
- (III) Calcular cualquier solución \mathbf{y} del sistema $\mathbf{R}_y = \mathbf{d}_r$;
- (IV) Formar $\mathbf{x} = \mathbf{P} \mathbf{y}$.

El número de operaciones requerido para resolver el problema de mínimos cuadrados de esta manera es del orden de $2 \cdot m \cdot n \cdot r - r^2(m + n) + \frac{2}{3}n^3$. Si la matriz \mathbf{A} posee columnas linealmente independientes ($r = n$), la matriz \mathbf{R} es no singular, la solución de $\mathbf{R} \mathbf{y} = \mathbf{d}_r$ es única, y la solución del problema de mínimos cuadrados es única. En este caso resolver el problema con el método QR es aproximadamente dos veces más costoso que resolviéndolo con las ecuaciones normales. Cuando las columnas de \mathbf{A} son linealmente dependientes, de manera que $r < n$, el sistema $r \times n$ $\mathbf{R} \mathbf{y} = \mathbf{d}_r$ posee un número infinito de soluciones. Dado que \mathbf{R} es de la forma $\mathbf{R} = (\mathbf{R}_{11} \mathbf{R}_{12})$, con \mathbf{R}_{11} triangular superior no singular, es posible escoger un vector $\mathbf{y} = (\mathbf{y}_B, \mathbf{0})^t$ tal que $\mathbf{R} \mathbf{y} = \mathbf{d}_r$ y con la propiedad de que $\mathbf{R}_{11} \mathbf{y}_B = \mathbf{d}_r$. Si \mathbf{y} posee esta forma, la solución $\mathbf{x} = \mathbf{P} \mathbf{y}$ es sencillamente una reordenación de

\mathbf{y} , y es llamada una solución básica del problema de mínimos cuadrados.

3. ANÁLISIS

4. OBSERVACIONES

5. CONCLUSIONES

-
1. I.K. Argyros, *Newton-like methods under mild differentiability conditions with error analysis*, Bull. Austral. Math. Soc. **37** (1988), 131-147.
 - 2.