

**T.C.
İstanbul Üniversitesi
Sosyal Bilimler Enstitüsü
Ekonometri Anabilim Dalı**

Yüksek Lisans Tezi

Matris Ayrışımı

Li Zhaoyang

2501010543

Tez Danışmanı: Yrd. Doç. Dr. Enis Sınıksaran

İstanbul 2006

ÖZ

Matris ayrışımı karmaşık bir matrisi daha basit matrislerin çarpımına dönüştüren bir yöntemdir. Örneğin Tekil değer Ayrışımı (Singular Value Decomposition – SVD), bir matrisi ortonormal bir matris, köşegen bir matris ve ortonormal bir matris olmak üzerine üçlü bir çarpıma ayıştıran bir algoritmadır. 1960'lı yıllardan önce sadece lineer sistem analizine uygulanmış olan matris ayrışım, son yıllarda yazılım, elektronik sinyal filtrelemesi, matris transformasyonu ve regresyon analizi gibi alanlarda da kullanılmaktadır. Tezde matris ayrışımında en önemli rolü oynayan SVD, QR ve LU 'nun algoritmaları ve uygulama alanları tanıtılmaktadır. Çalışmanın temel amacı ise SVD 'nin regresyon analizinde oynadığı rolün ayrıntılı olarak tartışılmasıdır. Tezdeki ispat ve uygulamalarda sadece reel sayılar kullanılmaktadır.

ABSTRACT

Matrix factorization is the algorithm factorizing a matrix into the product of several matrices with particular properties. For instance, with SVD algorithm, a matrix can be factorized into the product of an orthonormal, a diagonal and another orthonormal matrix, that magically facilitates our analysis involving this matrix. Before 1960's, matrix factorization was only used in the linear system analysis, but in the last few decades the quickly developed algorithms of matrix factorizations have been applied to solve a variety of problems, like the regression analysis and information technologies. In this thesis, we are concerned with the theoretical derivation of SVD, QR and LU decompositions of matrixes along with their applications in the regression analysis. Here the primary goal we aspire to meet is to present the magician roles of SVD in the regression analysis. And all discussions in this thesis are confined to the real number realm.

ÖNSÖZ

Sayısal analizin (Numerical Analysis) en önemli amaçlarından biri, karmaşık bir matrisin yerine, bu matrisin özelliklerini taşıyan daha basit bir matrisin incelenmesidir. Bu amacı yerine getiren en önemli yöntem ise, matris ayrışımıdır.

Çalışmada, matris ayrışımını en iyi temsil eden üç algoritma – SVD, QR, LU ve bunların regresyon analizindeki uygulamaları incelenmektedir.

Tezimin yazılma sürecinde bana yardım eden ve beni daima teşvik eden değerli hocam Yrd. Doç. Dr. Enis Sınıksaran 'a çok teşekkür ederim. Özellikle Yrd. Doç. Dr. Aylin Aktükün 'ün bir kaç kez titizce tezin başından sonuna kadar yaptığı düzeltmeleri için de çok teşekkür ederim. Ayrıca tezimdeki Türkçe ifadelerin düzeltilmesinde yardımda bulunan Dr. Ferda Yerdelen 'e, Araş. Gör. Leyla İşbilen 'e ve arkadaşım Mansur Ocakoğlu 'na da çok teşekkür ederim.

İÇİNDEKİLER

ÖZ	iii
ÖNSÖZ	iv
İÇİNDEKİLER	v
TABLolar ve ŞEKİLLER LİSTESİ	vii
KISALTMALAR ve SİMGELER LİSTESİ	ix
GİRİŞ	1
1. BÖLÜM TEKİL DEĞER AYRIŞIMI (SVD)	3
1.1. Tekil Değer Ayrışımı Kavramı	3
1.1.1 Öz Değer Ayrışımı	3
1.1.2 Tekil Değer Ayrışımı	7
1.2. Regresyon Analizine SVD Uygulamaları	14
1.2.1. Çoklu Doğrusal Bağlantının SVD ile İncelenmesi ve Çözülmesi	14
1.2.1.1. Tam Olmayan Çoklu Doğrusal Bağlantının SVD ile İncelenmesi ve Çözülmesi	15
1.2.1.2. Tam Çoklu Doğrusal Bağlantının SVD ile İncelenmesi ve Çözülmesi	25
1.2.2. SVD ile Elde Edilen Genelleştirilmiş Ters Matris Üzerine Parametre Tahminlerinin İncelenmesi	29
1.2.2.1. Genelleştirilmiş Ters Matris ve Moore – Penrose Ters Matrisinin SVD ile Açıklanması	29
1.2.2.2. Genelleştirilmiş Ters Matris ve Moore - Penrose Ters Matrisi ile Parametre Tahminlerinin Açıklanması	33
1.2.3. SVD ve Matris Rotasyonu Tekniğini Kullanarak Parametre Tahminlerinin ve Serbestlik Derecesinin Geometrik Açidan İncelenmeleri	38
2. BÖLÜM QR AYRIŞIMI	55
2.1. QR Ayrışımı Kavramı	55
2.2. Regresyon Analizine QR Ayrışımının uygulaması	62

3. BÖLÜM LU AYRIŞIMI	72
3.1. LU Ayrışımı Kavramı	72
3.1.1. Gauss Eliminasyonu ve LU Ayrışımı	72
3.1.2. Pivot Yöntemi	79
3.1. QR Ayrışımının Hesaplanmasına LU 'nun Katkısı	85
SONUÇ	88
KAYNAKÇA	89

TABLO ve ŞEKİLLER LİSTESİ

Tablo 1.1. Örnek 1.1'deki Değişkenlerin Verileri	18
Tablo 1.2. Örnek 1.1'deki X Matrisinin SVD 'si	18
Tablo 1.3. Örnek 1.1'deki Verinin Regresyon Analizinin Sonuçları	24
Tablo 1.4. Örnek 1.4 'teki Bağımsız Değişken X ve Bağımlı Değişken y 'nin Verileri	26
Tablo 1.5. Örnek 1.4 'teki X Matrisinin SVD 'si	26
Tablo 1.6. Örnek 2.1'deki Verinin Regresyon Analizinin Sonuçları	69
Şekil 1(a) Örnek 1.5 'teki A 'nın İki Tane f_1, f_2 Sütun Vektörü ve d Noktasının Standart Koordinat Sistemindeki Konumları	39
Şekil 1(b) Örnek 1.5 'teki f_1 ve f_2 Vektörlerinin Oluşturduğu Koordinat Sistemindeki d Noktasının Konumu	40
Şekil 2(a) Örnek 1.6 'daki f_1^*, f_2^* Vektörü ve d^* Noktasının Standart Koordinat Sistemindeki Konumları	41
Şekil 2(b) Örnek 1.6'daki d^* Noktasının f_1^* ve f_2^* Vektörlerinin Oluşturduğu Koordinat Sistemindeki Konumu	42
Şekil 2(c) θ^{-1} Transformasyon Matrisi ile Örnek 1.6 'daki d_1^* Noktasının Yatay ve Dikey Koordinat Değerleri Değişerek d_2^* Noktasına Dönüşmesi	43
Şekil 2(d) Örnek 1.6 'daki d_2^* Noktasının Standart Koordinat Sistemindeki Konumu	44
Şekil 3(a) Örnek 1.7 'deki u_1, u_2 ve u_3 Vektörlerinin Oluşturduğu Koordinat Sistemindeki Y Noktasının Konumu	46
Şekil 3(b) θ^+ Transformasyon Matrisi olurken Örnek 1.7'deki y_1 Noktasının Koordinat Değerleri Değişerek y_2 Noktasına Dönüştürülmesi	47
Şekil 3(c) Örnek 1.7'deki y_2 Noktasının (veya Vektörü), v_1^T ve v_2^T Vektörlerinin Oluşturduğu Uzayda $\hat{\beta}$ Noktası (veya Vektörü) olarak açıklanması	48
Şekil 3(d) Örnek 1.7'deki $\hat{y}$ ve $\hat{\varepsilon}$ Vektörünün Uzaydaki Konumları	51

KISALTMALAR ve SİMGELER LİSTESİ

EKKY	En Küçük Kareler Yöntemi
ESS	Hataların Kareleri Toplamı (Error Sum of Squares)
EVD	Öz Değer Ayrışımı (Eigen Value Decomposition)
SVD	Tekil Değer Ayrışımı (Singular Value Decomposition)
A^G	A Matrisinin Genelleştirilmiş Tersi
A^+	A Matrisinin Moore – Penrose Tersi
A^{-1}	A Matrisinin Tersi
A^T	A Matrisinin Devriği
$\text{diag}(a_1, a_2, \dots, a_n)$	Terimleri $a_1, a_2, \dots, a_n$ Olan A Köşegen Matrisi
$\det(A)$	A Matrisinin Determinantı
e_i	i 'nci Birim Vektör
I	Birim Matris
$\text{rank}(A)$	A Matrisinin Rankı
$S(A)$	A Matrisinin Sütun Uzayı
$\mathbb{R}^n$	n Boyutlu Reel Vektörler Uzayı
$\mathbb{R}^{m \times n}$	$m \times n$ Boyutlu Reel Matrisler Uzayı
$\ z\ _2$	z Vektörünün 2 - normu
$\ A\ _2$	A Matrisinin 2 – normu
ϕ	Boş Küme

GİRİŞ

1960'lı yıllardan bu yana bir matrisi daha basit matrise dönüştürebilmek için kullanılan matris ayrışımı, sayısal analizde hakim olan yöntemlerinden biridir.

Matrisin ayrışımında en çok kullanılan algoritmalar ise SVD, QR ve LU ayrışımı ; CS ayrışımı ¹; QLP ayrışımı ; Pivoted QR ve Pivoted QLP ayrışımı ; URV ve ULV ayrışımıdır ². Yapılan bu çalışmada, matris ayrışımından en yaygın olan üç algoritma SVD, QR ve LU 'nun teorileri ve regresyon analizinde uygulanmaları yer almaktadır.

Özellikle matris köşegenleştirilmesi ile sıkı bir ilişkisi olan SVD , tüm matris ayrışım algoritmalarının başında gelerek çalışmanın temelini oluşturmaktadır.

Birinci bölümde, bir matrisin SVD 'sinin varlığının ispatı ³ ve regresyon analizinde uygulamaları yer almaktadır. Bu bölümde SVD yönteminin hem bağımsız değişkenlerin aralarındaki çoklu doğrusallıkların analizi için, hem de en küçük kareler (EKK) yöntemine bir alternatif olarak kullanabileceğine ilişkin bulgular elde edilmiştir. Ayrıca regresyon modelinin parametre tahmincisi $\hat{\beta}$ ' ya , SVD uygulanması ile yeni bakış açıları sunulmaktadır.

Çalışmanın ikinci bölümünde tezin diğer bir konusu olan QR ayrışımının ispatı ⁴ ve regresyon analizinde uygulaması anlatılmaktadır. Bu uygulamada QR ayrışımının bakışından parametre tahmincisi $\hat{\beta}$ 'yı açıklamaktayız.

Üçüncü bölümde Gauss Eliminasyonu (Elimination) 'nun matematiksel olarak

¹ Stewart G.W, Matrix Algorithm, Volume I :Basic Decompositions, SIAM, 1998,s.73 – 77.

² a.g.e., s.367 – 416.

³ Bu ispat gerçekten Wedderburn Rank – One Reduction temelli bir uygulanmadır. Bu konudaki daha ayrıntılı bilgiler Chu Moody T (December 1995) 'te vardır. Ayrıca benzer bir ispat Golub Gene.H. (1996, s.55) 'te de vardır.

⁴ QR ayrışımı için iki tane önemli yöntem vardır.Birisi Klasik GramSchmid ortonormalleştirme yöntemidir, diğeri ise Householding Üçgenleştirme yöntemidir.Tezde Klasik Gram Schmidt ortonormalleştirme yöntemi kullanılmıştır. Ayrıntılı bilgi için bakınız. Stewart G.W (1998, Chap. 4).

LU ayrışımına denk olduğu gösterilmiş ve LU ayrışımı türetilmesi ele alınmıştır. Ayrıca QR ile LU ayrışımı arasındaki bağlantıyı göstermek üzere bir uygulama yapılmıştır.

1. BÖLÜM

TEKİL DEĞER AYRIŞIMI (SVD)

1.1 Tekil Değer Ayrışımı Kavramı

Matris ayrışımından en önemlisi olan Tekil Değer Ayrışımının ispatı ve Öz Değer Ayrışımı ile arasındaki ilişki bu kısımda tartışılmaktadır.

1.1.1 Öz Değer Ayrışımı

Tekil Değer Ayrışımı (SVD) girmeden önce SVD 'ye benzeyen bir ayrışım olan Öz Değer Ayrışımını (Eigen Value Decomposition-EVD) inceleyelim. Aşağıdaki matris ve vektörler verilsin :

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}, p_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, p_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$A p_1 = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 4 \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 4 p_1 \quad (1.1)$$

$$A p_2 = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1 p_2 \quad (1.2)$$

Denklem (1.1) ve (1.2) 'den denklem (1.3) 'ü elde edebiliriz,

$$A [p_1 \ p_2] = [p_1 \ p_2] \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix} \quad (1.3)$$

$$[p_1 \ p_2] = P, \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix} = \theta. \text{ Dolayısı ile denklem (1.3)' ü aşağıdaki gibi}$$

yeniden yazabiliriz :

$$A P = P \theta \quad (1.4)$$

P bir tekil matris olmadığı için, denklem (1.4)'ü aşağıdaki gibi yazabiliriz.

$$A = P \theta P^{-1} \quad (1.5)$$

Denklem (1.5), A kare matrisinin öz değer ayrışımının (EVD) açıklanmasıdır.¹ Denklem (1.1) ve (1.2) 'yi sağlayan 4 ve -1 ise, A matrisinin öz değerleridir. Denklem (1.4) ve (1.5) 'teki gibi, A 'nın öz değerleri 4 ve -1 oluşturdukları θ köşegen matris ise, A 'nın öz değerler matrisidir.

Denklem (1.3) 'teki gibi, A 'nın öz değeri 4 ve -1 'e karşılık gelen p_1 ve p_2 vektörü ise, A matrisinin öz vektörleridir. p_1 le p_2 vektörünün oluşturduğu P matrisi, A 'nın öz vektör matrisidir.

Kompleks cisim üzerinde herhangi bir kare matrisin öz değeri vardır. Fakat reel cisim üzerinde her kare matrisin öz değeri vardır diyemeyiz ². Örneğin,

$\begin{bmatrix} 2 & -5 \\ 1 & -2 \end{bmatrix}$ 'nın reel öz değeri yoktur.

Burada kare matrisin öz değerinin türetilmesini inceleyelim. Öz değerin türetilmesinde genel olarak matrisin öz denklemini kullanırız.

n mertebeli kare matris B aşağıdaki gibi olsun.

$$B = \begin{bmatrix} b_{11} & b_{12} & . & . & b_{1n} \\ b_{21} & b_{22} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ b_{n1} & b_{n2} & . & . & b_{nn} \end{bmatrix}$$

$\Delta(\gamma) = \det(\gamma I - B) = 0$ denklemi, B 'nin öz denklemi olarak adlandırılır ³. B 'nin öz denkleminin kökü γ ise, B 'nin öz değeridir ⁴. Ayrıca $\Delta(\gamma) = \det(\gamma I - B)$ ise, B 'nin öz polinomu olarak da adlandırılır. Burada B 'nin öz polinomunu bulalım.

¹ EVD' nin ayrıntıları için bakınız. Lipschutz Seymour, Lineer Cebir, Çev. Hilmi Hacısalihioğlu, Nobel Yayın Dağıtım, Ankara, 1991, s.286.

² a.g.e., s.284.

³ a.g.e., s.281.

⁴ a.g.e., s.284.

$$\begin{aligned}
\Delta(\gamma) &= \det(\gamma I - B) \\
&= \gamma^n - \gamma^{n-1} \sum_{i=1}^n b_{ii} + \gamma^{n-2} \left\{ \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ b_{31} & b_{33} \end{vmatrix} + \dots + \begin{vmatrix} b_{(n-1)(n-1)} & b_{(n-1)n} \\ b_{n(n-1)} & b_{nn} \end{vmatrix} \right\} \\
&\quad - \gamma^{n-3} \left\{ \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \dots + \begin{vmatrix} b_{(n-2)(n-2)} & b_{(n-2)(n-1)} & b_{(n-2)n} \\ b_{(n-1)(n-2)} & b_{(n-1)(n-1)} & b_{(n-1)n} \\ b_{n(n-2)} & b_{n(n-1)} & b_{nn} \end{vmatrix} \right\} \dots \\
&\quad \dots \mp \gamma^1 \left\{ \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1(n-1)} \\ b_{21} & b_{22} & \dots & b_{2(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{(n-1)1} & b_{(n-1)2} & \dots & b_{(n-1)(n-1)} \end{vmatrix} + \dots + \begin{vmatrix} b_{22} & b_{23} & \dots & b_{2n} \\ b_{32} & b_{33} & \dots & b_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n2} & b_{n3} & \dots & b_{nn} \end{vmatrix} \right\} \\
&\quad \pm \gamma^0 \det(B)
\end{aligned} \tag{1.6}$$

Denklem (1.6) 'daki $\Delta(\gamma)$ 'nin son öğesinin işareti ' $\pm$ 'dir. n bir çift sayı ise, işareti ' + ' olur. n bir tek sayı ise, işareti ' - ' olur.

Denklem (1.6) şu şekilde de özetlenebilir:

$$\begin{aligned}
\Delta(\gamma) &= \gamma^n - \gamma^{n-1} \sum \det(B_{1 \times 1}) + \gamma^{n-2} \sum \det(B_{2 \times 2}) - \gamma^{n-3} \sum \det(B_{3 \times 3}) + \dots \\
&\quad \dots \mp \gamma^1 \sum \det(B_{(n-1) \times (n-1)}) \pm \gamma^0 \det(B)
\end{aligned} \tag{1.7}$$

Burada $B_{1 \times 1}$, B matrisinin köşegenindeki 1 mertebeli kare altmatrisidir, yani B 'nin köşegeninin terimidir. $B_{2 \times 2}$ ise, B matrisinin köşegenindeki 2 mertebeli kare altmatrisidir.

$$\text{Denklem (1.7) 'ye göre } A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ 'nin öz polinomu } \Delta(\gamma)^*$$

aşağıdaki gibi açıklanabilir,

$$\Delta(\gamma)^* = \gamma^2 - \gamma(a_{11} + a_{22}) + \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \gamma^2 - 3\gamma + (-4)$$

A 'nın öz denklemi ise, $\Delta(\gamma)^* = \gamma^2 - 3\gamma + (-4) = 0$ 'dir. Dolayısı ile bu denklemin iki kökleri $\gamma_1 = 4$ ve $\gamma_2 = -1$, A matrisinin iki tane öz değeridir.

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \begin{bmatrix} 2 & -5 \\ 1 & -2 \end{bmatrix} \text{ olsun, } C \text{ 'nin öz denklemi ise:}$$

$$\Delta(\gamma)^{**} = \gamma^2 - \gamma(c_{11} + c_{22}) + \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} = \gamma^2 + 1 = 0$$

Açıktır ki, $\Delta(\gamma)^{**} = \gamma^2 + 1 = 0$ 'ın reel kökü yoktur. Bu yüzden reel cisim üzerinde C matrisinin öz değerleri söz konusu değildir.

Reel öz değerlere sahip olan bir kare matrise denklem (1.5)'deki gibi öz değer ayrışımı (EVD) uygulanabileceği düşünülebilir. Fakat bir matrisin öz değerlerinden ikisi aynıysa ve de karşılık gelen iki tane öz vektörü de aynıysa, o zaman öz vektör matrisi tekil olacağı için tersi yoktur, dolayısı ile bu matris EVD 'ye ayrıştırılmaz.

$$\text{Örneğin, } D = \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix} \text{ olsun, } D \text{ 'nin öz denklem } \Delta(\gamma)^{***} = (\lambda - \gamma)^4 = 0$$

dır. D 'nin 4 tane öz değerinin hepsi λ olarak aynıdır. Dolayısı ile D 'nin öz vektör matrisi tekildir, D kare matrisi EVD 'ye ayrıştırılmaz. Fakat herhangi bir simetrik matris EVD 'ye ayrıştırılabilir ve öz vektör matrisi de ortogonaldır⁵. Üstelik ortogonal matris lineer cebir alanındaki araştırmamız için ideal bir matristir.

$$\text{Örneğin, } E = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \text{ olsun. } E \text{ 'nin öz değeri } r = 6, r = 1 \text{ dir, } E \text{ 'nin öz vektör}$$

matrisi ise $\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$ 'dir ve bir ortogonal matristir.

$$E \text{ 'nin öz değer ayrışımı ise, } E = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}^{-1} \text{ 'dir.}$$

⁵ a.g.e., s.281.

Sadece E matrisi değil, her hangi bir simetrik matrisin öz vektör matrisi bir ortogonal matris olmak üzere EVD 'ye ayrıştırılabilir . Fakat simetrik olmayan bir kare matris EVD 'ye ayrıştırılmasına rağmen, öz vektör matrisi bir ortogonal matris olmayabilir. Mesela, denklem (1.5) 'teki A , EVD 'ye ayrıştırılmasına rağmen öz vektör matrisi V bir ortogonal matris değil.

Kare yada dikdörtgen her türlü matris, simetrik matrisin EVD 'si gibi ayrıştırılabilirse, regresyon analizinde kolaylık sağlanabileceği de doğal bir düşüncedir. Bu düşüncüyü gerçekleştirmek için Tekil Değer Ayrışımına başvurmamız gerekmektedir.

1.1.2 . Tekil Değer Ayrışımı

Geçen kısımda kare matrisin öz değer ayrışımının üzerinde tartıştık, bu kısımda simetrik matrisin EVD 'sine benzer bir matris ayrışımı olan tekil değer ayrışımını (SVD) inceleyelim.

$$X = \begin{bmatrix} 3.01 & 0.01 & -2.99 \\ 2.99 & -0.01 & -3.01 \\ 2.00 & -4.00 & 2.00 \end{bmatrix}$$

olsun. X matrisi aşağıdaki gibi ayrıştırılabilir :

$$X = \begin{bmatrix} 3 & 0 & -3 \\ 3 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & -4 & 2 \end{bmatrix} + \begin{bmatrix} 0.01 & 0.01 & 0.01 \\ -0.01 & -0.01 & -0.01 \\ 0 & 0 & 0 \end{bmatrix} =$$

$$6 \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} + 2\sqrt{6} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} + \frac{\sqrt{6}}{100} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{aligned}
&= 6 \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \end{bmatrix} + 2\sqrt{6} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \\
&+ \sqrt{6}/100 \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} \\
&= \theta_1 U_1 V_1 + \theta_2 U_2 V_2 + \theta_3 U_3 V_3 = \sum_i^3 \theta_i U_i V_i^T \quad (1.8)
\end{aligned}$$

$$\begin{aligned}
&= \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2\sqrt{6} & 0 \\ 0 & 0 & \sqrt{6}/100 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} \\
&= U \theta V^T \quad (1.9)
\end{aligned}$$

Görülüyor ki, denklem (1.8) 'deki U_i , denklem (1.9) 'daki U ortonormal matrisinin sütunudur. Denklem (1.8) 'deki V_i^T ise, (1.9)'daki V^T ortonormal matrisinin satırıdır. θ_i ise, θ köşegen matrisinin köşegenindeki öğesidir. Denklem (1.8) ve (1.9) ise, X matrisinin tekil değer ayrışımının (SVD) açıklanmasıdır. Sadece buradaki X değil, her türlü matris denklem (1.9) gibi ayrıştırılabilir. Bu ifade Teorem 1.1'de daha ayrıntılı açıklanmaktadır.

Teorem 1.1. $m \times n$ boyutlu A matrisi, rankı r olan reel bir matris ise, $U^T A V = \theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_r, 0, \dots, 0)$ eşitliğini sağlayan ortonormal matris U ve V vardır. $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$, $\theta \in \mathbb{R}^{m \times n}$. Ayrıca $\theta_1 \geq \theta_2 \geq \dots \geq \theta_r \geq 0$.

Bu teoremin ispatları Golub Gene.H. (1996) ⁶ ve Stewart G.W. (1998) ⁷ 'de

⁶ Golub, Gene. H., Matrix Computation, Third Edition, The Johns Hppkins University Pres, Baltimore, 1996, s.70.

⁷ Stewart G.W., Matrix Algorithms, Volume I : Basic Decompositions, SIAM, 1998, s.62

yer almaktadır. Burada, iki kitaptaki ispatlara benzer bir ispat sunulmaktadır.

Teorem 1.1'in ispatını yapmadan önce kullanacağımız vektör ve matrisin 2-norm'u ile ilgili kavramlara biraz değinelim. Bu konudaki ayrıntıları Golub Gene. H. (1996)⁸ 'da yer almaktadır.

Teorem 1.2. Vektör $\chi^T = [\chi_1 \ \chi_2 \ \dots \ \chi_n]^T$ olsun. χ vektörünün 2-normu ise, $\|\chi\|_2 = \sqrt{\chi_1^2 + \chi_2^2 + \dots + \chi_n^2}$ 'dir⁹.

Teorem 1.3. A matrisi $\in \mathbb{R}^{m \times n}$ olsun, $A^T A z = \theta^2 z$ olduğundan, 2- normu birim olan n boyutlu bir vektör z varsa, skaler θ , A 'nın 2- normudur, yani $\|A\|_2 = \theta$ ¹⁰.

Gerçekten A matrisinin 2- normu $\|A\|_2$, $A^T A$ 'nın en büyük öz değerinin kare köküdür¹¹.

Y.Teorem 1.3.1. A matrisi $\in \mathbb{R}^{m \times n}$, χ vektörü $\in \mathbb{R}^n$ ise, $\|A\chi\|_2 \leq \|A\|_2 \|\chi\|_2$ olarak tutarlıdır¹².

Y.Teorem 1.3.2.

1.3.2.1. Q bir ortonormal matris ve χ uygun boyutlu bir vektör ise, $\|Q\chi\|_2^2 = \chi^T Q^T Q \chi = \chi^T \chi = \|\chi\|_2^2$ dir, yani $\|Q\chi\|_2 = \|\chi\|_2$ ¹³.

1.3.2.2. Q ve Z iki ortonormal matris olsun. $Q \in \mathbb{R}^{m \times m}$, $Z \in \mathbb{R}^{n \times n}$. A uygun boyutlu bir matris ise, $\|Q^T A Z\|_2^2 = \|A\|_2^2$, yani $\|Q^T A Z\|_2 = \|A\|_2$ ¹⁴.

⁸ Golub, Gene. H.: Matrix Computation, Third Edition, The Johns Hppkins University Pres, Baltimore, 1996, s.52 – 57.

⁹ a.g.e., s.52

^{10,11} a.g.e., s.57

¹² a.g.e., s.55

^{13,14} a.g.e., s.70

Teorem 1.4. $V_1 \in \mathbb{R}^{m \times n}$ bir ortonormal matris olsun. $V = [V_1 \ V_2]$ bir ortonormal matris olmak üzere ortonormal matris V_2 vardır ¹⁵.

Şimdi Teorem 1.2. , 1.3. , 1.4. ve Y.Teorem 1.3.1. , 1.3.2. 'yi kullanarak Teorem 1.1. 'in ispatını yapalım. İspat sürecinde kolaylık sağlaması için, bu ispatta kullandığımız A örnek matrisi , rankı tam olan bir matris olsun.

İspat: $m \times n$ boyutlu A matrisinin rankı $\text{rank}(A) = \min(m, n)$; $\|A\|_2 = \theta_1$ ve 2 - norm birim vektör $\chi \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ olsun. χ ve y vektörünün değerlerini ayarlayarak aşağıdaki denklemi kurabiliriz : $A \chi = \theta_1 y$. Teorem 1.4. 'ten biliyoruz ki χ 'le birlikte $n \times n$ boyutlu ortonormal matris V 'yi oluşturabilen ortonormal matris V_1 vardır. Yani $V = [\chi \ V_1]$ 'dir.

Keza, y ile birlikte $m \times m$ boyutlu ortonormal matris U 'yu oluşturabilen ortonormal matris U_1 de vardır. Yani $U = [y \ U_1]$ 'dir . Dolayısı ile

$$\begin{aligned}
 U^T A V &= \begin{bmatrix} y^T \\ U_1^T \end{bmatrix} A [\chi \ V_1] \\
 &= \begin{bmatrix} y^T A \chi & y^T A V_1 \\ U_1^T A \chi & U_1^T A V_1 \end{bmatrix} \\
 &= \begin{bmatrix} y^T \theta_1 y & w^T \\ U_1^T \theta_1 y & B_1 \end{bmatrix} \\
 &= \begin{bmatrix} \theta_1 & w^T \\ 0 & B_1 \end{bmatrix} \\
 &= \tilde{A}
 \end{aligned} \tag{1.10}$$

Denklem (1.10)'dan $y^T A V_1 = w^T \in \mathbb{R}^{1 \times (n-1)}$; $U_1^T A V_1 = B_1 \in \mathbb{R}^{(m-1) \times (n-1)}$ dir.

¹⁵ a.g.e., s.

Dolayısı ile,

$$\begin{aligned}
\left\| \tilde{A} \begin{bmatrix} \theta_1 \\ w \end{bmatrix} \right\|_2^2 &= \left\| \begin{bmatrix} \theta_1 & w^T \\ 0 & B_1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ w \end{bmatrix} \right\|_2^2 \\
&= \left\| \begin{bmatrix} \theta_1^2 + w^T w \\ B_1 w \end{bmatrix} \right\|_2^2 \\
&= (\theta_1^2 + w^T w)^2 + (B_1 w)^T B_1 w \geq (\theta_1^2 + w^T w)^2
\end{aligned} \tag{1.11}$$

Y.Teorem 1.3.1. ve denklem (1.11) 'den biliyoruz ki

$$\left\| \tilde{A} \right\|_2^2 \left\| \begin{bmatrix} \theta_1 \\ w \end{bmatrix} \right\|_2^2 \geq \left\| \tilde{A} \begin{bmatrix} \theta_1 \\ w \end{bmatrix} \right\|_2^2 \geq (\theta_1^2 + w^T w)^2 \tag{1.12}$$

$$\left\| \begin{bmatrix} \theta_1 \\ w \end{bmatrix} \right\|_2^2 = \theta_1^2 + w^T w \text{ olduğu için eşitsizlik (1.12) aşağıdaki gibi yazılabilir.}$$

$$\left\| \tilde{A} \right\|_2^2 \geq \theta_1^2 + w^T w \tag{1.13}$$

Y.Teorem1.3.2.2. ve denklem (1.10)'a dayanarak denklem (1.14)'ü elde ederiz.

$$\left\| A \right\|_2^2 = \left\| U^T A V \right\|_2^2 = \left\| \tilde{A} \right\|_2^2 \tag{1.14}$$

Denklem (1.13) ve denklem (1.14) 'den denklem (1.15)'i elde ederiz.

$$\left\| A \right\|_2^2 = \left\| \tilde{A} \right\|_2^2 \geq \theta_1^2 + w^T w \tag{1.15}$$

İspatın başlangıcında $\left\| A \right\|_2^2 = \theta_1^2$ diye varsaydık. Bu varsayımı temel alarak denklem (1.15)' i elde ederiz. Bu yüzden $w = 0$ olmalıdır.

Dolayısı ile denklem (1.10) şu şekilde yazılabilir:

$$U^T A V = \begin{bmatrix} \theta_1 & 0 \\ 0 & B_1 \end{bmatrix} \tag{1.16}$$

Gerçekten V_1 'le V 'yi oluşturan χ ve U_1 'le U 'yu oluşturan y vektörü, ayrı ayrı olarak A matrisinin tekil değeri θ_1 'le karşılık gelen sağ tekil vektörü (right singular vector) ve sol tekil vektörüdür (left singular vector) . θ_1 hem A 'nın hem $U^T A V$ 'nin 2 – normudur, yani $\theta_1 = \|A\|_2 = \|U^T A V\|_2$ 'dir, buradaki $U \in \mathbb{R}^{m \times m}$ ve $V \in \mathbb{R}^{n \times n}$ dir. Ayrıca $B_1 = U_1^T A V_1$ 'in 2 – normu $\|B_1\|_2 = \|U_1^T A V_1\|_2 = \theta_2$ 'dir. Buradaki $U_1 \in \mathbb{R}^{m \times (m-1)}$ ve $V_1 \in \mathbb{R}^{n \times (n-1)}$, dolayısı ile $\|B_1\|_2 = \|U_1^T A V_1\|_2 \leq \|U^T A V\|_2 = \|A\|_2$ 'dir, yani $\theta_2 \leq \theta_1$ 'dir.

Denklem (1.16)'ya benzer şekilde B_1 de

$$B_1 = U_1^T A V_1 = \begin{bmatrix} \theta_2 & 0 \\ 0 & B_2 \end{bmatrix} \quad (1.17)$$

olarak açıklanabilir. Buradaki $B_2 \in \mathbb{R}^{(m-2) \times (n-2)}$.

Denklem (1.17) 'yi denklem (1.16) 'ya dönüştürdükten sonra, denklem (1.18)'i elde ederiz.

$$U^T A V = \begin{bmatrix} \theta_1 & 0 & 0 \\ 0 & \theta_2 & 0 \\ 0 & 0 & B_2 \end{bmatrix} \quad (1.18)$$

Aynı mantıkla denklem (1.18) 'daki B_2 de bir blok köşegen matrisine dönüştürülebilir. Böyle devam ederek son olarak B_{r-1} köşegen matrisini elde ederiz. $B_{r-1} = \text{diag} (\theta_r, 0 \dots 0)$ 'dir. r ise A matrisinin rankıdır.

Dolayısı ile denklem (1.18) şu şekilde yeniden yazılabilir,

$$U^T A V = \left[\begin{array}{ccc|ccc} \theta_1 & & & & & \\ & \theta_2 & & \phi & & \\ & & \ddots & & \phi & \\ & \phi & & & & \\ & & & \theta_r & & \\ \hline & & & & & \\ & & \phi & & \phi & \end{array} \right] = \text{diag}(\theta_1, \theta_2, \dots, \theta_r, 0 \dots 0) \quad (1.19)$$

Denklem (1.19) ise, A matrisinin tekil değer ayrıştırılmasının (SVD) bir açıklanmasıdır. Ayrıca $\theta_1 \geq \theta_2 \geq \dots \geq \theta_r \geq 0$ dır.

Burada U ve V rankı tam olan kare ortonormal matris olduğu için denklem (1.19) şu şekilde de yazılabilir :

$$A = U \theta V^T$$

U , A matrisinin sol tekil vektör matrisidir. Yani U 'nun sütunları A matrisinin sol tekil vektörleridir. Ayrıca U matrisi AA^T 'nin öz vektör matrisidir ¹⁶.

V ise, A matrisinin sağ tekil vektör matrisidir. Yani V 'nin sütunları A matrisin sağ tekil vektörleridir. Ancak aynı zamanda V de $A^T A$ 'nın öz vektör matrisidir ¹⁷.

Biliyoruz ki θ , A matrisinin tekil değer matrisi dir. Aynı zamanda θ_i^2 ($0 \leq i \leq r$) terim olarak oluşturduğu köşegen matris, hem matris AA^T hem $A^T A$ 'nin öz değeri matrisidir. Yani simetrik matris AA^T 'ninki ile simetrik matris $A^T A$ 'nın sıfır olmayan öz değerleri aynıdır.

Teorik olarak A 'nın sol tekil vektör matrisi U ile sağ tekil vektör matrisi V , AA^T 'nin öz vektör matrisi ve $A^T A$ 'nın öz vektör matrisi olması ile

^{16,17} Golub Gene. H.: Matrix Computation, Third Edition, The Johns Hppkins University Pres, Baltimore, 1996, s.448.

hesaplanabilirler. Keza, θ ise, AA^T veya $A^T A$ 'nın sıfır olmayan öz değerleri olması ile hesaplanabilir. Fakat pratikte böyle yaparak bile doğru sonuçları her zaman kesin olarak ortaya çıkaramayız,

U , V ve θ 'nin hesaplanmaları, yani A 'nın SVD 'nin pratik hesaplanması Golub Gene. H. (1996) 'nın 8 'inci bölümünde yer almaktadır. Ama SVD 'nin hesaplanmasına yardımcı olan bazı bilgisayar programları mevcut olduğu için¹⁸ burada SVD 'nin hesaplanmasındaki algoritmaların üzerinde durmayacağız.

1.2. Regresyon Analizine SVD Uygulamaları

Bu kısımda , üç konuda SVD 'nin regresyon analizine uygulamaları tartışılmaktadır. Birinci konu, EKKY 'ye alternatif olan SVD yönteminin regresyon analizine uygulanması, çoklu doğrusal bağlantının incelenmesi ve çözülmesine katkılardır. İkinci konu SVD ve Genelleştirilmiş ters kavramını kullanarak β 'nin regresyon parametrelerinin tahmin vektörü $\hat{\beta}$ 'nin incelenmesidir. Üçüncüsü ise, SVD 'yi kullanarak $\hat{\beta}$ 'nin geometrik açıdan incelenmesidir.

1.2.1. Çoklu Doğrusal Bağlantılarının (Collinearity) SVD ile İncelenmesi ve Çözülmesi

Bu kısımda bağımsız değişkenlerin aralarındaki tam olmayan ve tam olan doğrusal bağlantıların doğrusal regresyon modelinin üzerine etkilerinin SVD ile incelenmesi ve çözülmesi tartışılacaktır.

¹⁸ SVD 'nin hesaplanması için Matematik 4.0 ve 5.0 'da ilgili programlar mevcuttur. Ayrıca Golub Gene H.(1996)'nın xiii –xiv sayfaları arasında SVD'nin hesaplanması üzere LAPACK EISPACK ve LINPACK programları da tavsiye edildi. Daha ayrıntılı bilgileri için, web sitesi – <http://www.netlib.org/index.html> 'e başvurulabilir.

1.2.1.1 Tam Olmayan Çoklu Doğrusal Bağlantıların (Near Collinearity) SVD ile İncelenmesi ve Çözülmesi

Bu kısımda doğrusal denklem sisteminin bağımsız değişken matrisinin tekil değerlerindeki hastalığın (ill condition), denklemin çözümü üzerine etkisinin ne olacağını inceleyeceğiz. Tekil değerindeki bu hastalık bağımsız değişkenlerin aralarındaki çoklu doğrusal bağlantının yansıtıcısıdır. Geçen kısımdan biliyoruz ki $m \times n$ boyutlu X matrisi aşağıdaki gibi ayrıştırılabilir:

$$X = U\theta V^T$$

X bağımsız değişkenler matrisi ve Y bağımlı değişken vektörü olmak üzere doğrusal çoklu regresyon modeli aşağıdaki gibi olsun. ε , teorik hata payı yada disturbanstır.

$$Y = X\beta + \varepsilon \quad (1.20)$$

$X = U\theta V^T$, denklem(1.20)'de yerine konulursa,

$$Y = U\theta V^T \beta + \varepsilon \quad (1.21)$$

elde edilir. Denklem (1.21)'e "Asal Bileşen Regresyon Model" denir. ¹⁹

$$a = \theta V^T \beta \quad (1.22)$$

olsun. Denklem (1.21) ve denklem (1.22)'den denklem (1.23)'ü elde ederiz.

$$Y = Ua + \varepsilon \quad (1.23)$$

$U^T U = I$, birim matris olarak tekil olmadığı için En Küçük Kareler Yöntemi uygulayarak denklem (1.24)'teki gibi a 'nın tahmincisini elde edebiliriz,

$$\hat{a} = (U^T U)^{-1} U^T Y \quad (1.24)$$

Buradan

¹⁹ Mandel J., Use of the Singular Value Decomposition in Regression Analysis, the American Statistician, Vol.36, s.16.

$$\hat{a} = U^T Y \quad (1.25)$$

olur. Denklem (1.23), denklem (1.25)'e dönüştürüldükten sonra,

$$\hat{a} = U^T (Ua + \varepsilon) = a + U^T \varepsilon \quad (1.26)$$

elde ederiz. Dolayısı ile ,

$$\hat{a}_i - a_i = u_i^T \varepsilon \quad (1.27)$$

elde ederiz. $\hat{a}_i$ ve a_i ise, sırasıyla $\hat{a}$ ve a 'nın i 'nci terimidir. u_i^T ise, U^T 'nin i 'nci satırıdır.

Denklem (1.27)'den $E(\hat{a}_i - a_i) = E(u_i^T \varepsilon) = 0$ elde edebiliriz. Yani,

$$E(\hat{a}_i) = a_i \quad (1.28)$$

Denklem(1.28)' den $\hat{a}$, a 'nın sapmasız tahmin edicisi olduğunu görebiliriz.

Burada $\hat{a}_i$ 'nin varyansı $Var(\hat{a}_i)$ ' yi türetilim.

$$\begin{aligned} Var(\hat{a}_i) &= E(\hat{a}_i - a_i)^2 = E(u_i^T \varepsilon)^2 \\ &= E(u_{i1}\varepsilon_1 + u_{i2}\varepsilon_2 + \dots + u_{in}\varepsilon_n)^2 \\ &= E\left(\sum_{k=1}^m u_{ik}^2 \varepsilon_k^2 + 2\sum_{k=1}^m \sum_{j=1}^m u_{ik} u_{ij} \varepsilon_k \varepsilon_j\right), \quad k \neq j \\ &= \sum_{k=1}^m u_{ik}^2 E(\varepsilon_k^2) + 2\sum_{k=1}^m \sum_{j=1}^m u_{ik} u_{ij} E(\varepsilon_k \varepsilon_j) \\ &= \sum_{k=1}^m u_{ik}^2 E(\varepsilon_k^2) \quad (\text{Burada } E(\varepsilon_k^2) = \delta^2 \text{ olsun}) \\ &= \delta^2 \sum_{k=1}^m u_{ik}^2 \\ &= \delta^2 \end{aligned} \quad (1.29)$$

Burada δ^2 , ε_i 'nin varyansıdır ; ε_i , ε 'nin i 'nci terimidir. Ayrıca $\sum_{k=1}^m u_{ik}^2 = 1$ 'dir.

Denklem (1.22) 'den denklem (1.30) 'u elde edebiliriz. Burada $\hat{\beta}$, β 'nın tahmincisidir.

$$\hat{a} = \theta V^T \hat{\beta} \quad (1.30)$$

$\text{rank}(X) = n$ ise, θV^T tekil olmayan bir kare matristir. Dolayısı ile

$$\hat{\beta} = (V^T)^{-1} \theta^{-1} \hat{a} = V \theta^{-1} \hat{a} \quad (1.31)$$

Denklem (1.31) 'den $\hat{\beta}$ 'nın terimi $\hat{\beta}_i$ ($0 \leq i \leq n$) ,

$$\hat{\beta}_i = \sum_{k=1}^n v_{ik} \hat{a}_k / \theta_k \quad (1.32)$$

olarak elde edilir. Burada $\hat{a}_k$, $\hat{a}$ 'nın k 'inci terimidir ; v_{ik} , V 'nin i 'nci satırı v_i 'nin k 'inci terimidir ; θ_k ise θ 'nın köşegenindeki k 'inci en büyük terimidir. Denklem (1.32) 'den $\hat{\beta}_i$ 'nin varyansı $\text{Var}(\hat{\beta}_i)$ 'yi türetelim.

$$\begin{aligned} \text{Var}(\hat{\beta}_i) &= E(\hat{\beta}_i - E(\hat{\beta}_i))^2 \\ &= E\left(\sum_{k=1}^n \hat{a}_k v_{ik} / \theta_k - E\left(\sum_{k=1}^n \hat{a}_k v_{ik} / \theta_k\right)\right)^2 \\ &= E\left(\sum_{k=1}^n \hat{a}_k v_{ik} / \theta_k - \sum_{k=1}^n a_k v_{ik} / \theta_k\right)^2 \\ &= E\left(\sum_{k=1}^n (\hat{a}_k - a_k) v_{ik} / \theta_k\right)^2 \\ &= E\left(\sum_{k=1}^n (\hat{a}_k - a_k)^2 v_{ik}^2 / \theta_k^2 + 2 \sum_{k=1}^n \sum_{j=1}^n (\hat{a}_k - a_k)(\hat{a}_j - a_j) v_{ik} v_{ij} / \theta_k \theta_j\right) \\ &\quad (k \neq j) \\ &= E\left(\sum_{k=1}^n (\hat{a}_k - a_k)^2 v_{ik}^2 / \theta_k^2\right) + 0 \quad (E(\hat{a}_k - a_k)^2 = \delta^2 \text{ 'dir}) \\ &= \delta^2 \sum_{k=1}^n v_{ik}^2 / \theta_k^2 \end{aligned} \quad (1.33)$$

Denklem (1.33) 'teki $Var(\hat{\beta}_i)$ ile yaygın olarak bilinen $Var(\hat{\beta}_i) = (X^T X)^{-1} \delta^2$

ye gerçekte denktir . SVD ile elde ettiğimiz $\hat{\beta}$ ve $Var(\hat{\beta}_i)$ 'nın açıklanmasını aşağıdaki örneğe uygulayalım.

Örnek 1.1. Tablo1.1. 'deki X bağımsız değişkenler matrisini ve y ise bağımlı değişken vektörünü gösterebiliriz . Tablo 1.2. 'deki ise, X 'in tekil değer ayrışımını göstermektedir. ²⁰

Tablo 1.1.

Nokta	X			y
	x_1	x_2	x_3	
1	1	16.85	1.46	41.38
2	1	24.81	-4.61	31.01
3	1	18.85	-0.21	37.41
4	1	12.63	4.93	50.05
5	1	21.38	-1.36	39.17
6	1	18.78	-0.08	38.86
7	1	15.58	2.98	46.14
8	1	16.30	1.73	44.47

Tablo 1.2.

U			θ	V		
u_1	u_2	u_3		v_1	v_2	v_3
0.3225	0.1761	0.1937	52.3478 0 0 0 7.8538 0 0 0 0.0556	0.0530 0.0673 0.9963 0.9985 -0.0083 - 0.0526 0.0047 0.9976 -0.0676		
0.4738	-0.6034	0.4988				
0.3605	-0.0381	0.0362				
0.4087	-0.1869	-0.6591				
0.3592	-0.0215	0.2418				
0.2984	0.3705	-0.4536				
0.3121	0.2109	0.3853				

Tablo 1.1. ve 1.2. 'nin verilerini kullanarak X ve y 'nin üzerine doğrusal regresyon modeli kuralım. Denklem (1.32) 'ye dayanarak, parametre tahmincisi

²⁰ Sadece SVD'nin uygulanmalarının teorileri anlatılmak için Tablo 1.1 ve Tablo 1.2 'de kullanılan suni veriler "Mandel J., Use of the Singular Value Decomposition in Regression Analysis, the American Statistician, Vol.36, s.15 -24" ten iktibas edilmektedir.

$\hat{\beta}$ 'yi hesaplayarak y 'nin regresyon tahmincisi $\hat{y}$ aşağıdaki gibi açıklanabilir.

$$\hat{y} = 0.4263 x_1 + 2.0839 x_2 + 4.6551 x_3$$

Gerçekten SVD ile elde ettiğimiz bu sonuç, en küçük kareler yöntemi ile, yani $\hat{\beta} = (X^T X)^{-1} X^T Y$ formülü ile elde ettiğimiz sonuç ile aynıdır. Çünkü: denklem (1.31) 'den $\hat{\beta} = V\theta^{-1}\hat{a}$ 'dır, denklem (1.25) 'ten $\hat{a} = U^T Y$ olur. Dolayısı ile $\hat{\beta} = V\theta^{-1}U^T Y$ 'dir. Ayrıca $\hat{\beta} = (X^T X)^{-1} X^T Y$ formülüne $X = U\theta V^T$ 'yi dönüştürerek $\hat{\beta} = (V\theta^T U^T U\theta V^T)^{-1} V\theta^T U^T Y = (V\theta^2 V^T)^{-1} V\theta^T U^T Y = V\theta^{-1}U^T Y$ 'yi elde ederiz, $\hat{\beta}$ 'nin bu açıklanmasının, SVD ile açıklanan $\hat{\beta} = V\theta^{-1}U^T Y$ 'ile aynı olduğunu görebiliriz.

Denklem (1.33) 'ü kullanarak $Var(\hat{\beta}_i)$ 'yi ve standart hatası $Sh(\hat{\beta}_i)$ 'yi elde edebiliriz. Tablo 1.1. 'deki verilere göre $\delta^2 = 3.075$ 'dir. Ayrıca,

$$Var(\hat{\beta}_1) = 984.1540 \quad Sh(\hat{\beta}_1) = 31.3712$$

$$Var(\hat{\beta}_2) = 2.7466 \quad Sh(\hat{\beta}_2) = 1.6572$$

$$Var(\hat{\beta}_3) = 4.5919 \quad Sh(\hat{\beta}_3) = 2.1428$$

olur. Burada t testi ile $\hat{\beta}_1$, $\hat{\beta}_2$ ve $\hat{\beta}_3$ 'ü sınavalım.

$$t_1 = (\hat{\beta}_1 - 0) / sh(\hat{\beta}_1) = 0.01358$$

$$t_2 = (\hat{\beta}_2 - 0) / sh(\hat{\beta}_2) = 1.2574$$

$$t_3 = (\hat{\beta}_3 - 0) / sh(\hat{\beta}_3) = 2.1723$$

Serbestlik derecesi $8 - 3 = 5$ 'e göre, $t_{0.05} = 2.571$ 'dir. Dolayısı ile $\hat{\beta}_1$, $\hat{\beta}_2$ ve $\hat{\beta}_3$ 'nin sıfıra eşit olduğu hipotezini reddedemeyiz, yani elde ettiğimiz $\hat{\beta}_i$ 'nin hiç biri istatistiksel olarak anlamlı değildir. Fakat F testi sonucu $F = 26.2110$ ile

tablodaki $F_{0.05(2,5)} = 5.79$ karşılaştırılarak $\hat{\beta}_1$, $\hat{\beta}_2$ ve $\hat{\beta}_3$ 'ün hepsinin sıfır olduğu hipotezini reddedebiliriz. Bu yüzden bağımsız değişkenler x_1 , x_2 ve x_3 'ün aralarında çoklu doğrusal bağlantı olduğundan şüphelenilebilir.

Tablo 1.2. 'deki X 'in tekil değerleri aşağıdaki gibi gösterilmektedir.

$$\theta = \text{diag}(\theta_1 \ \theta_2 \ \theta_3) = \text{dia}(52.3478, 7.8538, 0.0556) = \begin{bmatrix} 52.3478 & 0 & 0 \\ 0 & 7.8538 & 0 \\ 0 & 0 & 0.0556 \end{bmatrix}$$

$\theta_3 = 0.0556$, θ_1 ve θ_2 'ye kıyasla çok küçük gözüküyor. Burada $\text{Var}(\hat{\beta}_1)$ 'i inceleyerek $\text{Var}(\hat{\beta}_1)$ 'in üzerine θ_3 'ün etkisine bakalım.

Denklem (1.33) 'ten biliyoruz ki

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \sigma^2 \sum_{k=1}^3 v_{1k}^2 / \theta_k^2 \\ &= \sigma^2 (v_{11}^2 / \theta_1^2 + v_{12}^2 / \theta_2^2 + v_{13}^2 / \theta_3^2) \\ &= 3.075 \times ((0.0530)^2 / (52.3478)^2 + (0.0673)^2 / (7.8538)^2 \\ &\quad + (0.9963)^2 / (0.0556)^2) \\ &= 3.075 \times (1.035 \times 10^{-5} + 7.35 \times 10^{-5} + 320.05) \\ &= 984.1540 \end{aligned}$$

Yukarda $1/\theta_3^2 = 320.29$ dur, çok büyüktür. İşte bu büyük değer olan $1/\theta_3^2$, $\text{Var}(\hat{\beta}_1)$ 'nin büyük değerine neden olur ve regresyonun β_1 tahmincis $\hat{\beta}_1$ 'in kesinliğini düşürerek $\hat{\beta}_1$ 'i istatistiki olarak anlamsız hale getirmiştir.

Aynı şekilde $\text{Var}(\hat{\beta}_1)$ 'in üzerine negatif etkisi olan θ_3 'ün de $\text{Var}(\hat{\beta}_2)$ ve $\text{Var}(\hat{\beta}_3)$ 'yı şişirerek $\hat{\beta}_2$ ve $\hat{\beta}_3$ 'yı istatistiksel olarak anlamsız hale getirmiştir.

Küçük değer olan θ_3 , sadece $\text{Var}(\hat{\beta}_1)$ 'nin üzerine negatif etki oluşturmamakta

aynı zamanda bağımlı değişken y 'nin tahmincisi $\hat{y}$ 'nin terimi $\hat{y}_i$ 'nin varyansı $Var(\hat{y}_i)$ 'nin büyük değer olmasına etki etmektedir. Ayrıca bu durumdaki X bağımsız değişken matrisine hastalıklı matris (ill conditioned matrix) denilir. SVD ile $Var(\hat{y}_i)$ 'yı türeterek hastalıklı X matrisinin negatif etkisini inceleyelim.

X 'in i 'nci satırı χ_i olsun. X 'in sol tekil vektör matrisi U 'nun i 'nci satırı ise $\bar{u}_i$ olsun. $X = U\theta V^T$ 'ten

$$\chi_i = \bar{u}_i \theta V^T \quad (1.34)$$

yazılır. $rank(X) = n$ ise, θV^T tekil olmayan bir kare matris olur. Dolayısı ile denklem (1.34) aşağıdaki gibi yazılabilir ki

$$\bar{u}_i = \chi_i (V^T)^{-1} \theta^{-1} = \chi_i V \theta^{-1} \quad (1.35)$$

$\hat{y} = X \hat{\beta}$ olduğundan $\hat{y}$ 'nin her bir terimi $\hat{y}_i$, $\hat{y}_i = \chi_i \hat{\beta}$ olarak açıklanabilir.

Dolayısı ile denklem (1.30) ve (1.34) 'ten

$$\hat{y}_i = \chi_i \hat{\beta} = \bar{u}_i \theta V^T \hat{\beta} = \bar{u}_i \hat{a} \quad (1.36)$$

elde ederiz. Şimdi denklem(1.36) 'yı kullanarak $Var(\hat{y}_i)$ 'yı bulalım.

$$\begin{aligned} Var(\hat{y}_i) &= E(\hat{y}_i - E(\hat{y}_i))^2 = E(\bar{u}_i \hat{a} - \bar{u}_i a)^2 \\ &= E(\bar{u}_{i1}(\hat{a}_1 - a_1) + \bar{u}_{i2}(\hat{a}_2 - a_2) + \dots + \bar{u}_{in}(\hat{a}_n - a_n))^2 \\ &= E\left(\sum_{j=1}^n \bar{u}_{ij}^2 (\hat{a}_j - a_j)^2 + 2 \sum_{j=1}^n \sum_{k=1}^n \bar{u}_{ij} (\hat{a}_j - a_j) \bar{u}_{ik} (\hat{a}_k - a_k)\right) \quad k \neq j \\ &= \sum_{j=1}^n \bar{u}_{ij}^2 E(\hat{a}_j - a_j)^2 + 2 \sum_{j=1}^n \sum_{k=1}^n \bar{u}_{ij} \bar{u}_{ik} E(\hat{a}_j - a_j)(\hat{a}_k - a_k) \\ &= \sum_{j=1}^n \bar{u}_{ij}^2 E(\hat{a}_j - a_j)^2 \\ &= \delta^2 \sum_{j=1}^n \bar{u}_{ij}^2 \end{aligned} \quad (1.37)$$

Denklem (1.37) 'daki $\bar{u}_{ij}$ 'nin mutlak değeri $|\bar{u}_{ij}|$ büyükse, $Var(\hat{y}_i)$ de büyür. Burada büyük $|\bar{u}_{ij}|$ 'nin nereden kaynaklayabileceğini Tablo1.1. ve Tablo1.2.'deki verileri kullanarak inceleyelim.

Tablo 1.2. 'deki X 'nin tekil değer matrisi θ 'yı bir blok matris yapalım:

$$\theta = \left[\begin{array}{c|c} \theta_A & 0 \\ \hline 0 & \theta_B \end{array} \right] = \left[\begin{array}{cc|c} 52.3478 & 0 & 0 \\ 0 & 7.8380 & 0 \\ \hline 0 & 0 & 0.0556 \end{array} \right]$$

Tablo1.2.'deki X 'nin sağ tekil vektör matrisi V 'yi bir blok matris yapalım.

$$V = [V_A \mid V_B] = \left[\begin{array}{cc|c} 0.5306 & 0.0673 & 0.9963 \\ 0.9985 & -0.0083 & -0.0526 \\ 0.0047 & -0.0526 & -0.0676 \end{array} \right]$$

Yukardaki bloklaşmış θ ile V 'yi denklem (1.35) 'e dönüştürdükten sonra

$$\bar{u}_i = \chi_i V \theta^{-1} = \chi_i [V_A \mid V_B] \left[\begin{array}{c|c} \theta_A^{-1} & 0 \\ \hline 0 & \theta_B^{-1} \end{array} \right] = [\chi_i V_A \theta_A^{-1} \quad \chi_i V_B \theta_B^{-1}] \quad (1.38)$$

elde ederiz. Tablo1.2.'deki θ değerlerine göre $\theta_B = \theta_3 = 0.0556$ 'dır ve çok küçüktür.

Dolayısı ile θ_B^{-1} çok büyüktür ve $\bar{u}_i$ hayli şişirilmiştir. Şişirilen $\bar{u}_i$, $Var(\hat{y}_i)$ 'nin değerini de esasen büyüttüğünü denklem (1.37) 'den görebiliriz.

Az önce küçük değer olan θ_3 , $Var(\hat{\beta}_i)$ 'nin üzerine negatif etkide de bulunduğunu tartıştık. Hem $Var(\hat{\beta}_i)$ hem de $Var(\hat{y}_i)$ 'yı esasen büyüten küçük değer olan θ_3 'nin nereden kaynaklandığını burada inceleyelim.

Gerçekte sıfıra hayli yakın olan bu küçük θ_3 , x_1 , x_2 ve x_3 bağımsız değişkenlerinin aralarındaki tam olmayan çoklu doğrusal bağlantıyı (nearcollinearity)

yansıtıyor ²¹.

Denklem (1.35) aşağıdaki gibi yeniden yazılabilir,

$$\bar{u}_i \theta = \chi_i V$$

Bloklaşmış θ ile V 'yi bu denkleme dönüştürdükten sonra

$$\bar{u}_i \begin{bmatrix} \theta_A & 0 \\ 0 & \theta_B \end{bmatrix} = \chi_i [V_A \quad V_B]$$

elde ederiz. Yani,

$$\begin{bmatrix} \bar{u}_i \begin{bmatrix} \theta_A \\ 0 \end{bmatrix} & \bar{u}_i \begin{bmatrix} 0 \\ \theta_B \end{bmatrix} \end{bmatrix} = [\chi_i V_A \quad \chi_i V_B] \quad (1.39)$$

Denklem (1.39) 'dan $\bar{u}_i \begin{bmatrix} \theta_A \\ 0 \end{bmatrix} = \chi_i V_A$ ve $\bar{u}_i \begin{bmatrix} 0 \\ \theta_B \end{bmatrix} = \chi_i V_B$ 'yi yazabiliriz.

Görebiliriz θ_B sıfıra çok yakın olunca $\bar{u}_i \begin{bmatrix} 0 \\ \theta_B \end{bmatrix} = \chi_i V_B \cong 0$ 'dır.

V_B 'nın değeri $\begin{bmatrix} 0.9963 \\ -0.0526 \\ -0.0676 \end{bmatrix}$ 'yı yukarıdaki yaklaşığın dönüştürdükten sonra

$$0.9963 \chi_{i1} - 0.0526 \chi_{i2} - 0.0676 \chi_{i3} \cong 0$$

elde ederiz. χ_{i1} , χ_{i2} ve χ_{i3} sırasıyla X bağımsız değişken matrisinin i 'nci satırı

χ_i 'nin 1'inci , 2'nci ve 3'üncü terimidir. Dolayısı ile

$$0.9963 x_1 - 0.0526 x_2 - 0.0676 x_3 \cong 0 \quad (1.40)$$

olur. x_1 , x_2 ve x_3 ise 3 tane bağımsız değişkendir, yani X 'in 3 tane sütunudur.

Önceden varsaydığımız x_1 , x_2 ve x_3 aralarındaki tam olmayan çoklu doğrusal

²¹ Mandel J., Use of the Singular Value Decomposition in Regression Analysis, the American Statistician, Vol.36, s.20

bağlantı, (1.40) 'tan daha somut şekilde gözüküyor. Böyle bir ilişki daha önce elde ettiğimiz $\hat{\beta}_1$, $\hat{\beta}_2$ ve $\hat{\beta}_3$ 'ü istatistiksel olarak anlamsız hale getirmiştir.

Bu sorunu aşabilmek için θ_3 (yani θ_B) sıfıra eşitleyelim, yani $XV_B = 0$ olsun.

Dolayısı ile,

$$0.9963x_1 - 0.0526x_2 - 0.0676x_3 = 0$$

elde ederiz . x_3 ise

$$x_3 = 14.7193x_1 - 0.7774x_2 \quad (1.41)$$

olur. Daha önce elde ettiğimiz $\hat{y} = 0.4263x_1 + 2.0839x_2 + 4.6551x_3$ denklemini denklem (1.41) 'i dönüştürerek

$$\hat{y} = 68.9460x_1 - 1.5349x_2 \quad (1.42)$$

elde ederiz. X bağımsız değişken matrisinin SVD 'sini inceleyerek ve kullanarak elde ettiğimiz denklem (1.42) ile SPSS programını kullanarak elde ettiğimiz sonuç arasında nasıl bir ilişki olduğuna bakalım. Aşağıda Tablo 1.3.'te SPSS 'den elde ettiğimiz sonuçlar yer almaktadır.

Tablo 1.3.

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	X2, X3		Enter

a All requested variables entered. b Dependent Variable: Y. Predictors: (Constant), X2, X3

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,955	,913	,878	2,04904

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	220,101	2	110,051	26,211	,002
	Residual	20,993	5	4,199		
	Total	241,094	7			

Coefficients

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	64,024	8,889		7,202	,001
	X2	-1,291	,445	-,822	-2,900	,034
	X3	,413	,793	,148	,521	,624

Tablo1.3.'ün Katsayılar (Coefficients) Tablosundaki verileri kullanarak tahmin edilen regresyon doğrusunu aşağıdaki

$$\hat{y} = 64.024 x_1 - 1.291 x_2 + 0.413 x_3$$

gibi elde ederiz. Bu denklemi, denklem (1.41) 'e dönüştürdükten sonra,

$$\hat{y} = 70.103 x_1 - 1.612 x_2$$

eşitliğini elde ederiz. SPSS ile elde ettiğimiz bu sonucu $\hat{y} = 70.103 x_1 - 1.612 x_2$ eşitliği ile X bağımsız değişken matrisinin SVD 'sini inceleyerek ve kullanarak elde ettiğimiz doğrusal regresyon modeli $\hat{y} = 68.9460 x_1 - 1.5349 x_2$ hemen hemen aynı olduğu görebiliriz. Bu yüzden regresyon analizinde SVD 'nin uygulamasının güvenilir bir yöntem olduğunu da görebiliriz.

1.2.1.2 Tam Çoklu Doğrusal Bağlantının SVD ile İncelenmesi

Bu kısımda bir örnekle bağımsız değişkenlerin aralarında tam çoklu doğrusal bağlantıyı SVD yöntemi ile inceleyeceğiz.

Örnek 1.2. Tablo1.4. 'te çoklu doğrusal bağlantının söz konusu olduğu X bağımsız değişken matrisi ve y bağımlı değişkeni yer almaktadır. Tablo1.5. 'tekiler ise X matrisinin SVD 'si dir.

Tablo 1.4.

Nokta	X			y
	x_1	x_2	x_3	
1	1	16.85	2.36	41.38
2	1	24.81	-3.60	31.01
3	1	18.85	0.86	37.41
4	1	12.63	5.52	50.05
5	1	21.38	-1.03	39.17
6	1	18.78	0.91	38.86
7	1	15.58	3.31	46.14
8	1	16.30	2.77	44.47

Tablo 1.5.

U			θ			V		
u_1	u_2	u_3				v_1	v_2	v_3
0.3238	0.1931	0.9329	524.63 0 0	0 80.364 0	0 0 0	0.0530	0.0638	0.9931
0.4699	-0.5981	-0.0398				0.9974	-0.5140	-0.0492
0.3605	-0.5600	-0.1258				0.0480	0.9966	-0.0661
0.2464	-0.5600	-0.1258						
0.4069	0.6126	-0.2144						
0.3592	-0.2571	-0.0893						
0.3005	0.0012	-0.1289						
0.3137	0.3193	-0.1717						
0.3137	0.2478	-0.1625						

Tablo 1.5.'teki verileri kullanarak $\theta V^T = \begin{bmatrix} 2.7814 & 52.2718 & 2.5179 \\ 0.5132 & -0.4131 & 8.0094 \end{bmatrix}$ 'yı elde

ederiz. Ayrıca Tablo 1.4. ve 1.5.'teki verileri kullanarak denklem (1.25) 'ten

$$\hat{a} = U^T y = \begin{bmatrix} 111.5345 \\ 35.6283 \end{bmatrix}$$

elde ederiz. Dolayısı ile denklem (1.30)'dan

$$\begin{bmatrix} 111.5345 \\ 35.6283 \end{bmatrix} = \begin{bmatrix} 2.7814 & 52.2718 & 2.5179 \\ 0.5132 & -0.4131 & 8.0094 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

eşitliğini elde ederiz. Bu eşitliğinden $\hat{\beta}$ 'nın terimi $\hat{\beta}_i$ 'nin belirli tek bir çözümü

olmadığını fark edebiliriz. Tablo1.5. ‘teki θ ‘nın değerlerinden X bağımsız değişken matrisinin rankı 2 olan bir matris olduğunu de görebiliriz. Burada θV^T ve $\hat{\beta}$ ‘yı ayrı ayrı bloklaştıralım. $Z = \theta V^T$ olsun, dolayısı ile

$$Z = \theta V^T = \left[\begin{array}{cc|c} 2.7814 & 52.2718 & 2.5179 \\ 0.5132 & -0.4131 & 8.0094 \end{array} \right] = [Z_A \mid Z_B]$$

olur. Bloklaşmış $\hat{\beta}$ ise

$$\hat{\beta} = \left[\begin{array}{c} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hline \hat{\beta}_3 \end{array} \right] = \left[\begin{array}{c} \hat{\beta}_A \\ \hline \hat{\beta}_B \end{array} \right]$$

olur. Dolayısı ile denklem (1.30) $\hat{a} = \theta V^T \hat{\beta}$ aşağıdaki gibi yazılabilir,

$$\hat{a} = [Z_A \mid Z_B] \left[\begin{array}{c} \hat{\beta}_A \\ \hline \hat{\beta}_B \end{array} \right] = Z_A \hat{\beta}_A + Z_B \hat{\beta}_B \quad (1.43)$$

Denklem (1.43) ‘ün iki tarafını Z_A^{-1} ‘le çarptıktan sonra

$$Z_A^{-1} \hat{a} = \hat{\beta}_A + Z_A^{-1} Z_B \hat{\beta}_B \quad (1.44)$$

elde ederiz. X bağımsız değişken matrisinin i ‘nci satırı χ_i ve χ_i ‘nin terimleri

sırasıyla χ_{i1}, χ_{i2} ve χ_{i3} olsun. Ayrıca χ_i ‘nin bloklaşmış şekli ise

$$\chi_i = [\chi_{i1} \quad \chi_{i2} \mid \chi_{i3}] = [\chi_{iA} \mid \chi_{iB}]$$

olsun. y bağımlı değişkeninin tahmincisi $\hat{y}$ ‘nın herhangi bir terimi aşağıdaki gibi

$$\hat{y}_i = \chi_i \hat{\beta}$$

olur, yani

$$\hat{y}_i = [\chi_{iA} \quad \chi_{iB}] \begin{bmatrix} \hat{\beta}_A \\ \hat{\beta}_B \end{bmatrix} = \chi_{iA} \hat{\beta}_A + \chi_{iB} \hat{\beta}_B \quad (1.45)$$

dir. Denklem (1.44)'ten elde ettiğimiz $\hat{\beta}_A = Z_A^{-1} \hat{a} - Z_A^{-1} Z_B \hat{\beta}_B$ 'yi denklem (1.45) 'e dönüştürdükten sonra

$$\hat{y}_i = \chi_{iA} (Z_A^{-1} \hat{a}) + (\chi_{iB} - \chi_{iA} Z_A^{-1} Z_B) \hat{\beta}_B \quad (1.46)$$

elde ederiz. $\hat{\beta}$ rassal bir değişken olduğu için denklem (1.46) 'daki $\hat{y}_i$ herhangi bir değere sahip olabilir. Fakat $\hat{y}_i$, y bağımlı değişkeninin terimi y_i 'nin sapmasız bir tahmincisidir ve tek değeri vardı. Bunu garanti altına alabilmek için denklem (1.46)'daki $(\chi_{iB} - \chi_{iA} Z_A^{-1} Z_B) \hat{\beta}_B$ sıfıra eşit olmalıdır. Yani

$$\chi_{iB} = \chi_{iA} Z_A^{-1} Z_B$$

olur. Burada $Z_A^{-1} Z_B = \begin{bmatrix} 15.016 \\ -0.749 \end{bmatrix}$, $\chi_{iA} = [\chi_{i1} \quad \chi_{i2}]$ ve $\chi_{iB} = \chi_{i3}$ 'dir. Dolayısı ile

$$\chi_{i3} = 15.016 \chi_{i1} - 0.749 \chi_{i2}$$

elde ederiz. χ_{i1} , χ_{i2} ve χ_{i3} sırası ile X 'in herhangi bir satırı χ_i 'nin 1'inci , 2'nci ve 3'üncü terimi olduğu için

$$x_3 = 15.016 x_1 - 0.749 x_2 \quad (1.47)$$

eşitliğini elde ederiz. x_1 , x_2 ve x_3 ise, X bağımsız değişken matrisinin 3 tane sütunudur, yani 3 tane bağımsız değişkenidir. Denklem (1.47) ise tam olarak bu 3 bağımsız değişkenin aralarındaki ilişkiyi göstermektedir.

Denklem (1.47) 'nin gösterdiği x_1 ve x_2 ile x_3 aralarındaki ilişkinin üzerine denklem (1.46) 'yı aşağıdaki

$$\hat{y}_i = \chi_{iA} (Z_A^{-1} \hat{a})$$

olarak yazarız. Tablo 1.4. ve 1.5.'teki verileri kullanarak

$$\hat{y}_i = 68.207 \chi_{i1} - 1.496 \chi_{i2}$$

elde ederiz. $\hat{y}_i$, $\hat{y}$ bağımsız değişkeninin herhangi bir terimi olduğu için

$$\hat{y} = 68.207 x_1 - 1.496 x_2 \quad (1.48)$$

yazılabilir. Denklem (1.48), $\hat{y}$ 'nin tek açıklanması değildir, x_1 ile x_3 veya x_2 ile x_3 bağımsız değişken olarak denklem (1.48)'i yeniden yazabiliriz.

Çoklu doğrusal bağlantının söz konusu olduğu bağımsız değişken matrisi ile tahmin edilen regresyon doğrusunu kurmakta en etkili ve kolay yöntemlerden birisi – gereksiz değişkenlerin dışlanması süreci burada SVD ile gerçekleştirildi.²²

1.2.2. SVD ile Elde Edilen Genelleştirilmiş Ters Matrisin Üzerine Parametre Tahminlerinin İncelenmesi

Bu kısımda SVD ile X matrisinin genelleştirilmiş tersi X^G ve Moore – Penrose tersi X^+ 'i açıklayacağız, yanı sıra SVD tarafından açıklanmış X^G ve X^+ sayesinde doğrusal regresyon modelinin parametrelerinin tahmincisi $\hat{\beta}$ ve regresyon tahmincisi $\hat{y}$ 'yı inceleyeceğiz.

1.2.2.1. Genelleştirilmiş Ters Matris ve Moore - Penrose Ters Matrisinin SVD ile Açıklanması

Bir X matrisinin, ($X \in \mathbb{R}^{m \times n}$) aşağıdaki gibi SVD şeklinde açıklanabildiğini geçen kısımda zaten tartıştık:

$$X = U \begin{bmatrix} \theta & 0 \\ 0 & 0 \end{bmatrix} V^T$$

²² Diğer yöntemlerle ilgili ayrıntılar için bakınız.Gujarati Damodar N.(1999) s.340

Burada X matrisinin genelleştirilmiş tersinin X 'in SVD 'nin üzerine nasıl açıklanabildiğini inceleyeceğiz.

$\text{rank}(X) = r$ olsun, dolayısı ile $\theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_r)$ 'dir. $\text{rank}(X) = r$ 'ye göre U ve V^T 'yi bloklaştıralım.

$$U = [U_1 \quad U_2]$$

$$U_1 \in \mathbb{R}^{m \times r}, \quad U_2 \in \mathbb{R}^{m \times (m-r)} \text{ 'dir.}$$

$$V = [V_1 \quad V_2]$$

$$V_1 \in \mathbb{R}^{n \times r}, \quad V_2 \in \mathbb{R}^{n \times (n-r)} \text{ 'dir.}$$

Dolayısı ile X matrisi aşağıdaki gibi

$$X = [U_1 \quad U_2] \begin{bmatrix} \theta & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \theta V_1^T$$

olur. Dolayısı ile, X 'in genelleştirilmiş tersi X^G aşağıdaki gibi ifade edilebilir:

$$\begin{aligned} X^G &= [V_1 \quad V_2] M \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} = [V_1 \quad V_2] \begin{bmatrix} \theta^{-1} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \\ &= V_1 \theta^{-1} U_1^T + V_2 M_{21} U_1^T + V_1 M_{12} U_2^T + V_2 M_{22} U_2^T \end{aligned} \quad (1.49)$$

Buradaki M_{12} , M_{21} ve M_{22} , her biri ayrı ayrı olarak uygun boyutlu rassal matrislerdir. Örneğin, rassal matris $M_{12} \in \mathbb{R}^{r \times (m-r)}$.

Genelleştirilmiş ters matris kavramından biliyoruz ki bir matrisinin genelleştirilmiş tersi X^G ise, $X X^G X = X$ eşitliğini sağlamalıdır. Denklem (1.49) 'la açıklanmış X^G , $X X^G X = X$ eşitliğini sağlayıp sağlamadığını sınavalım,

$$X X^G X = U_1 \theta V_1^T (V_1 \theta^{-1} U_1^T + V_2 M_{21} U_1^T + V_1 M_{12} U_2^T + V_2 M_{22} U_2^T) U_1 \theta V_1^T$$

$$\begin{aligned}
&= U_1 \theta \theta^{-1} U_1^T U_1 \theta V_1^T + U_1 \theta V_1^T V_2 M_{21} U_1^T U_1 \theta V_1^T \\
&\quad + U_1 \theta V_1^T V_1 M_{12} U_2^T U_1 \theta V_1^T + U_1 \theta V_1^T V_2 M_{22} U_2^T U_1 \theta V_1^T \\
&= U_1 \theta V_1^T \\
&= X
\end{aligned}$$

U ile V ortonormal matris olduğu için yukarıdaki denklemde $V_1^T V_2 = 0$, $U_2^T U_1 = 0$ 'dır. Bu yüzden $X X^G X = X$ eşitliğini sağlayan denklem (1.49) 'daki varsaydığımız X^G 'nin X 'in genelleştirilmiş tersi olduğundan eminiz.

Ayrıca, denklem (1.49)'daki X^G 'nin açıklanmasında rassal matrisler M_{12} , M_{21} ve M_{22} de vardır. Bu yüzden rankı r olan ($r \leq \min(m,n)$) $m \times n$ boyutlu X matrisinin sonsuz tane genelleştirilmiş tersi X^G vardır. Ayrıca bu sonsuz tane X^G 'nin, M_{12} , M_{21} ile M_{22} 'nin hepsinin sıfır olduğu özel durumda denklem (1.49) 'da açıklanmış X^G , X matrisinin Moore – Penrose tersi X^+ 'dir.

Moore - Penrose tersinin tanımları için bakınız. Wang Guorong, Wei Yimin and Qiao S. (2003)²³, Campbell S. L. and Meyer C. D. (1991)²⁴ ve Boullion Thomas L. and Odell Patrick L. (1971)²⁵.

Kısaca belirtilirse, aşağıda verilen 4 denklemin hepsini sağlayan X 'in genelleştirilmiş tersi X^G , X 'in Moore - Penrose tersidir.

- (a) $X X^G X = X$
- (b) $X^G X X^G = X^G$
- (c) $(X X^G)^T = X X^G$
- (d) $(X^G X)^T = X^G X$

Şart (a), X 'in herhangi bir genelleştirilmiş tersi X^G tarafından sağlanabilir.

²³ Wang Guorong, Wei Yimin and Qiao S., Generalized Inverse: Theory and Computations, Science Pres, Beijing, 2003, s.2

²⁴ Campbell S.L. and Meyer C.D., Generalized Inverse of Linear Transformations, Dover Publications, New York, 1979, s.9

²⁵ Boullion T. L. and Odell Patrick L., Generalized Inverse Matrics, Wiley & Sons, Inc, New York, 1971, s.1

Şart (b) 'yi sağlayabilmek için ne gerekeceğine bakalım.

Denklem (1.49)'da açıklanan X^G 'nin üzerine

$$X^G X X^G = V_1 \theta^{-1} U_1^T + V_2 M_{21} U_1^T + V_1 M_{12} U_2^T + V_2 M_{21} \theta M_{12} U_2^T$$

$$X^G = V_1 \theta^{-1} U_1^T + V_2 M_{21} U_1^T + V_1 M_{12} U_2^T + V_2 M_{22} U_2^T$$

sonucunu elde ederiz. Yukarıdaki iki eşitliğin denk olması için $M_{21} \theta M_{12} = M_{22}$ olması gerekecektir.

Şart (c) 'nin sağlanabilmesi için ise

$$X X^G = U_1 U_1^T + U_1 \theta M_{12} U_2^T$$

$$(X X^G)^T = U_1 U_1^T + U_2 M_{12}^T \theta U_1^T$$

eşitliklerinin denk olması gerekir. Dolayısıyla $U_1 \theta M_{12} U_2^T = U_2 M_{12}^T \theta U_1^T$ olması gerekir. Bu yüzden $M_{12} = 0$ olmalıdır. Fakat $U_1 \theta M_{12} U_2^T$ bir simetrik matris ise, $M_{12} = 0$ olmadan $U_1 \theta M_{12} U_2^T = U_2 M_{12}^T \theta U_1^T$ tutarlı olabileceğinin düşünülmesi de çok doğaldır. $U_1 \theta M_{12} U_2^T = U_2 M_{12}^T \theta U_1^T$ eşitliğinin sağ ve sol tarafını U_2^T ile çarptıktan sonra, $0 = M_{12}^T \theta U_1^T$ elde ettik. Dolayısı ile $U_1 \theta M_{12} U_2^T = U_2 M_{12}^T \theta U_1^T$ tutarlı olabilmesi için $M_{12} = 0$ olmalıdır.

Şart (d)' ye bakalım.

$$X^G X = V_1 V_1^T + V_2 M_{21} \theta^T V_1^T$$

$$(X^G X)^T = V_1 V_1^T + V_1 \theta M_{21} V_2^T$$

iki eşitliğin denk olması , yani $V_2 M_{21} \theta^T V_1^T = V_1 \theta M_{21} V_2^T$ 'nin tutarlı olması, Bu yüzden $M_{21} = 0$ olmalıdır. Dolayısı ile (a) 'dan (d) 'ye kadar 4 denklemin tutarlı olması için denklem (1.49) gibi X 'in genelleştirilmiş tersi X^G 'nin tanımındaki M_{12} ,

M_{21} ve M_{22} hepsinin sıfıra eşit olması gerekir. Yani $M_{12} = M_{21} = M_{22} = 0$ 'dir. Bu şartı baz alarak ve denklem (1.49) 'u yeniden yazarak X^+

$$X^+ = V_1 \theta^{-1} U_1^T \quad V_1 \in \mathbb{R}^{n \times r}, \quad U_1^T \in \mathbb{R}^{m \times r} \quad (1.50)$$

olarak yazılabilir.

1.2.2.2. Genelleştirilmiş Ters Matris ve Moore – Penrose Ters Matrisi ile Parametre Tahminlerinin Açıklanması

X , sütun rankı tam olan bir matris ise, $y = X\beta + \varepsilon$ doğrusal regresyon modelindeki parametre β 'nın EKKY tahmincisi $\hat{\beta} = (X^T X)^{-1} X^T y$ 'dır.

Ayrıca $\hat{\beta} = (X^T X)^{-1} X^T y$ eşitliğini başka bir yöntemle de sağlayabiliriz.

Amacımız X 'in sütun uzayı $S(X)$ 'in projeksiyonu ile $\hat{\beta}$ 'yı açıklamaktadır.

X 'in sütun uzayı $S(X)$ 'in projektörü $P_{S(X)} = X(X^T X)^{-1} X^T$ olarak tanımlanır. Ayrıca regresyon tahmincisi $\hat{y}$, y vektörünün X 'in sütun uzayı $S(X)$ 'e düşen projeksiyonu diye de biliyoruz. Dolayısı ile $\hat{y} = P_{S(X)} y = X(X^T X)^{-1} X^T y = X\hat{\beta}$ olarak yazılabilir. Fakat X sütun rankı tam olmayan bir matris olursa, $X^T X$ 'in tekil olmasından $P_{S(X)} = X(X^T X)^{-1} X^T$ olarak $P_{S(X)}$ açıklanamaz. Bu durumda EKKY 'yi de kullanamayız. Dolayısı ile daha geniş bir şekilde $P_{S(X)}$ 'i tarif etmemiz gerekiyor, yani Moore - Penrose tersi kavramı kullanmamız gerekir.

Moore - Penrose tersinden biliyoruz ki: XX^+ , X 'in sütun uzayı $S(X)$ 'inin projektörüdür, yani $P_{S(X)} = XX^+$ ²⁶.

Dolayısı ile $\hat{y}$ şu şekilde açıklanabilir.

$$\hat{y} = P_{S(X)} y = XX^T y$$

X $m \times n$ boyutlu bir matris ve $\text{rank}(X) = r$ olsun. Denklem (1.50) 'den

²⁶ Bu konudaki ayrıntılı bilgi için bakınız. Campbell S.L and Meyer C.D. (1991), s.9.

$$\hat{y} = P_{S(X)} y = XX^T y = U_1 \theta V_1^T (V_1 \theta^{-1} U_1^T) y = U_1 U_1^T y \quad U_1 \in \mathbb{R}^{m \times r} \quad (1.51)$$

elde ederiz. Denklem (1.51)'den $\hat{y}$ 'nin $\hat{\beta}$ 'dan etkilenmediğini de görebiliriz. Yani $\hat{\beta}$ ne olsa olsun $\hat{y}$ değişmeden daima denklem (1.51)'deki gibi açıklanabilir.

Gerçekten sütun rankı tam olmayan bağımsız değişken matrisi için sonsuz tane parametre tahmini $\hat{\beta}, \|\hat{\varepsilon}\|^2 = \|y - X\hat{\beta}\|^2$ en küçük olmasını sağlayabilir, yani sonsuz tane en küçük kareler çözüm vardır. Buradaki $\hat{\varepsilon}$ ise ε 'nin tahmincisidir. Bu konu Wang Guorong, Wei Yimin and Qiao S. (2003)²⁷'da daha ayrıntılı tartışılmaktadır.

Bu sonsuz tane $\hat{\beta}$, Z rassal bir matris olmak üzere

$$\hat{\beta} = X^G y + Z = (V_1 \theta^{-1} U_1^T + V_2 M_{21} U_1^T + V_1 M_{12} U_2^T + V_2 M_{22} U_2^T) y + Z \quad (1.52)$$

olarak ifade edilebilir. Dolayısı ile,

$$\begin{aligned} X\hat{\beta} &= X(X^G y + Z) \\ &= U_1 \theta V_1^T ((V_1 \theta^{-1} U_1^T + V_2 M_{21} U_1^T + V_1 M_{12} U_2^T + V_2 M_{22} U_2^T) y + Z) \\ &= (U_1 U_1^T + U_1 \theta M_{12} U_2^T) y + U_1 \theta V_1^T Z \end{aligned}$$

olur. Az önce tartıştık ki $\hat{\beta}$ ne olursa olsun $\hat{y}$ değişmeden daima denklem (1.51)'deki gibi açıklanabilir. Yani $X\hat{\beta} = \hat{y} = U_1 U_1^T y$ 'dir. Dolayısı ile $M_{12} = 0$ ve $Z = V_2 W$ olması gerekir. Burada W rassal bir matristir. $M_{12} = 0$ ve $Z = V_2 W$ olmak üzere denklem (1.52) 'yi yeniden yazarak

$$\hat{\beta} = (V_1 \theta^{-1} U_1^T + V_2 M_{21} U_1^T + V_2 M_{22} U_2^T) y + V_2 W \quad (1.53)$$

elde ederiz. Denklem (1.53) 'te açıklanan $\hat{\beta}$ 'nın sonsuz tane değeri olduğunu görebiliriz ve $\hat{\beta}$ 'nın bu sonsuz tane değerinin hepsi $\hat{y} = X\hat{\beta} = U_1 U_1^T y$ eşitliğini

²⁷ Wang Guorong, Wei Yimin and Qiao S., Generalized Inverse: Theory and Computations, Science Pres, Beijing, 2003, s.5.

sağlar.

Genel olarak doğrusal regresyon modelinin parametre tahmini $\hat{\beta}$, denklem (1.53) 'teki gibi açıklanabilir. Ayrıca X , sütun rankı tam olan bir matris ise (burada kolaylık sağlamak için $m \geq n$ olsun, $V_1 = V$, $V_2 = 0$ olduğu için $\hat{\beta} = V\theta^{-1}U_1^T y$ 'dir. Bu durumdaki X^+ denklem (1.50) 'deki gibi $X^+ = V\theta^{-1}U_1^T$ olarak yazılabilir. Ayrıca sütun rankı tam olan X 'in Moore – Penrose tersi $X^+ = (X^T X)^{-1} X^T$ 'tir. Dolayısı ile $\hat{\beta} = V\theta^{-1}U_1^T y = X^+ y = (X^T X)^{-1} X^T y$ 'dir. Bu sonuç ile EKKY 'den elde ettiğimiz sonuç aynıdır. Gerçekten EKKY 'den elde ettiğimiz parametre tahmini $\hat{\beta} = (X^T X)^{-1} X^T y$, denklem (1.53) 'ün bir özel durumudur.

Burada iki örnekle denklem(1.53) 'ü uygulayalım.

Örnek 1.3. Bağımsız değişken matrisi $X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, bağımlı değişken $y = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$ olsun.

X , sütun rankı tam olan bir matris olduğu için, tahmin edilen regresyon doğrusunun parametresinin tahmini $\hat{\beta}$ EKKY ile açıklanabilir,

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}$$

olur. Ayrıca X matrisinin SVD 'si de şu şekilde açıklanabilir,

$$X = U \begin{bmatrix} \theta \\ 0 \end{bmatrix} V^T = \begin{bmatrix} 0.324 & 0.854 & 0.420 \\ 0.549 & 0.186 & -0.820 \\ 0.774 & -0.490 & 0.410 \end{bmatrix} \begin{bmatrix} 4.07 & 0 \\ 0 & 0.6 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.403 & 0.915 \\ 0.915 & -0.403 \end{bmatrix}^T$$

$$\text{rank}(X) = 2 \text{ olduğu için } V_1 = V, V_2 = 0 \text{ ve } U_1 = \begin{bmatrix} 0.324 & 0.854 \\ 0.549 & 0.186 \\ 0.774 & -0.490 \end{bmatrix} \text{ 'dir.}$$

Denklem (1.53)'ten

$$\hat{\beta} = V\theta^{-1}U_1^T y = \begin{bmatrix} 0.403 & 0.915 \\ 0.915 & -0.403 \end{bmatrix} \begin{bmatrix} 1/4.07 & 0 \\ 0 & 1/0.6 \end{bmatrix} \begin{bmatrix} 0.324 & 0.854 \\ 0.549 & 0.186 \\ 0.774 & -0.490 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}$$

olur.

Yukarıdaki denklem (1.53) 'le elde ettiğimiz $\hat{\beta}$ 'nın değeri ile EKKY 'yi uygulayarak elde ettiğimiz $\hat{\beta}$ 'nın değerinin aynı olduğunu görebiliriz. Çünkü bağımsız eğişken matrisi X , sütun rankı tam olan bir matristir. Fakat X sütun rankı tam olmayan bir matris ise denklem (1.53) 'le sağlanan $\hat{\beta}$ 'nın değeri tek bir tane değil, sonsuz tanedir. $\hat{\beta}$ 'nın bu sonsuz tane değerlerinden herhangi biri denklem (1.51) 'yi yerine getirebilir. Bir örnekle bu durumu inceleyelim.

Örnek 1.4. Bağımsız ve bağımlı değişkenler Tablo 1.4. 'teki gibi olsun. X 'in SVD 'si Tablo 1.5. 'teki gibi olsun.

Geçen kısımdan biliyoruz ki rank (X)= 2 'dir. Burada Tablo 1.5. 'teki X 'in sağ tekil vektör matrisi U ve sol tekil vektör matrisi V 'yi ayrı ayrı bloklaştıralım.

$$U = [U_1 \mid U_2] = \begin{bmatrix} 0.3238 & 0.1931 & 0.9329 \\ 0.4699 & -0.5981 & -0.0398 \\ 0.3605 & -0.5600 & -0.1258 \\ 0.2464 & 0.6126 & -0.2144 \\ 0.4069 & -0.2571 & -0.0893 \\ 0.3592 & 0.0012 & -0.1289 \\ 0.3005 & 0.3193 & -0.1717 \\ 0.3137 & 0.2478 & -0.1625 \end{bmatrix}$$

$$V = [V_1 \mid V_2] = \begin{bmatrix} 0.0530 & 0.0638 & 0.9931 \\ 0.9974 & -0.5140 & -0.0492 \\ 0.0480 & 0.9966 & -0.0661 \end{bmatrix}$$

Denklem (1.53)'ten

$$\begin{aligned}
 \hat{\beta} &= (V_1 \theta^{-1} U_1^T + V_2 M_{21} U_1^T + V_2 M_{22} U_2^T) y + V_2 W \\
 &= (V_1 \theta^{-1} U_1^T) y + (V_2 M_{21} U_1^T + V_2 M_{22} U_2^T) y + V_2 W \\
 &= \begin{bmatrix} 0.3943 \\ 2.0200 \\ 4.4929 \end{bmatrix} + \left(\begin{bmatrix} 0.9931 \\ -0.0492 \\ -0.0661 \end{bmatrix} M_{21} U_1^T + \begin{bmatrix} 0.9931 \\ -0.0492 \\ -0.0661 \end{bmatrix} M_{22} U_2^T \right) y + \begin{bmatrix} 0.9931 \\ -0.0492 \\ -0.0661 \end{bmatrix} W \quad (1.54)
 \end{aligned}$$

olur. Buradaki M_{21} , M_{22} ve W rassal değerlerdir. Denklem (1.54) ise Tablo 1.4. 'teki sütun rankı tam olmayan X , bağımsız değişken matrisi olmak üzere elde ettiğimiz sonsuz tane doğrusal regresyon modelinin parametre tahmini $\hat{\beta}$ 'nın açıklanmasıdır. Bu sonsuz tane $\hat{\beta}$ 'nın değerlerinin hepsi doğrusal regresyon modelinin en küçük kareler çözümü dur. Çünkü denklem (1.54) 'teki $\hat{\beta}$ 'nın hepsi $\hat{y} = X\hat{\beta}$ olmak üzere denklem (1.51) 'i sağlayabilir, yani

$$\begin{aligned}
 \hat{y} &= X\hat{\beta} = U_1 \theta V_1^T ((V_1 \theta^{-1} U_1^T + V_2 M_{21} U_1^T + V_2 M_{22} U_2^T) y + V_2 W) \\
 &= U_1 U_1^T y
 \end{aligned}$$

olur, çünkü $V_1^T V_2 = 0$ 'dır.

Fakat bu sonsuz tane en küçük kareler çözümlerinden bir tanesi $\hat{\beta}$ 'nın minimum - norm en küçük kareler çözümüdür (minimum - norm least squares resolution). Yani

$$\hat{\beta} = V_1 \theta^{-1} U_1^T y = X^+ y.$$

olur ²⁸.

Örnek 1.4. 'ün parametre tahmini $\hat{\beta}$ 'nın minimum - norm en küçük kareler çözümü olduğunu ortaya çıkaralım.

²⁸ Minimum – norm en küçük kareler çözümü ile ilgili ayrıntılı bilgiler için bakınız. Wang Gourong, Wei Yimin and Qiao S. (2003), s.5.

$$\hat{\beta} = V_1 \theta^{-1} U_1^T y = \begin{bmatrix} 0.3943 \\ 2.0200 \\ 4.4929 \end{bmatrix}$$

elde ederiz. $\hat{\beta}$ 'nın bu minimum - norm en küçük kareler çözümü ile tahmin edilen regresyon doğrusunu kurarak

$$\hat{y} = 0.3943x_1 + 2.0200x_2 + 4.4929x_3 \quad (1.55)$$

eşitliğini elde ederiz. Tablo 1.4'teki bağımsız değişkenler aralarında, yani x_1 , x_2 ve x_3 aralarında denklem (1.47) 'deki $x_3 = 15.016x_1 - 0.749x_2$ gibi doğrusal bağlantıların olduğunu biliyoruz. Denklem (1.47) 'yi denklem (1.55) 'e dönüştürdükten sonra

$$\hat{y} = 67.860x_1 - 1.346x_2 \quad (1.56)$$

elde ederiz.

Denklem (1.56) ile denklem (1.48) 'in hemen hemen aynı olduğunu görebiliriz.

Bu rastlantısal bir sonuç değildir. Çünkü her ikisi de $\hat{\beta}$ 'nın minimum - norm en küçük kareler çözümüne dayanmaktadır.

1.2.3. SVD ve Matris Rotasyonu Tekniğini Kullanarak Parametre Tahminlerinin ve Serbestlik Derecesinin Geometrik Açıdan İncelenmeleri

Bu bölümde matris rotasyonu tekniği ile birlikte SVD uygulayarak doğrusal regresyon modelinin parametre tahmini $\hat{\beta}$ ve serbestlik derecesini geometrik açıdan inceleyeceğiz. Bu incelemeyi yürütmeden önce bir örnekle matris rotasyonu kavramından biraz bahsedelim ²⁹

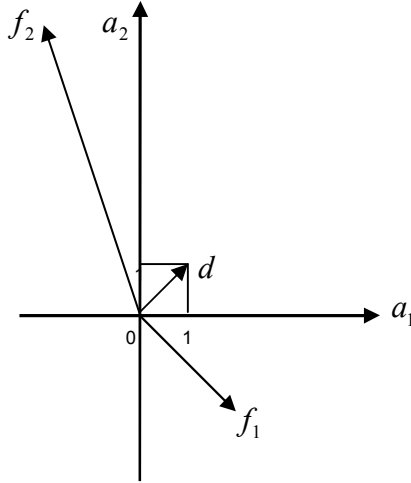
²⁹ Daha ayrıntılı bilgi için bakınız. Green Paul E. (1976), s.127 – 156.

Örnek1.5 $A = \begin{bmatrix} 2 & -2 \\ -2 & 5 \end{bmatrix}$ olsun. A 'nın sütun vektörleri $\begin{bmatrix} 2 \\ -2 \end{bmatrix} = f_1$,

$\begin{bmatrix} -2 \\ 5 \end{bmatrix} = f_2$ olsun. Şekil 1(a)'da a_1 le a_2 ($a_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ve $a_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$) oluşturduğu koordinat

sisteminde bir nokta $d = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ olsun.

A^{-1} bir transformasyon matrisi olarak, d noktasının rotasyonunu nasıl yapacağı ve hangi geometrik anlam ortaya çıkaracağına bakalım.



Şekil 1(a) Örnek 1.5 'teki A 'nın İki Tane f_1 , f_2 Sütun Vektörü ve d Noktasının Standart Koordinat Sistemindeki Konumları

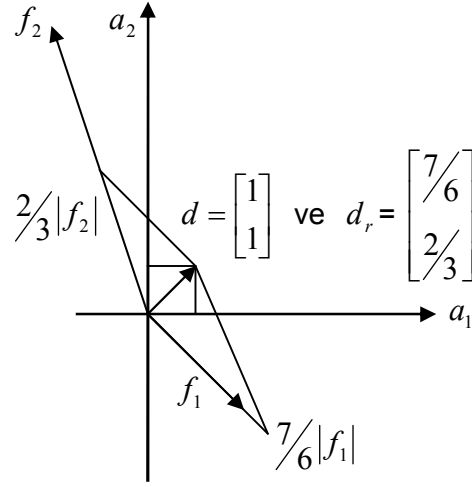
f_1 ekseninin uzunluğu $|f_1| = \sqrt{2^2 + (-2)^2} = \sqrt{8}$, f_2 ekseninin uzunluğu

$|f_2| = \sqrt{5^2 + (-2)^2} = \sqrt{29}$. d noktasının f_1 'le f_2 'nin eksen olarak oluşturduğu

koordinat sistemindeki konumu ise

$$d_r = A^{-1}d = \begin{bmatrix} 5/6 & 2/6 \\ 2/6 & 2/6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7/6 \\ 2/3 \end{bmatrix} \quad (1.57)$$

olur.



Şekil 1(b) Örnek 1.5 'teki f_1 ve f_2 Vektörü Oluşturduğu Uzaydaki d Noktasının Konumu

d_r noktası Şekil 1(b)'de gösterildiği gibi d noktası ile aynı uzay noktasındadır, sadece referans koordinat sistemleri farklı olduğu için, koordinat değerleri farklıdır. d noktası a_1 'le a_2 'nın oluşturduğu koordinat sisteminde açıklanırken d_r noktası, f_1 'le f_2 koordinat sisteminde açıklanmaktadır.

Burada diğer bir bakıştan d noktasının rotasyonunu inceleyelim.

Örnek 1.6 Burada yine Örnek 1.5 'teki A matrisini bir örnek olarak kullanarak

$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ noktasının rotasyonunu A 'nın öz değer ayrışımı (EVD) ile gerçekleştirelim. A

simetrik bir matris olduğu için EVD 'ye ayrıştırılabilir, yani:

$$A = P \theta P^{-1} = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}^{-1}$$

olur. Dolayısı ile

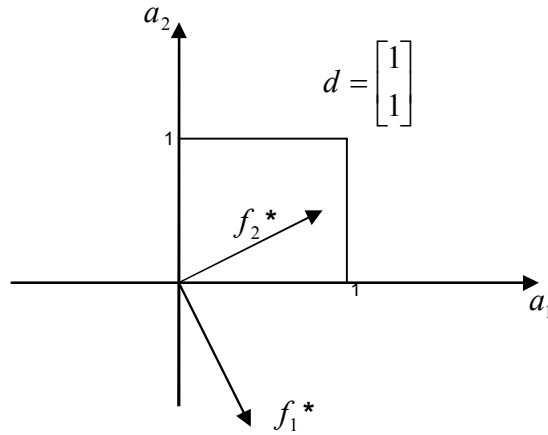
$$A^{-1} = P\theta^{-1}P^{-1} = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/6 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}^{-1} \quad (1.58)$$

$$A \text{ 'nin öz vektörü } \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix} = f_1^*, \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} = f_2^* \text{ olsun. } a_1 \text{ 'le } a_2 \text{ 'nin oluşturduğu}$$

koordinat sisteminde bir nokta d^* olsun. Bu örnekte (Ömek 1.6) kullandığımız

birim uzunluğu Örnek 1.5 'tekenden farklı olduğu için $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ noktasının notasyonu d^*

olarak kullanılmaktadır. d^* noktasının, f_1^* ve f_2^* vektörlerinin ve a_1 ile a_2 'nin oluşturduğu koordinat sistemindeki konumları Şekil 2(a)'da görülmektedir.



Şekil 2(a) Örnek 1.6'daki f_1^* , f_2^* Vektörü ve d^* Noktasının Standard Koordinat Sistemindeki Konumları

Denklem (1.58) 'deki EVD 'nin üzerine A^{-1} 'i denklem (1.57) 'ye dönüştürdükten sonra

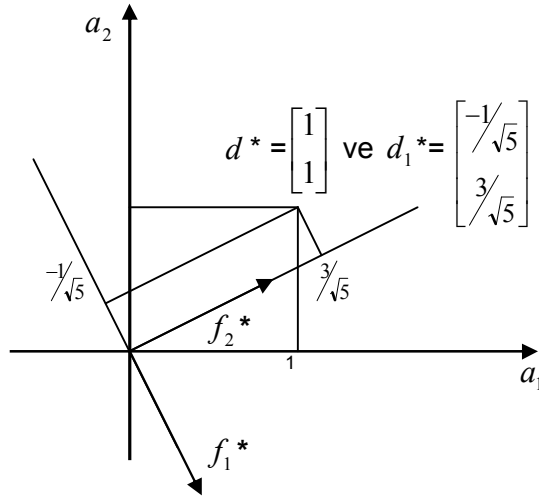
$$d_r = A^{-1} d^* = P\theta^{-1}P^{-1} d^* = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/6 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7/6 \\ 2/3 \end{bmatrix}$$

elde ederiz. Bu eşitlikteki $P\theta^{-1}P^{-1}d^*$ kısmını sağdan sola çarpalım. Yani önce $P^{-1}d^*$, sonra $\theta^{-1}P^{-1}d^*$, son olarak ise $P\theta^{-1}P^{-1}d^*$ çarpımını gerçekleştirelim.

Bu süreci üç aşamaya bölelim.

Rotasyon 1.6.1. $d_1^* = P^{-1}d^* = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{5} \\ 3/\sqrt{5} \end{bmatrix}$

Bu eşitliğin geometrik anlamı ise, a_1 ve a_2 oluşturdukları koordinat sistemindeki nokta $d^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, f_1^* ve f_2^* vektörlerinin oluşturdukları koordinat sistemindeki koordinat konumu ise, $d_1^* = \begin{bmatrix} -1/\sqrt{5} \\ 3/\sqrt{5} \end{bmatrix}$ 'dir. Bu süreç Şekil 2(b)'de görülmektedir.

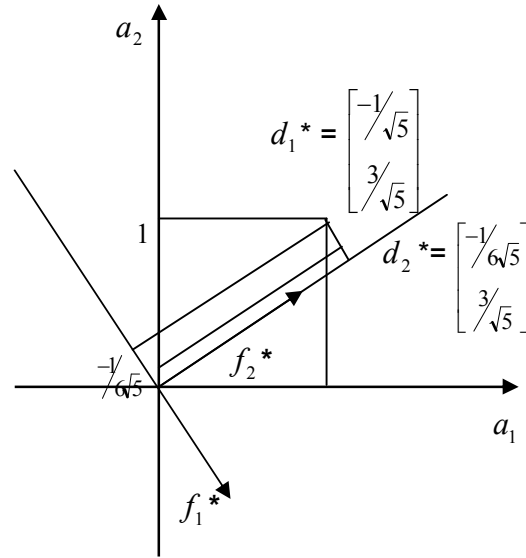


Şekil 2(b) Örnek 1.6'daki d^* Noktasının f_1^* ve f_2^* Vektörlerinin Oluşturduğu Koordinat Sistemindeki Konumu

Rotasyon 1.6.2. $d_2^* = \theta^{-1} P^{-1} d^* = \theta^{-1} d_1^* = \begin{bmatrix} 1/6 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} \\ 3/\sqrt{5} \end{bmatrix} = \begin{bmatrix} -1/6\sqrt{5} \\ 3/\sqrt{5} \end{bmatrix}$

Bu eşitliğin geometrik anlamı ise, uzatma (stretching) matrisi $\theta^{-1} = \begin{bmatrix} 1/6 & 0 \\ 0 & 1 \end{bmatrix}$ ile

d_1^* noktasının koordinat değerleri (f_1^* 'deki $-1/\sqrt{5}$ ve f_2^* 'deki $3/\sqrt{5}$) büyütülerek veya küçültülerek Şekil 2(c) 'de gösterildiği gibi d_2^* noktasına dönüşmesidir.



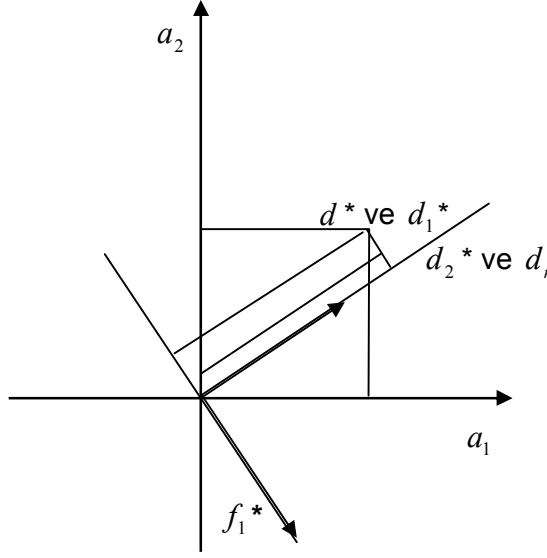
Şekil 2(c) θ^{-1} Transformasyon Matrisi ile Örnek1.6 'daki d_1^* Noktasının f_1^* 'deki ve f_2^* 'deki Koordinat Değerleri Değişerek d_2^* Noktasına Dönüştürülmesi

Şekil 2(c) 'de gösterildiği gibi hem d_1^* hem de d_2^* noktası f_1^* ve f_2^* 'nin oluşturduğu koordinat sisteminde açıklanmaktadır.

Rotasyon 1.6.3. $d_r = P \theta^{-1} P^{-1} d^* = P d_2^* = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} -1/6\sqrt{5} \\ 3/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 7/5 \\ 2/3 \end{bmatrix}$

Bu eşitliğin geometrik anlamı ise, f_1^* ve f_2^* vektörlerin oluşturdukları koordinat sistemindeki d_2^* noktası, a_1 ve a_2 'nin oluşturdukları koordinat sistemi ile

açıklanıyorsa, nokta $d_r = \begin{bmatrix} 7/5 \\ 2/3 \end{bmatrix}$ 'tür. d_r 'nin koordinat konumu Şekil 2(d)'deki gibidir.



Şekil 2(d) Örnek 1.6 'daki d_2^* Noktasının Standart Koordinat Sistemindeki Konumu

Şimdiye kadar simetrik kare matrisin tersi bir transformasyon matris olmak üzere bir noktanın rotasyonunu inceledik. Dikdörtgen matrisin genelleştirilmiş tersi bir transformasyon matrisi olmak üzere nokta rotasyonunun nasıl gerçekleştirilebileceğini de inceleyeceğiz. Aynı anda rotasyon tekniği ve SVD ile serbestlik derecesi ve parametre tahmini $\hat{\beta}$ geometrik açıdan incelenecektir.

Örnek 1.7 $X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ ve $y = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$ bir doğrusal regresyon modelin bağımsız

değişken matrisi ve bağımlı değişken vektörü olsun. X 'nin SVD'si

$$X = U\theta V^T = \begin{bmatrix} 0.324 & 0.854 & 0.420 \\ 0.549 & 0.186 & -0.820 \\ 0.774 & -0.490 & 0.410 \end{bmatrix} \begin{bmatrix} 4.07 & 0 \\ 0 & 0.6 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.403 & 0.915 \\ 0.915 & -0.403 \end{bmatrix}^T$$

olur.

1.2.2. kısmında bağımsız değişken matrisi sütun rankı tam olan bir matris ise parametre tahmininin $\hat{\beta} = X^+ y$ olarak açıklanabildiğini tartıştık. Burada X 'in sütun rankı tam olan bir matris olması için doğrusal regresyon model $y = X \hat{\beta} + \hat{\varepsilon}$ 'nin parametre tahmini $\hat{\beta}$ de şu şekilde tanımlanmalıdır:

$$\hat{\beta} = X^+ y = (U\theta V^T)^+ y = (V^T)^{-1} \theta^+ U^{-1} y \quad ^{30}$$

Bu eşitliğin en sağ tarafı yani $(V^T)^{-1} \theta^+ U^{-1} y$ çarpımını yani y noktasının rotasyonunu üç aşamaya bölelim.

Rotasyon 1.7.1. $y_1 = U^{-1} y = \begin{bmatrix} 0.324 & 0.854 & 0.420 \\ 0.549 & 0.186 & -0.820 \\ 0.774 & -0.490 & 0.410 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.86 \\ -0.12 \\ 1.24 \end{bmatrix} \quad (1.59)$

Ayrıca U 'nun sütunları $\begin{bmatrix} 0.324 \\ 0.549 \\ 0.774 \end{bmatrix} = u_1$, $\begin{bmatrix} 0.854 \\ 0.186 \\ -0.490 \end{bmatrix} = u_2$ ve $\begin{bmatrix} 0.420 \\ -0.820 \\ 0.410 \end{bmatrix} = u_3$

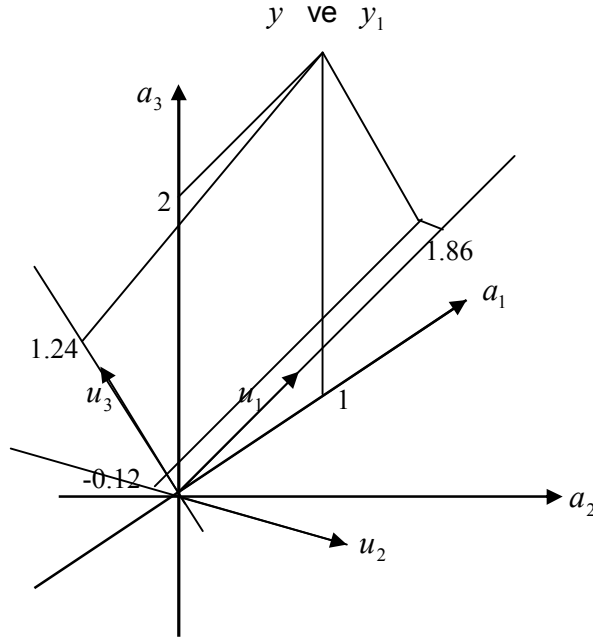
olsunlar.

Denklem (1.59)'un anlamı ise $a_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $a_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ ve $a_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ vektörlerinin

eksenler olarak oluşturdukları koordinat sistemindeki y noktası, u_1 , u_2 ve u_3 'ün oluşturdukları koordinat sistemindeki karşılık gelen koordinat konumu ise,

$y_1 = \begin{bmatrix} 1.86 \\ -0.12 \\ 1.24 \end{bmatrix}$ 'dır. Bu sonuç Şekil 3(a)'da somut şekilde görülmektedir.

³⁰ Denklem(1.91)'deki $(U\theta V^T)^+ = (V^T)^{-1} \theta^+ U^{-1}$ olduğun nedeni ile ilgili ayrıntılar için bakınız. Bu çalışmanın 1.2.2.1 kısmındaki denklem(1.50)'si.



Şekil 3(a) Örnek 1.7'deki u_1 , u_2 ve u_3 Vektörlerinin Oluşturduğu Koordinat Sistemindeki y Noktasının Konumu

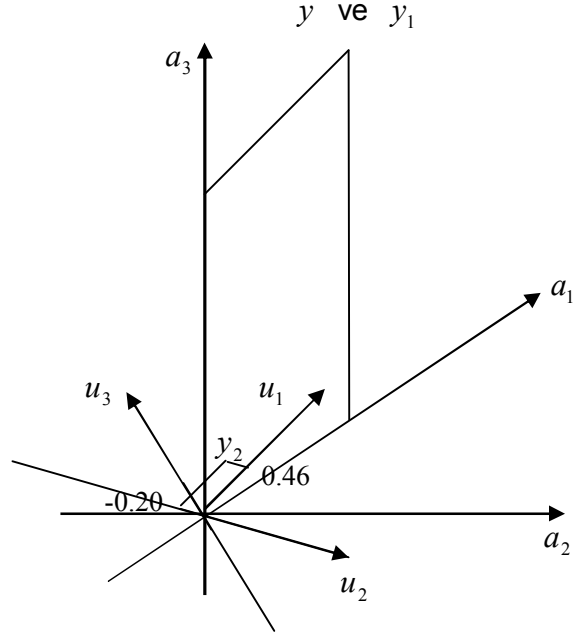
Rotasyon 1.7.2.³¹ $y_2 = \theta^+ U^{-1} y = \theta^+ y_1 = \begin{bmatrix} 4.07 & 0 \\ 0 & 0.6 \\ 0 & 0 \end{bmatrix}^+ \begin{bmatrix} 1.86 \\ -0.12 \\ 1.24 \end{bmatrix} = \begin{bmatrix} 0.46 \\ -0.20 \\ 0 \end{bmatrix}$

Bu aşamada u_1 , u_2 ve u_3 'ün oluşturdukları koordinat sistemindeki y_1 noktası X 'in tekil değer matrisi θ 'nın Moore – Penrose tersi θ^+ bir transformasyon matrisi olmak üzere rotasyon yaparak y_2 noktasına dönüştürülür. Bu süreç Şekil 3(b) 'de görülmektedir. Bu süreçte kullanılan rotasyon tekniği ise, matrisin 'stretching' dir³².

y_1 noktasının y_2 noktasına dönüştüğü süreçte, y_1 'in u_1 eksenindeki koordinat değeri küçülürken, u_2 eksenindeki koordinat değeri büyümektedir. Aynı sırada y_1 'in u_3 eksenindeki koordinat değeri sıfır olmakta, yani y_1 , y_2 noktasına dönüşmekte ve böylece y_1 bir boyutunu kaybetmektedir. Bu süreç Şekil 3(b)'de gösteriliyor.

³¹ θ 'nın Moore – Penrose tersi θ^+ 'ın hesaplanması ile ilgili ayrıntılar için bakınız. Campbell, S.L (1979), s.12 – 19.

³² Ayrıntılar için bakınız. Green Paul E. and Carroll J.Douglas (1976), s.339 – 341



Şekil 3(b) θ^+ Transformasyon Matrisi iken Örnek 1.7 'deki y_1 Noktasının Koordinat Değerlerinin Değişerek y_2 Noktasına Dönüşmesi

y_1 'in bu kayıp boyutu, X bağımsız değişken matrisinin serbestlik derecesinin geometrik bir açıklanmasıdır.

$$\begin{aligned} \text{Rotasyon 1.7.3. } \hat{\beta} &= (V^T)^{-1} \theta^+ U^{-1} y = (V^T)^{-1} y_2 = \left(\begin{bmatrix} 0.403 & 0.915 \\ 0.915 & -0.403 \end{bmatrix}^T \right)^{-1} \begin{bmatrix} 0.46 \\ -0.20 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0.5 \end{bmatrix} \end{aligned} \quad (1.60)$$

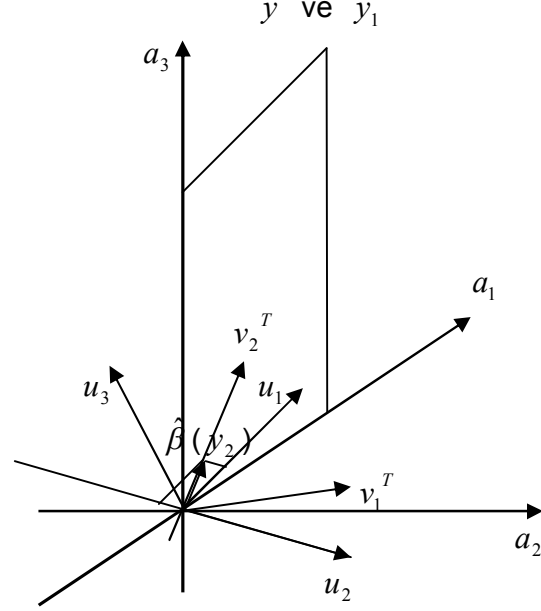
Yukarıdaki $\hat{\beta}$ 'nın sonucu ile EKKY $\hat{\beta} = (X^T X)^{-1} X^T y$ 'yi uygulayarak elde ettiğimiz sonucun aynı olduğu da kolayca ispat edilebilir.

$$X \text{ 'in sağ tekil vektörleri (right singular vectors) } \begin{bmatrix} 0.403 \\ 0.915 \end{bmatrix} = v_1 \text{ ve } \begin{bmatrix} 0.915 \\ -0.403 \end{bmatrix} = v_2$$

olsun. v_1^T ve v_2^T vektörleri, u_1 ve u_2 vektörlerinin oluşturdukları uzayda yer alırlar.

Denklem (1.60) 'in anlamı, u_1 ve u_2 vektörlerinin oluşturdukları koordinat sisteminde

yer alan y_2 noktası , v_1^T ve v_2^T vektörlerinin oluşturdukları koordinat sisteminde açıklanırsa $\hat{\beta}$ noktası olmaktadır. $\hat{\beta}$ noktası (veya vektörü) Şekil 3(c)'de gösteriliyor.



Şekil 3(c) Örnek1.7'deki y_2 Noktası (veya Vektörü), v_1^T ve v_2^T Vektörlerinin Gerdiği Uzayda $\hat{\beta}$ Noktası (veya Vektörü) olarak açıklanması

Şekil 3(c) 'de görüldüğü gibi $\hat{\beta}$ ile y_2 aynı noktada yer alırlar , fakat farklı koordinat sistemde yer alırlar . y_2 noktası u_1 ve u_2 'nin oluşturduğu koordinat sistemindedir ; $\hat{\beta}$ noktası ise v_1^T ve v_2^T 'nin oluşturduğu koordinat sistemindedir.

Burada y noktasının $\hat{\beta}$ 'ya dönüştüğü süreci tekrarlayalım.

Rotasyon 1.7.1.'de a_1 , a_2 ve a_3 ' ün oluşturduğu koordinat sistemindeki y noktası, u_1 , u_2 ve u_3 ' ün oluşturduğu koordinat sisteminde karşılık gelen y_1 noktası 'dır. Bu süreç cebirsel olarak $y_1 = U^{-1} y$ eşitliği ile açıklanıyor ve geometrik olarak Şekil 3(a)'da gösterilmektedir.

Rotasyon 1.7.2.'de u_1 , u_2 ve u_3 oluşturdukları koordinat sistemindeki y_1

noktası, X 'in tekil değer matrisi θ 'nın Moore – Penrose tersi θ^+ bir transformasyon matrisi olmak üzere rotasyon yaparak aynı koordinat sistemindeki y_2 noktasına dönüştü. Bu süreçte y_1 'in u_3 üzerindeki boyutu kayboldu. y_1 'in kaybolan bu boyutu, X bağımsız değişken matrisinin serbestlik derecesinin geometrik bir açıklamasıdır. Bu süreç cebirsel olarak $y_2 = \theta^+ y_1$ eşitliği ile açıklanmakta ve geometrik olarak Şekil 3(b)'de gösterilmektedir.

Rotasyon 1.7.3.'te u_1 le u_2 'nin oluşturduğu koordinat sistemindeki y_2 noktası, v_1^T ile v_2^T 'nin oluşturduğu koordinat sisteminde karşılık gelen $\hat{\beta}$ noktasıdır. Bu süreç cebirsel olarak $\hat{\beta} = (V^T)^{-1} y_2$ eşitliği ile açıklanmakta ve geometrik olarak Şekil 3(c)'de gösteriliyor.

$\hat{\beta}$ 'nın uzaydaki konumunu Rotasyon 1.7.1 , 1.72 ve 1.73 gibi üç aşamalı rotasyon yaparak tespit ettik, $\hat{y}$ noktasının uzayda nerede yer alacağına şimdi bakalım.

1.1.2.2.'de doğrusal regresyon modelinin regresyon tahmini $\hat{y} = U_1 U_1^T y$ olarak açıklanabildiğini zaten tartıştık ³³.

$$\text{Ayrıca bu örnekteki (Ömek 1.7) } U = \begin{bmatrix} 0.324 & 0.854 & 0.420 \\ 0.549 & 0.186 & -0.820 \\ 0.774 & -0.490 & 0.410 \end{bmatrix} \text{ 'dır.}$$

$$\text{Rank}(X) = 2 \text{ olmak üzere, } U_1 = [u_1 \mid u_2] = \begin{bmatrix} 0.324 & 0.854 & 0.420 \\ 0.549 & 0.186 & -0.820 \\ 0.774 & -0.490 & 0.410 \end{bmatrix}$$

³³ $\hat{y} = U_1 U_1^T y$ ile ayrıntılar için bakınız. Bu çalışmanın denklem (1.51) 'i olarak bloklştırılsın.

Dolayısı ile

$$\hat{y} = U_1 U_1^T y = \begin{bmatrix} 0.324 & 0.854 \\ 0.549 & 0.186 \\ 0.774 & -0.490 \end{bmatrix} \begin{bmatrix} 0.324 & 0.854 \\ 0.549 & 0.186 \\ 0.774 & -0.490 \end{bmatrix}^T \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \\ 1.5 \end{bmatrix} \quad (1.61)$$

olur. $\hat{y}$ 'nın bu sonucu ile EKKY yani $\hat{y} = X (X^T X)^{-1} X^T y$ ile elde edeceğimiz sonuç aynıdır. Bu kolayca ispat edilebilir. Ayrıca U_1 ortonormal bir matris olduğu için $U_1^+ = U_1^T$ 'dir ³⁴. Dolayısı ile denklem (1.61) şu şekilde de yazılabilir,

$$\hat{y} = U_1 U_1^+ y = \begin{bmatrix} 0.324 & 0.854 \\ 0.549 & 0.186 \\ 0.774 & -0.490 \end{bmatrix} \begin{bmatrix} 0.324 & 0.854 \\ 0.549 & 0.186 \\ 0.774 & -0.490 \end{bmatrix}^+ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \\ 1.5 \end{bmatrix}$$

Bu eşitliğin üzerine iki aşamalı matris rotasyonu yaparak $\hat{y}$ 'nın geometrik anlamını inceleyelim.

$$\textbf{Rotasyon 1.7.4} \quad y^* = U_1^+ y = \begin{bmatrix} 0.324 & 0.854 \\ 0.549 & 0.186 \\ 0.774 & -0.490 \end{bmatrix}^+ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.86 \\ -0.12 \end{bmatrix} \quad (1.62)$$

Bu aşamada U_1^+ transformasyon matrisi olmak üzere a_1 , a_2 ve a_3 'ün oluşturdukları koordinat sistemindeki y noktası, u_1 ile u_2 'nin oluşturduğu koordinat sistemindeki y^* noktasına dönüşmekte ve bir boyutu kaybolmaktadır. Bu süreç Şekil 3(d) 'de gösterilmektedir.

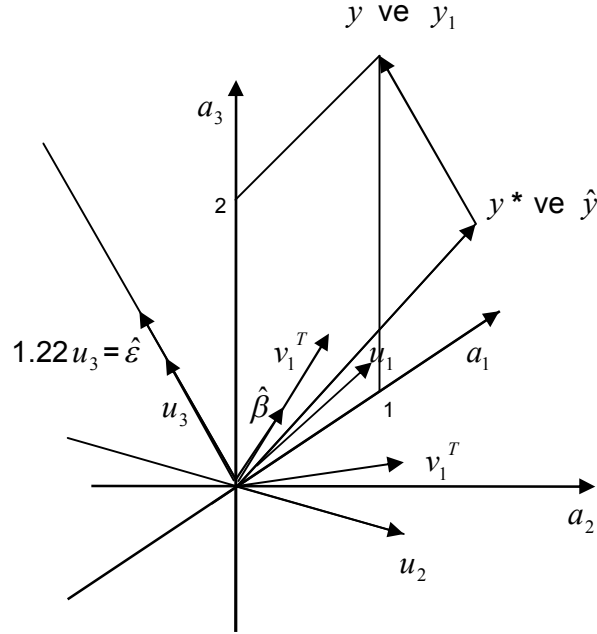
Rotasyon (1.7.1) 'de $y_1 = U^{-1} y$ olmak üzere a_1 , a_2 ve a_3 'ün oluşturdukları koordinat sistemindeki y noktası , u_1 , u_2 ve u_3 'ün oluşturdukları koordinat

sistemindeki karşılık gelen koordinat konumu $y_1 = \begin{bmatrix} 1.86 \\ -0.12 \\ 1.24 \end{bmatrix}$ 'dır. y^* ile y_1 'i

kıyaslayınca iki noktanın u_1 ve u_2 'nin üzerindeki koordinat değeri ile aynı olduğunu

³⁴ U_1 bir orthonormal matris olduğu için U_1^T Moore-Penrose tersinin 4 şartına uyar. Ayrıntılar için bu çalışmanın 31'inci sayfasına bakınız.

fark edebiliriz. Fakat u_3 'ün üzerindeki y^* 'nin koordinat değeri sıfırdır, yani y^* noktası u_1 ile u_2 'nin oluşturduğu düzlemde yer alır. Ayrıca ilginç bir geometrik anlam da ortaya çıkıyor ; y_1 noktasının u_1 ile u_2 'nin oluşturduğu düzleme düşen projeksiyonu y^* noktasıdır. Dolayısı ile y noktasının u_1 ile u_2 'nin oluşturduğu düzleme düşen projeksiyonu da y^* 'dir. Çünkü y ve y_1 aynı uzay noktasındadır, sadece yer aldıkları referans koordinat sistemleri farklıdır.



Şekil 3(d) Örnek 1.7 'deki $\hat{y}$ ve $\hat{e}$ Vektörünün Uzaydaki konumları

y^* ile y_1 noktasının uzaydaki konumu Şekil 3(d)'de görülmektedir.

Ayrıca X matrisinin sol tekil vektörleri olan u_1 ve u_2 , X matrisinin temel sütun uzayı yani sol uzayını (left space) oluştururlar ³⁵. Dolayısı ile y 'nin X matrisinin sol uzayına düşen projeksiyonu ise de y^* 'dir. X matrisinin sol uzayının tamamlayıcı altuzayı (complement subspace) yani sol sıfır uzayı (left null space) u_3 vektörünün tarafından oluşturulmakta olup, X 'in sol sıfır uzayı bir boyutludur.

Bu da Şekil 3(d)'de görülmektedir.

$$\textbf{Rotasyon 1.7.5} \quad \hat{y} = U_1 U_1^+ y = U_1 y^* = \begin{bmatrix} 0.324 & 0.854 \\ 0.549 & 0.186 \\ 0.774 & -0.490 \end{bmatrix} \begin{bmatrix} 1.86 \\ -0.12 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \\ 1.5 \end{bmatrix}$$

Bu eşitliğin geometrik anlamı şudur : u_1 ile u_2 'nin oluşturduğu koordinat sistemindeki y^* noktası , a_1 , a_2 ve a_3 'ün oluşturdukları koordinat sisteminde açıklanıyorsa , $\hat{y}$ 'dır. y^* ile $\hat{y}$ aynı noktadadır, sadece yer aldıkları referans koordinat sistemleri farklıdır. y^* ile $\hat{y}$ arasındaki ilişki Şekil 3(d) 'de daha somut şekilde gösterilmektedir.

Rotasyon 1.7.4 'te u_1 ile u_2 'nin oluşturduğu düzlemde yer alan y^* 'ın , y noktasının bu düzleme düşen projeksiyonu olduğunu gördük. y^* ile $\hat{y}$ aynı uzay noktasında yer almak için $\hat{y}$ da y noktasının u_1 ile u_2 'nin oluşturduğu düzleme düşen projeksiyonudur. Yani $\hat{y}$ noktası, y noktasının X matrisinin sol uzayına düşen projeksiyonudur.

Ayrıca y ile $\hat{y}$ noktasının arasında kalan vektörün , teorik hata payının tahmincisi $\hat{\varepsilon}$ vektörü olduğunu da biliyoruz. $\hat{\varepsilon}$ ise

$$\hat{\varepsilon} = y - \hat{y} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 0.5 \\ 1 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -1 \\ 0.5 \end{bmatrix}$$

olur. Elde ettiğimiz $\hat{\varepsilon}$ 'ya, $\hat{y}$ 'nın dik olduğu açıktır :

$$\hat{y}^T \hat{\varepsilon} = \begin{bmatrix} 0.5 & 1 & 1.5 \end{bmatrix} \begin{bmatrix} 0.5 \\ -1 \\ 0.5 \end{bmatrix} = 0$$

Ayrıca $\hat{\varepsilon}$ vektörünün X matrisinin sol sıfır uzayında yer aldığını da biliyoruz.

³⁵ Ayrıntılar için bakınız. Golub, Gene.H (1996), s.71.

X matrisinin sol sıfır uzayının u_3 vektörü tarafından oluşturulduğunu az önce tartıştık. Dolayısıyla $\hat{\varepsilon}$ ile u_3 aynı uzayda yer alır, yani aynı doğruya yer alır. $\hat{\varepsilon}$ ile u_3 arasındaki ilişki Şekil 3(d) 'de daha somut olarak açıklamaktadır. Cebirsel olarak $\hat{\varepsilon}$ ile u_3 arasındaki ilişki şu şekilde açıklanabilir :

$$\hat{\varepsilon} = \begin{bmatrix} 0.5 \\ -1 \\ 0.5 \end{bmatrix} \cong 1.22 \begin{bmatrix} 0.42 \\ 0.82 \\ 0.41 \end{bmatrix} = 1.22 u_3$$

y 'nin tahmincisi $\hat{y}$ vektörü, parametre tahmini $\hat{\beta}$ vektörü ve X 'in sol sıfır uzayı u_3 vektöründe yer alan teorik hata payının tahmincisi $\hat{\varepsilon}$ vektörü Şekil 3(d)' de görülmektedir. Ayrıca bu örnekte serbestlik derecesinin geometrik anlamı da Şekil 3(b)'de gösterilmektedir.

2. BÖLÜM QR AYRIŞIMI

2.1 QR Ayrışımı Kavramı

Sütun veya satır rankı tam olsun yada olmasın, her türlü matris, sütun rankı tam olan bir matris ile satır rankı tam olan bir matrisin çarpımına ayrıştırılabilir. Böyle bir ayrışım ise matrisin tam rank ayrışımı (full-rank factorization) demektir. Tam rank ayrışımın ispatı Wang Guorong, Wei Yimim and Qiao S. (2003)¹ 'de yer almaktadır.

Bu kısımda matrisin tam rank ayrışımının özel bir durumunu inceleyeceğiz, yani sütun rankı tam olan matrisin bir ortonormal matrisle, satır rankı tam olan bir üst üçgen matrisin çarpımına ayrıştırılmasının üzerinde tartışacağız. Bu ayrışım matrisin QR ayrışımıdır.

Matrisin SVD 'sine benzemeden matrisin QR ayrışımı sadece sütun rank tam olan matrisler için uygundur. Mesela A , $m \times n$ boyutlu bir matris ve $\text{rank}(A) = n$ olsun (Bu bölümdeki her tartışmada $m \geq n$ olsun. $m \leq n$ ise, A^T 'nin QR ayrışımını inceleyeceğiz). A 'nın QR ayrışımı, basitçe $A = QR$ 'dir. Q , $m \times n$ boyutlu bir ortonormal matris, R ise, $n \times n$ boyutlu bir üst üçgen matristir. QR ayrışımı Teorem 2.1. 'de daha ayrıntılı olarak tanımlanmaktadır.

Teorem 2.1. $m \times n$ boyutlu A matrisinin rankı n ise, A şöyle ayrıştırılabilir : $A = QR$ 'dir. Q , $m \times n$ boyutlu ortonormal bir matristir. R ise, köşegen terimleri pozitif olan $n \times n$ boyutlu üst üçgen matristir. Ayrıca Q ve R tek (unique)dir.

Ayrıca Teorem 2.1.'in ispatı aşağıda yer almaktadır². Bu ispat sürecinde kullanılan

¹Wang Guorong, Wei Yimim and Qiao S. (2003),s.4

² Benzeri bir ispat, Stewart G. W. (1998),s.58 'inde bulunmaktadır.

algoritma ise klasik Gram – Schmidt algoritmasıdır³. Bu algoritmadan başka algoritmalar da vardır. Örneğin Householder Triangularization algoritması⁴ ve Wedderburn Rank - One Reduction Formula algoritması⁵.

İspat

İspata kolaylık sağlamak için $m \times 3$ boyutlu X matrisini bir örnek olarak kullanıyoruz. Ayrıca $m \geq 3$ ve $\text{rank}(X) = 3$ olsun. X vektörleri x_1 , x_2 ve x_3 olan ve birbirine lineer bağımsız vektörler olsun. X matrisi $X = [x_1 \ x_2 \ x_3]$ olarak da yazılabilir.

İspatın başlangıcında x_1 'in QR ayrışımı, yani $x_1 = Q_1 R_1$ eşitliğini sağlayan Q_1 ve R_1 'i bulalım.

$x_1 \in \mathbb{R}^{m \times 1}$ 'dir. Bu yüzden $Q_1 \in \mathbb{R}^{m \times 1}$ ve $R_1 \in \mathbb{R}^{1 \times 1}$ olmalı. Dolayısı ile $x_1 = Q_1 R_1$ eşitliğini sağlayan en uygun Q_1 ve R_1 ise, $Q_1 = x_1 / \|x_1\|_2$ ve $R_1 = \|x_1\|_2$ 'dir, yani Q_1 normleştirilmiş olsun. $\|x_1\|_2$ ise, x_1 'in 2- normudur⁶. Ayrıca Q_1 , 2 – norm bir vektördür, yani $\|Q_1\|_2 = 1$ 'dir.

Matris $x^* = [x_1 \ x_2]$ olsun. x^* 'in QR ayrışımını

$$x^* = [x_1 \ x_2] = [Q_1 \ Q_2] \begin{bmatrix} R_1 & p_1 \\ 0 & r_{22} \end{bmatrix} \quad (2.1)$$

olarak varsayalım. Buradaki R_1 , p_1 ve r_{22} 'nin hepsi skaler olsun. Denklem (2.1) 'den

³ Ayrıntılı bilgi için bakınız. Stewart G.W. (1998),s.254 ve Golub Gene.H.(1996),s.230 – 232.

⁴ Stewart G.W. (1998),s.254 - 259 ve Golub Gene.H.(1996),s.224 – 225.

⁵ Ayrıntılı bilgi için bakınız. Chu Moody Y, Funderlic Robert E.and Golub Gene H., A Rank – One Reduction Formula and Its Application to Matrix Factorization, SIAM Review, Vol. 37, No.4, s.512 – 530.

⁶ Ayrıntılı bilgi için bakınız. Golub Gene.H. (1996),s.52.

$$\begin{cases} x_1 = Q_1 R_1 \\ x_2 = Q_1 p_1 + Q_2 r_{22} \end{cases} \quad (2.2)$$

elde ederiz. Ayrıca denklem (2.2) 'den

$$Q_2 r_{22} = x_2 - Q_1 p_1 \quad (2.3)$$

bulunur. Denklem (2.3) 'ün iki tarafını Q_1^T ile çarptıktan sonra ise

$$Q_1^T Q_2 r_{22} = Q_1^T x_2 - Q_1^T Q_1 p_1 \quad (2.4)$$

olur.

$Q_1^T Q_1 = I$ olduğu için denklem (2.4) şu şekilde yazılabilir,

$$Q_1^T Q_2 r_{22} = Q_1^T x_2 - Q_1^T Q_1 p_1 = Q_1^T x_2 - p_1$$

$p_1 = Q_1^T x_2$ olsun. Dolayısı ile

$$Q_1^T Q_2 r_{22} = Q_1^T x_2 - p_1 = p_1 - p_1 = 0 \quad (2.5)$$

olur. Denklem (2.5) 'ten görüldüğü gibi $p_1 = Q_1^T x_2$ ise , Q_1 'le $Q_2 r_{22}$ arasında ortogonal ilişki vardır, yani Q_1 vektörü ile Q_2 vektörü ortogonaldır, çünkü r_{22} bir skalerdir. Fakat Q_2 sıfıra eşit olursa, $Q_1^T Q_2 r_{22}$ de sıfıra eşit olacağından Q_1 'le $Q_2 r_{22}$ arasında ilişki tespit edilemez. Şimdi Q_2 'nin sıfıra eşit olup olmadığına bakalım. $x_1 = Q_1 R_1$ 'den

$$Q_1 = x_1 R_1^{-1} \quad (2.6)$$

elde ederiz.

Denklem (2.6) 'yı denklem (2.3)'e dönüştürdükten sonra,

$$Q_2 r_{22} = x_2 - x_1 (R_1^{-1} p_1) \quad (2.7)$$

elde ederiz. Bu ispatın başlangıcında x_2 ile x_1 'in birbirleriyle lineer bağımsız olduğunu belirtmiştik. Bu yüzden denklem (2.7) 'deki $Q_2 r_{22}$ sıfıra eşit değildir.

Dolayısı ile Q_2 de sıfır değildir. Bu yüzden $p_1 = Q_1^T x_2$ iken Q_1 ile Q_2 ortogonal olmalıdır.

Q_1 ile Q_2 'nin ortogonalliği p_1 , $p_1 = Q_1^T x_2$ olarak belirtilebilir. Şimdi denklem (2.1) 'de belirsiz kalan Q_2 ve r_{22} 'yi tespit edelim.

$p_1 = Q_1^T x_2$ 'yi denklem (2.3) 'e dönüştürdükten sonra

$$Q_2 r_{22} = x_2 - Q_1 Q_1^T x_2 \quad (2.8)$$

elde ederiz. Denklem (2.8)' den normalleştirilmiş olan Q_2 aşağıdaki gibi olur :

$$Q_2 = Q_2 r_{22} / r_{22} = (x_2 - Q_1 Q_1^T x_2) / \|x_2 - Q_1 Q_1^T x_2\|_2 \quad (2.9)$$

r_{22} ise,

$$r_{22} = \|Q_2 r_{22}\|_2 = \|x_2 - Q_1 Q_1^T x_2\|_2 \quad (2.10)$$

olur. Şimdiye kadar denklem (2.1) 'deki Q_2 , p_1 ve r_{22} 'nin hepsi tespit edildi ve bunların üzerine Q_1 'le Q_2 'nin ortogonalliği de garantilemektedir.

$$X^* = [x_1 \ x_2] = [Q_1 \ Q_2] \begin{bmatrix} R_1 & p_1 \\ 0 & r_{22} \end{bmatrix} = Q^* R^*$$

olsun. Buradaki $Q^* = [Q_1 \ Q_2]$ bir ortonormal matristir, çünkü hem Q_1 hem Q_2 2-

norm vektördür ve birbirleri ile ortogonaldır. $R^* = \begin{bmatrix} R_1 & p_1 \\ 0 & r_{22} \end{bmatrix}$ 'dir ve köşegenindeki

terim R ve r_{22} 'nin hepsi pozitiftir.

Şimdi $X = [x_1 \ x_2 \ x_3]$ 'e, QR ayrışımının nasıl uygulanabileceğine bakalım.

X matrisinin QR ayrışımının denklem (2.11) 'deki gibi olduğunu varsayalım.

$$X = [x_1 \ x_2 \ x_3] = [X^* \ x_3] = [Q^* \ Q_3] \begin{bmatrix} R^* & p_2 \\ 0 & r_{33} \end{bmatrix} \quad (2.11)$$

Denklem (2.11)'den

$$\begin{cases} X^* = Q^* R^* \\ x_3 = Q^* p_2 + Q_3 r_{33} \end{cases} \quad (2.12)$$

elde ederiz. Denklem (2.12)' den

$$Q_3 r_{33} = x_3 - Q^* p_2 \quad (2.13)$$

olur. Burada Q^* ile Q_3 'ün ortogonal olması için ne gerekeceğine bakalım.

Denklem (2.13)'ün iki tarafını Q^{*T} ile çarptıktan sonra

$$Q^{*T} Q_3 r_{33} = Q^{*T} x_3 - Q^{*T} Q^* p_2 \quad (2.14)$$

elde ederiz. $Q^{*T} Q^* = I$ olduğu için denklem (2.14) şu şekilde de yazılabilir :

$$Q^{*T} Q_3 r_{33} = Q^{*T} x_3 - p_2 \quad (2.15)$$

Denklem (2.15)'ten görüldüğü gibi Q^* ile Q_3 'ün ortogonal olabilmesi için , yani Q_1 ve Q_2 'ile Q_3 'ün ortogonal olabilmesi için denklem (2.15) 'in sağ tarafı sıfıra eşit olmalı, yani $Q^{*T} x_3 - p_2 = 0$ olmalıdır. Buradan da

$$Q^{*T} x_3 = p_2$$

olur ve $p_2 \in \mathbb{R}^2$ dir. Ayrıca Q_3 ve r_{33} şu şekilde de açıklanabilir :

$$Q_3 = Q_3 r_{33} / r_{33} = (x_3 - Q^* Q^{*T} x_3) / \|x_3 - Q^* Q^{*T} x_3\|_2 \quad (2.16)$$

$$r_{33} = \|Q_3 r_{33}\|_2 = \|x_3 - Q^* p_2\|_2 = \|x_3 - Q^* Q^{*T} x_3\|_2 \quad (2.17)$$

Şimdiye kadar denklem (2.11) 'deki Q_3 , p_2 ve r_{33} 'ün hepsi tespit edildi ve bunların üzerine Q^* ile Q_3 'ün ortogonal olduğu garanti altına alınır. Dolayısı ile denklem (2.11) aşağıdaki gibi yeniden yazılabilir :

$$X = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} = \begin{bmatrix} X^* & x_3 \end{bmatrix} = \begin{bmatrix} Q^* & Q_3 \end{bmatrix} \begin{bmatrix} R^* & p_2 \\ 0 & r_{33} \end{bmatrix}$$

$$= [Q_1 \quad Q_2 \quad Q_3] \left[\begin{array}{cc|c} R_{11} & p_1 & p_2 \\ 0 & r_{22} & \\ \hline 0 & 0 & r_{33} \end{array} \right] = QR \quad (2.18)$$

Burada $p_2 = Q^{*T} x_3$ olduğu için p_2 2 boyutlu bir vektördür. Ayrıca

$$Q = [Q_1 \quad Q_2 \quad Q_3] \text{ bir ortonormal matristir ; } R = \left[\begin{array}{cc|c} R_{11} & p_1 & p_2 \\ 0 & r_{22} & \\ \hline 0 & 0 & r_{33} \end{array} \right] \text{ ise köşegenindeki}$$

terimleri pozitif olan üst üçgen matristir.

Aynı mantıkla her boyutlu sütun rankı tam olan matris denklem (2.18) 'deki X matrisinin QR ayrışımı gibi bir ortonormal matrisi ile köşegen terimleri pozitif olan bir üst üçgen matrisin çarpımına ayrıştırılabilir.

Denklem (2.18)'deki X 'in herhangi bir sütunu x_i ,

$$x_i = Q r_i$$

olarak da açıklanabilir. r_i ise, R 'nın i 'inci sütunudur. Bu eşitlikten X 'in herhangi bir sütunu x_i , Q tarafından açıklanabilir, yani Q , X matrisinin sütun temel uzayını oluşturduğunu görebiliriz.

Teorem 1.2. 'yi ispat ettiğimiz süreçte kullandığımız Gram–Schmidt algoritması matrisin QR ayrışımının hesaplanmasında da kullanılabilen ideal bir algoritmadır. Bunun dışında Householder, Block Householder, Givens ve Fast Givens QR ayrışımının hesaplanması için diğer iyi algoritmalar⁷. Bu algoritmaların arasındaki Fast Givens algoritması LU ayrıştırması ile QR ayrıştırılması arasındaki ilginç ilişkiyi ortaya koyalım :

A , sütun rankı tam olan bir matris ve $A \in \mathbb{R}^{m \times n}$ olsun.

⁷ Golub, Gene.H. (1996),s.223

$A^T A$ 'nın LU ayrışımı $A^T A = LU$ olsun. L , köşegeninin terimleri '1' olan alt üçgen matristir. U ise tekil olmayan üst üçgen matristir. Ayrıca $\hat{A} = A (L^{-1})^T$ olur. Dolayısı ile

$$\hat{A}^T \hat{A} = L^{-1} A^T A (L^{-1})^T = L^{-1} LU (L^{-1})^T = U (L^{-1})^T$$

yazılabilir.

$\hat{A}^T \hat{A} = D$ olsun. D hem simetrik hem üst üçgen bir matris olduğu için bir köşegen matris olmalı. A 'nın QR ayrışımındaki Q ve R aşağıdakiler gibi hesaplanabilir ⁸ :

$$Q = \hat{A} D^{-1/2} = A(L^{-1})^T (U(L^{-1})^T)^{-1/2} \quad (2.19)$$

$$R = D^{1/2} L^T = (U(L^{-1})^T)^{1/2} L^T \quad (2.20)$$

Burada QR ile SVD arasında nasıl bir ilişki olduğunu da inceleyelim.

Yine A 'yı bir örnek olarak kullanalım. $A = QR$, A 'nın QR ayrışımı olsun.

Dolayısı ile $Q^T A = R$ 'dir. Ayrıca R 'nin SVD ayrışımı

$$W^T R V = \theta \quad (2.21)$$

olsun. θ bir köşegen matristir. $Q^T A = R$ 'yi denklem (2.21)'e dönüştürdükten sonra,

$$W^T Q^T A V = \theta \quad (2.22)$$

elde ederiz. $Z = QW$ olsun. Denklem (2.22) şu şekilde yeniden yazılabilir :

$$Z^T A V = \theta \quad (2.23)$$

Denklem (2.21) ve (2.23) 'ten A ile R 'nin aynı tekil değer matrisi ve temel satır uzayına sahip olduğunu görebiliriz .

⁸ Ayrıntılı bilgi için bakınız. Golub Gene.H. (1996), s.228 ve Pursell Lyle and Trimble S.Y., American Mathematical Monthly, Vol.98, No.6, s.544 – 549.

2.2. Regresyon Analizine QR Ayırışımın Uygulanması

1.2.1. kısmında SVD uygulayarak çoklu doğrusal bağlantıyı inceledik ve EKKY 'nin yerine SVD ile doğrusal regresyon modeli kurduk. Ayrıca doğrusal regresyon modelinin kurulmasında QR ayırışım yöntemi⁹ ile EKKY 'nin temeli aynıdır, yani ikisi de kalıntı hata kareler toplamını en küçük yapacak parametre tahminini tespit etmektedir. Bu kısımda QR ayırışım yöntemi ile regresyon analizini gerçekleştirerek EKKY ile QR ayırışım yönteminin denk olduğunu göstereceğiz.

Bir doğrusal regresyon modeli aşağıdaki gibi açıklanmış olsun :

$$y = X\beta + \varepsilon \quad (2.24)$$

Bu modeldeki X bağımsız değişkeni sütun rankı tam olan bir matris ve $X \in \mathbb{R}^{m \times n}$, $m \geq n$ olsun. Ayrıca X 'in QR ayırışımı $X = Q_1 R$ ve $Q_1 \in \mathbb{R}^{m \times n}$, $R \in \mathbb{R}^{n \times n}$ olsun. Q_1 'in $\mathbb{R}^m$ vektör uzayındaki ortogonal tamamlayıcısı (complement) Q_2 olsun ve $Q_2 \in \mathbb{R}^{m \times (m-n)}$ 'dir. X 'in QR ayırışımı şu şekilde de yazılabilir :

$$X = Q_1 R = [Q_1 \quad Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} = Q \begin{bmatrix} R \\ 0 \end{bmatrix}$$

Burada $Q = [Q_1 \quad Q_2]$ 'dir. Ayrıca Q kare ortogonal matris olduğu için

$$Q^T X = \begin{bmatrix} R \\ 0 \end{bmatrix} \quad (2.25)$$

olur. Ayrıca denklem (2.25) 'e Stewart G. W. (1998) 'de QR Decomposition denmektedir ¹⁰.

⁹ Bağımsız değişken matrisinin QR ayırışımı incelenerek regresyon analizinin yürütülmesi burada geçici olarak QR ayırışım yöntemine densin.

¹⁰ Stewart G. W. (1998), s.250.

Denklem (2.24) 'ün iki tarafı Q_1^T ile çarptıktan sonra

$$Q^T y = Q^T X\beta + Q^T \varepsilon$$

elde ederiz. Bu denklemi $Q^T X = \begin{bmatrix} R \\ 0 \end{bmatrix}$ 'a dönüştürdükten sonra

$$Q^T y = \begin{bmatrix} R \\ 0 \end{bmatrix} \beta + Q^T \varepsilon \quad (2.26)$$

elde ederiz. $Q^T y = z$ ve $Q^T \varepsilon = \eta$ olsun. Dolayısı ile denklem (2.26) şu şekilde yazılabilir :

$$z = \begin{bmatrix} R \\ 0 \end{bmatrix} \beta + \eta \quad (2.27)$$

Denklem (2.27) 'nin hata kareler toplamı (ESS) aşağıdaki gibidir :

$$\begin{aligned} \|\eta\|_2^2 &= \left\| z - \begin{bmatrix} R \\ 0 \end{bmatrix} \beta \right\|_2^2 \\ &= \|Q^T y - Q^T X\beta\|_2^2 \\ &= \|Q^T (y - X\beta)\|_2^2 \\ &= (y - X\beta)^T Q Q^T (y - X\beta) \\ &= (y - X\beta)^T Q Q^{-1} (y - X\beta) \\ &= \|y - X\beta\|_2^2 \\ &= \|\varepsilon\|_2^2 \end{aligned}$$

Ayrıca denklem (2.24) ' ile denklem (2.27) 'nin teorik hata payının varyansı da aynıdır, yani $Var(\eta_i) = Var(\varepsilon_i)$ 'dir. Bunu gösterelim :

Q 'nün i 'nci sütununun notasyonu Q_i olsun. $Q^T \varepsilon = \eta$ olduğu için $Q_i^T \varepsilon = \eta_i$ 'dir. Dolayısı ile

$$\begin{aligned}
 Var(\eta_i) &= Var(Q_i^T \varepsilon) \\
 &= E\left(\sum_{j=1}^m Q_{ij} \varepsilon_j\right)^2 \\
 &= E\left(\sum_{j=1}^m Q_{ij}^2 \varepsilon_j^2\right) + 2E\left(\sum_{j=1}^m \sum_{k=1}^m Q_{ij} \varepsilon_j Q_{ik} \varepsilon_k\right) \quad j \neq k \\
 &= \sum_{j=1}^m Q_{ij}^2 E(\varepsilon_j^2) \\
 &= Var(\varepsilon_j) \sum_{j=1}^m Q_{ij}^2 \\
 &= Var(\varepsilon_j)
 \end{aligned} \tag{2.28}$$

olur. Denklem (2.28) 'den görebiliriz ki $Var(\eta_i) = Var(\varepsilon_i)$, yani $\varepsilon_i \sim N(0, \delta^2)$ ve aynı zamanda $\eta_i \sim N(0, \delta^2)$ 'dir. Diğer bir deyişle η_i ile ε_i 'nin olasılık dağılımı aynıdır .

Dolayısı ile hem ESS hem teorik hata payının varyansı açısından denklem (2.24) 'ün denklem (2.27) 'ye denk olduğu görülebilir. Şimdi denklem (2.24) 'ün yerine denklem (2.27) 'yi kullanarak regresyon analizini gerçekleştirelim :

$$\begin{aligned}
 z &= \begin{bmatrix} Q_1^T y \\ Q_2^T y \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \quad z_1 \in \mathbb{R}^n, \quad z_2 \in \mathbb{R}^{(m-n)} \text{ ve} \\
 \eta &= \begin{bmatrix} Q_1^T \varepsilon \\ Q_2^T \varepsilon \end{bmatrix} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}, \quad \eta_1 \in \mathbb{R}^n, \quad \eta_2 \in \mathbb{R}^{(m-n)} \text{ olsun.}
 \end{aligned}$$

Yukardaki bloklaşmış z ve η 'nın üzerine denklem (2.27) 'yi yeniden yazalım,

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} R \\ 0 \end{bmatrix} \beta + \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad (2.29)$$

Denklem (2.29) 'daki β 'nın sapmasız tahmincisi $\hat{\beta}$ olsun, dolayısı ile

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} R \\ 0 \end{bmatrix} \hat{\beta} + \begin{bmatrix} \hat{\eta}_1 \\ \hat{\eta}_2 \end{bmatrix} \quad (2.30)$$

olur. $\hat{\eta}_1 \in \mathbb{R}^n$, $\hat{\eta}_2 \in \mathbb{R}^{(m-n)}$ 'dir.

Denklem (2.30) 'daki ESS ise,

$$\begin{aligned} \text{ESS} &= \left\| \begin{bmatrix} \hat{\eta}_1 \\ \hat{\eta}_2 \end{bmatrix} \right\|_2^2 = \left\| \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} - \begin{bmatrix} R \\ 0 \end{bmatrix} \hat{\beta} \right\|_2^2 \\ &= (z_1 - R\hat{\beta})^T (z_1 - R\hat{\beta}) + z_2^T z_2 \end{aligned} \quad (2.31)$$

Bu denklemden görebiliriz ki en küçük ESS 'yi sağlayabilmek için $z_1 - R\hat{\beta} = 0$,

yani $z_1 = R\hat{\beta}$ olması gerekiyor. Bu yüzden

$$\hat{\beta} = R^{-1} z_1 \quad (2.32)$$

'dir. Gerçekten QR ayrışım yöntemini uygulayarak denklem (2.32) 'deki gibi elde ettiğimiz $\hat{\beta}$ ile EKKY ile elde ettiğimiz sonuç aynıdır. Çünkü

$$\hat{\beta} = R^{-1} z_1 = (R^T R)^{-1} R^T z_1 = (R^T Q_1^T Q_1 R)^{-1} R^T Q_1^T y = (X^T X)^{-1} X^T y$$

'dir.

Bu sonuç çok doğaldır, çünkü hem QR ayrışım yöntemi hem EKKY 'nin taban aldığı teori, hata kareler toplamının en küçük değerinin sağlamasıdır. Aynı kriter üzerine iki yöntemle elde ettiğimiz sonuçlar aynıdır, sadece işleme süreçleri

farklıdır. EKKY, $\frac{\partial(\|\hat{\epsilon}\|_2^2)}{\partial \hat{\beta}} = 0$ eşitliğini sağlayarak $\hat{\beta}$ 'yı elde etmektedir, yani $\hat{\beta}$ bir

değişken olmak üzere $\|\hat{\epsilon}\|_2^2 = y^T y - 2y^T X\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta}$ iki kare fonksiyonunun

eğrisindeki dip noktayı bularak $\hat{\beta}$ 'yı sağlamaktadır. Fakat QR ayrışım yöntemi,

denklem (2.31) deki en küçük ESS 'yi sağlayarak $\hat{\beta}$ 'yı elde etmektedir.

Şimdi $\hat{\beta}$ 'nın varyansını türetelim. Denklem (2.29) 'dan

$$z_1 = R\beta + \eta_1$$

olur. Bunu denklem (2.32) 'ye dönüştürdükten sonra

$$\hat{\beta} = R^{-1}(R\beta + \eta_1)$$

elde ederiz. Dolayısı ile $\hat{\beta} - \beta = R^{-1}\eta_1$ 'dır. β ile $\hat{\beta}$ 'nın i 'nci teriminin notasyonu

ayrı ayrı olarak $\hat{\beta}_i$ ve β_i olsun. Ayrıca $R^{-1} = C$ ve C 'nin i 'nci satırı c_i olsun.

$$\begin{aligned} Var(\hat{\beta}_i) &= E(\hat{\beta}_i - \beta_i)^2 \\ &= E(c_i \eta_1)^2 \\ &= E(c_i \eta_1)^2 \\ &= E\left(\sum_j^n c_{ij} \eta_{1j}\right)^2 \\ &= E\left(\sum_j^n c_{ij}^2 \eta_{1j}^2 + 2 \sum_{j=1}^n \sum_{k=1}^n c_{ij} \eta_{1j} c_{ik} \eta_{1k}\right) \quad j \neq k \\ &= E\left(\sum_j^n c_{ij}^2 \eta_{1j}^2\right) \\ &= \sigma^2 \sum_j^n c_{ij}^2 \\ &= \sigma^2 (c_i c_i^T) \end{aligned} \tag{2.33}$$

olur. $\hat{\beta}_i$ ve $\hat{\beta}_j$ arasındaki kovaryans ise,

$$\begin{aligned} Kov(\hat{\beta}_i, \hat{\beta}_j) &= E(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j) \quad i \neq j \\ &= E(c_i \eta_1)(c_j \eta_1) \end{aligned}$$

$$\begin{aligned}
&= E\left(\sum_{k=1}^n c_{ik} \eta_{1k}\right) \left(\sum_{k=1}^n c_{jk} \eta_{1k}\right) \\
&= E\left(\sum_{k=1}^n c_{ik} c_{jk} \eta_{1k}^2 + \sum_{k=1}^n \sum_{p=1, p \neq k}^n c_{ik} c_{jp} \eta_{1k} \eta_{1p}\right) \quad k \neq p \\
&= E\left(\sum_{k=1}^n c_{ik} c_{jk} \eta_{1k}^2\right) \\
&= \sum_{k=1}^n c_{ik} c_{jk} E(\eta_{1k}^2) \\
&= \delta^2 \sum_{k=1}^n c_{ik} c_{jk} \\
&= \delta^2 c_i c_j^T \quad (2.34)
\end{aligned}$$

olur. Denklem (2.33) 'le (2.34) 'ü birleştirdikten sonra $\hat{\beta}$ 'nın varyans - kovaryans matrisini

$$Var\ Kov(\hat{\beta}) = \delta^2 C C^T = \delta^2 R^{-1} (R^{-1})^T = \delta^2 (R^T R)^{-1}$$

olarak elde ederiz. Gerçekten yukarıdaki $Var - Kov(\hat{\beta})$ 'nın tanımı ile alıştığımız

$Var\ Kov(\hat{\beta}) = \delta^2 (X^T X)^{-1}$ formülü aynıdır, çünkü

$$Var\ Kov(\hat{\beta}) = \delta^2 (R^T R)^{-1} = \delta^2 (R^T Q_1^T Q_1 R)^{-1} = \delta^2 ((Q_1 R)^T Q_1 R)^{-1} = \delta^2 (X^T X)^{-1}$$

'dir. Burada QR ayrışım yöntemi ile regresyon analizi yaparak elde ettiğimiz denklem (2.32)'deki $\hat{\beta}$ 'nın açıklanması ve denklem (2.33)'teki $Var(\hat{\beta}_i)$ açıklanmasını bir örnekte kullanacağız .

Örnek 2.1 $y = X\hat{\beta} + \hat{\varepsilon}$ doğrusal regresyon modeli bağımsız değişken X ve bağımlı değişken y olmak üzere¹¹,

¹¹ Buradaki bağımsız değişken X ve bağımlı değişken y verileri Ansley Craig F.(1985) 'ten iktibas edilmektedir.

$$y = \begin{bmatrix} 2.99 \\ 5.69 \\ 7.02 \\ 9.45 \\ 9.28 \end{bmatrix} \quad \text{ve} \quad X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \quad \text{olsun.}$$

X 'in QR ayrışımı ,

$$X = Q \begin{bmatrix} R \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.4472 & -0.6325 & 0.1450 & 0.3318 & 0.5186 \\ 0.4472 & -0.3162 & -0.6290 & -0.4639 & -0.2987 \\ 0.4472 & 0 & 0.7395 & -0.3231 & 0.3857 \\ 0.4472 & 0.3162 & -0.1720 & 0.7107 & -0.4066 \\ 0.4472 & 0.6325 & -0.0835 & -0.2555 & 0.5725 \end{bmatrix} \begin{bmatrix} 2.236 & 6.708 \\ 0 & 3.162 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

ve

$$Q^T y = z = \begin{bmatrix} z_1 \\ - \\ z_2 \end{bmatrix} = \begin{bmatrix} 15.438 \\ 5.196 \\ -0.370 \\ 0.493 \\ -1.423 \end{bmatrix}$$

olur. Denklem (2.32) 'den $y = X\hat{\beta} + \hat{\varepsilon}$ regresyon modelinin parametre tahmini $\hat{\beta}$

$$\hat{\beta} = R^{-1}z_1 = \begin{bmatrix} 2.236 & 6.708 \\ 0 & 3.162 \end{bmatrix}^{-1} \begin{bmatrix} 15.438 \\ 5.196 \end{bmatrix} = \begin{bmatrix} 1.975 \\ 1.643 \end{bmatrix}$$

olur. Yukarıdaki $\hat{\beta}$ 'nın sonucu ile SPSS ile elde ettiğimiz Tablo 6 'nın Katsayı (Coefficients) 'daki sonuç kıyaslanırsa ikisinin de aynı olduğu görülebilir.

Ayrıca, denklem (2.31) 'den, $\hat{\beta} = R^{-1}z_1$ ise, $y = X\hat{\beta} + \hat{\varepsilon}$ 'nın en küçük hata kareleri toplamı sağlanabildiğini görebiliriz, yani

$$ESS = z_2^T z_2 = \begin{bmatrix} -0.370 \\ 0.493 \\ -1.423 \end{bmatrix}^T \begin{bmatrix} -0.370 \\ 0.493 \\ -1.423 \end{bmatrix} = 2.4048$$

olur. Ayrıca $\hat{\varepsilon}_i$ 'nin varyansı δ^2 olsun ve $\hat{\varepsilon} = \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \end{bmatrix}$, $\hat{\varepsilon}_1 \in \mathbb{R}^2$, $\hat{\varepsilon}_2 \in \mathbb{R}^3$ olsun.

$\hat{\beta} = R^{-1} z_1$ olmak üzere denklem (2.31) 'ten $\hat{\varepsilon}_2^T \hat{\varepsilon}_2 = z_2^T z_2$ olur, dolayısı ile

$ESS = z_2^T z_2 = \hat{\varepsilon}_2^T \hat{\varepsilon}_2$ 'dir. Ki – kare' nin tanımından $ESS = \hat{\varepsilon}_2^T \hat{\varepsilon}_2 \sim \delta^2 \chi(m-n)$ 'dir.

Ayrıca $ESS / m-n = \hat{\varepsilon}_2^T \hat{\varepsilon}_2 / m-n$ ise δ^2 'nin sapmasız tahmincisidir. $s = ESS / m-n$ olsun. Dolayısı ile

$$s^2 = ESS / m-n = 2.4048 / 5 - 2 = 0.8016$$

olur. Yukarıdaki δ^2 'nin sapmasız tahmincisi s^2 ile Tablo 6. 'nın Varyans Analizi (Anova) tablosundaki hata kalıntı (residual) ortalama karesi (mean square) ile kıyaslarsak ikisinin de hemen hemen aynı olduğu görülebilir.

Denklem (2.33) 'le elde ettiğimiz $\hat{\beta}$ 'nın varyansı aşağıdaki gibi bulunur :

$$VarKov(\hat{\beta}) = \delta^2 (R^T R)^{-1} = \begin{bmatrix} 0.882 & -0.240 \\ -0.240 & 0.080 \end{bmatrix}$$

Dolayısı ile $Var(\hat{\beta}_1) = 0.882$, $Sh(\hat{\beta}_1) = 0.939$; $Var(\hat{\beta}_2) = 0.080$, $Sh(\hat{\beta}_2) = 0.283$

olur. Burada QR ayrışımıyla elde ettiğimiz $Sh(\hat{\beta}_1)$ ve $Sh(\hat{\beta}_2)$ 'yi kullanarak $\hat{\beta}$ 'yı t student testi ile sınavalım :

$$t(\hat{\beta}_1) = \frac{\hat{\beta}_1 - 0}{Sh(\hat{\beta}_1)} = \frac{1.975}{0.939} = 2.103$$

$$t(\hat{\beta}_2) = \frac{\hat{\beta}_2 - 0}{Sh(\hat{\beta}_2)} = \frac{1.643}{0.283} = 5.803$$

Yukarıdaki $\hat{\beta}$ 'nın t student test sonucu ile Tablo 6 'nın Katsayı tablosundaki

$\hat{\beta}$ 'nın t student test sonucunun aynı olduğunu görebiliriz. Yani Tablo 6. 'nın Katsayı tablosundaki $\hat{\beta}$ 'nın t student test sonucunun istatistik anlamı ile QR ayrışım yöntemi ile elde ettiğimiz $Sh(\hat{\beta}_1)$ ve $Sh(\hat{\beta}_2)$ 'nın üzerine yaptığımız t student testinin sonucunun istatistik anlamı aynıdır.

Tablo 6. aşağıdaki gibi gösterilmektedir ¹² .

Tablo 1.6.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,958 ^a	,918	,891	,89555

a. Predictors: (Constant), X

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	26,994	1	26,994	33,659	,010 ^a
	Residual	2,406	3	,802		
	Total	29,401	4			

a. Predictors: (Constant), X

b. Dependent Variable: Y

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1,975	,939		2,103	,126
	X	1,643	,283	,958	5,802	,010

a. Dependent Variable: Y

Örnek 2.1.'de veriye klasik yöntem ile *QR* ayrışım yönteminin aynı sonucu verdiğini, böylece regresyon analizinde *QR* ayrışım yönteminin güvenilir bir yöntem olduğunu gösterdik. *QR* ayrışımı sadece regresyon analizine uygulanması

¹² Tablo 1.6.'daki sonuçlar, Örnek 2.1'deki x ve y her biri ayrı olarak bağımsız ve bağımlı değişken olarak SPSS 'yi kullanarak elde edilirler.

ile değil, rankı tam olan matrisin tekil değerini bulmak açısından da çok popüler bir algoritmadır¹³.

Ayrıca Shifted QR algoritması da ilginç bir konudur. Burada bu konuya biraz değineceğiz.

Matris QR Upper Hessenberg Reduction $H = U_0^T A U_0$ olarak açıklanmaktadır¹⁴. H , A 'nın Hessenberg matrisidir ve bir üst altköşegen (upper subdiagonal) matristir. $U_0^T U_0 = I$ ve $U_0 = P_1 P_2 \dots P_k$ 'dir. P_i ($P_1 \leq P_i \leq P_k$) ise A 'nın Householder matrisidir¹⁵.

s bir skaler olsun. Ayrıca $H - sI$ 'nin QR ayrışım $H - sI = QR$ olsun. Bir matris $\hat{H} = RQ + sI$ olsun. $\hat{H}$ ile H arasındaki ilişki ise,

$$\hat{H} = RQ + sI = Q^T (QR + sI) \quad Q = Q^T H Q$$

Yukarıdaki gibi $\hat{H}$ ile H arasındaki ilişki ile H 'nin altköşegenindeki terimler sıfırlanarak A 'nın Hessenberg matris H , $\hat{H}$ 'ya dönüştürülür.

A 'nın i 'nci özdeğerinin notasyonu γ_i olsun. H 'nin i 'nci altköşegen terimi,

$$\left| \frac{\gamma_{i+1} - s}{\gamma_i - s} \right|^k \text{ hızı ile sıfıra yakınsar. Buradaki } s \text{ ise " shift " tir }^{16}.$$

Yukardaki algoritmaya Shifted QR algoritması ismi verilir¹⁷.

¹³ Watkins David S. (1982), s.428.

¹⁴ Ayrıntılı bilgi için bakınız. Parlett B.N.(2000), Section 7 ve Golub Gene H.(1996), s.352 –253.

¹⁵ Bu konudaki ayrıntı bilgileri için bakınız. Golub Gene H. (1996)s. 344 – 346.

^{16,17} Ayrıntılı bilgi için bakınız. Golub Gene H. (1996) s.353 ve Demmel James W. (1997), s.161-173.

3. BÖLÜM LU AYRIŞIM

3.1. LU Ayrışımın Kavramı

Bu kısımda Gauss Eliminasyonunu ile LU ayrışımın arasındaki bağlantı ve bir matrisin LU ayrışımının türetilmesini tartışacağız. Ayrıca LU ayrışım algoritmasının özel bir durumu – Pivot yöntemi ile ilgili tartışma da bu kısımda yer almaktadır.

3.1.1. Gauss Eliminasyonu ve LU Ayrışım

Örnek 3.1. $A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix}$ olsun. Gauss Eliminasyon algoritması ile A matrisini

bir üstüçgen matrise dönüştürelim.

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix} \xrightarrow{p_1} \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{bmatrix} \xrightarrow{p_2} \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{bmatrix} = U$$

Burada aşama aşama yukardaki A 'nın Gauss Eliminasyonunu cebir dili ile gerçekleştirelim.

$$M_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \text{ olsun. Dolayısı ile}$$

$$M_1 A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{bmatrix}$$

olur. Yukarıdaki $M_1 A$ 'nın sonucu ile A 'nın Gauss Eliminasyonunun p_1 aşamasının sonucu aynıdır, yani A 'nın Gauss Eliminasyonunun p_1 aşamasını

cebirsel olarak gerçekleştirebiliriz. Keza, p_1 aşamasını Gauss Eliminasyonunu cebirsel olarak gerçekleştirelim.

$$M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \text{ olsun.}$$

$$M_2 M_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{bmatrix} = U$$

olur. Yukarıdaki M_1 ve M_2 , A matrisinin Gauss transformasyonu (transformation) ¹ 'dur.

Burada A matrisinin Gauss transformasyonunu M_1 ve M_2 'den türetelim ².

$$t^{(1)} = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix} \text{ bir vektör ve 1'nci birim vektör } e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ olsun. } M_1 \text{ 'i}$$

$$M_1 = I - t^{(1)} e_1^T = I - \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$$

$$\text{olarak elde ederiz. } t^{(2)} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \text{ bir vektör ve 2'nci birim vektör } e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ olsun. } M_2 \text{ 'yi}$$

$$M_2 = I - t^{(2)} e_2^T = I - \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$$

olarak elde ederiz. $t^{(1)}$ ve $t^{(2)}$, A matrisinin Gauss vektörü (Gauss vector) ³ dür.

Ayrıca köşegendeki altmatrislerin determinantları sıfır olmayan herhangi bir

^{1,2} Matrisinin Gauss transformasyonu ile ilgili ayrıntılar için bakınız. Golub. Gene. H. (1996), s.94 – 95.

³ Matrisinin Gauss vektörü tanımlanışı için bakınız. Golub.Gene.H. (1996), s.95.

kare matris X 'in ($X \in \mathbb{R}^{n \times n}$) k 'nci Gauss Transformasyonu şu şekilde açıklanabilir :

$$M_k = I - t^{(k)} e_k^T = I - \begin{bmatrix} 0 \\ \vdots \\ 0 \\ t_{k+1} \\ \vdots \\ t_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & t^{(k)}_{k+1} & 1 & 0 & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & t^{(k)}_n & 0 & \cdot & 0 & 1 \end{bmatrix} \quad (3.1)$$

$1 \leq k \leq n-1$

Buradaki $t^{(k)}$, X matrisinin k 'nci Gauss vektörüdür. e_k ise k 'nci birim vektördür.

Ayrıca $t^{(k)}$ 'nun i 'nci terimi $t^{(k)}_i$ ise $t^{(k)}_i = \frac{d_{ki}}{d_{kk}}$ 'dir, ayrıca $i = k+1 : n$ dir⁴. d_{ki} ise

D matrisinin k 'nci vektörünün i 'nci terimidir ve $D = M_{k-1} \dots M_2 M_1 X$ 'dir :

X matrisinin Gauss Eliminasyonunu U şu şekilde açıklanabilir,

$$U = M_{n-1} \dots M_2 M_1 X \quad (3.2)$$

Aynı zamanda U matrisi de X matrisinin LU ayrışımındaki U ögesidir. Ayrıca

M_k tekil olmayan bir kare matris olduğu için denklem (3.2) 'den

$$X = M_1^{-1} \dots M_{n-2}^{-1} M_{n-1}^{-1} U$$

elde ederiz. $M_1^{-1} \dots M_{n-2}^{-1} M_{n-1}^{-1} = L$ olsun, dolayısı ile yukarıdaki denklem şu

şekilde yazılabilir :

$$X = LU$$

Bu X matrisinin LU ayrışımıdır. Matrisin LU ayrışımının daha ayrıntılı tanımlanışı aşağıdaki Teorem 3.1. 'de verilmiştir :

⁴Buradaki $k+1 : n$ ise, $k+1$ 'den n 'ye kadar sıralanmaktadır.

Teorem 3.1. Bir $n \times n$ boyutlu X matrisinin köşegenindeki altmatrislerin determinantlarının hepsi sıfır olmuyorsa, yani $\det X(1:i, 1:i) \neq 0$ ise ($i = 1: n-1$), X 'in LU ayrışımı söz konusudur. Buradaki $X(1:i, 1:i)$, X matrisinin köşegenindeki altmatrislerdir. Ayrıca L köşegeninin terimleri '1' olan $n \times n$ boyutlu alt üçgen matris, U ise $n \times n$ boyutlu üst üçgen matristir.

Teorem 3.1. 'in ispatı Golub.Gene.H (1996)'da bulunabilir ⁵.

Burada L 'yi X 'in Gauss vektörü $t^{(k)}$ ile açıklayalım. Denklem (3.1) 'deki

$$M_k = I - t^{(k)} e_k^T \text{ olduğu için}$$

$$M_k^{-1} = I + t^{(k)} e_k^T$$

olur. Çünkü $M_k^{-1} = I + t^{(k)} e_k^T$ olmak üzere

$$M_k^{-1} M_k = (I + t^{(k)} e_k^T) (I - t^{(k)} e_k^T) = I - t^{(k)} e_k^T t^{(k)} e_k^T = I - t^{(k)} (e_k^T t^{(k)}) e_k^T = I$$

olur, buradaki $e_k^T t^{(k)} = 0$ 'dır. Dolayısı ile $M_k^{-1} = I + t^{(k)} e_k^T$ 'dir. Bu yüzden

$$\begin{aligned} L &= M_1^{-1} \dots M_{n-2}^{-1} M_{n-1}^{-1} \\ &= (I + t^{(1)} e_1^T) (I + t^{(2)} e_2^T) \dots (I + t^{(n-1)} e_{n-1}^T) \\ &= (I + t^{(1)} e_1^T + t^{(2)} e_2^T + t^{(1)} e_1^T t^{(2)} e_2^T) (I + t^{(3)} e_3^T) \dots (I + t^{(n-1)} e_{n-1}^T) \\ &= I + \sum_{k=1}^{n-1} r^{(k)} e_k^T \end{aligned} \quad (3.3)$$

olur ⁶. Denklem (3.3) 'teki $t^{(1)} e_1^T t^{(2)} e_2^T = 0$, çünkü $e_1^T t^{(2)} = 0$ 'dır. Keza,

$$t^{(k)} e_k^T t^{(j)} e_j^T = 0 \quad k \neq j$$

olur. Dolayısı ile denklem (3.3) 'teki $I + \sum_{k=1}^{n-1} r^{(k)} e_k^T$ 'ten başka öğelerin hepsi

⁵ Golub.Gene.H. (1996), s.97 - 98.

⁶ İlgili ispat için bakınız. a.g.e.,s.99.

sıfırdır.

Denklem (3.3) 'ün üzerine Örnek 3.1. 'deki A 'nın LU ayrışımının L ögesini bulalım.

$$\begin{aligned} L &= I + t^{(1)}e_1^T + t^{(2)}e_2^T \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix} \end{aligned}$$

olur. Dolayısı A 'nın LU ayrışımı şu şekilde açıklanabilir :

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{bmatrix} = LU$$

Örnek 3.2 $D = \begin{bmatrix} 1 & 3 & 9 \\ 2 & 6 & 7 \\ 3 & 8 & 3 \end{bmatrix}$ olsun. D 'nin LU 'ya ayrıştırılabildiğini varsayalım,

yani,

$$D = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

olsun. Bu denklemin tutarlı olabilmesi için L ve U 'nun öğelerinin aşağıdaki değerlere sahip olmaları gerekiyor:

$u_{11} = 1, u_{12} = 3, u_{22} = 0, l_{21} = 2, l_{31} = 3$. Bu değerlere göre D matrisinin (3,2)

terimi $d_{32} = l_{31}u_{12} + l_{32}u_{22} = 9$ 'dur. Fakat bu sonuca aykırı olarak d_{32} 'nin gerçek değeri 8 'dir. Bu yüzden D 'nin LU ayrışımı yoktur diyebiliriz.

D 'yi Gauss Eliminasyon algoritma ile üst üçgen matris şekline dönüştürmeyi

deneyelim.

$$D = \begin{bmatrix} 1 & 3 & 9 \\ 2 & 6 & 7 \\ 3 & 8 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 3 & 9 \\ 0 & 0 & -11 \\ 0 & -1 & -24 \end{bmatrix} = D_1$$

Görüldüğü gibi 2'nci ile 3'üncü satırı yer değiştirmeden D_1 üst üçgen matris şekline dönüştürülmez. Çünkü D 'nin köşegenindeki bir altmatris $\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$ 'nın

determinantı sıfıra eşit olduğu için D_1 'in 2 'nci pivot ⁷, yani (2,2) ögesi sıfıra eşittir.

Burada denklem (3.2) 'ile D 'nin LU ayrışımının U ögesini bulmayı deneyelim.

$$M_1 = I - t^{(1)} e_1^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$$

olur. Dolayısı ile

$$D_1 = M_1 D = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 9 \\ 2 & 6 & 7 \\ 3 & 8 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 9 \\ 0 & 0 & -11 \\ 0 & -1 & -24 \end{bmatrix}$$

olur. D 'nin 2'nci Gauss transformasyonu $M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -t^{(2)}_3 & 1 \end{bmatrix}$ 'dir. $t^{(2)}_3$ ise D 'nin

2 'nci Gauss vektörü $t^{(2)} = \begin{bmatrix} 0 \\ 0 \\ t^{(2)}_3 \end{bmatrix}$ 'nın 3 'üncü terimidir. Gauss vektörünün

tanımlanışına göre, $t^{(2)}_3 = \frac{-1}{0}$ olması gerekiyor. Fakat 0 bir payda olmaz. Bu

yüzden $t^{(2)}_3$ söz konusu değildir. Dolayısı ile D 'nin 2'nci Gauss Transformasyonu

⁷Pivot, Gauss Elimination 'un her aşamasından sonra elde ettiği matrisin Gauss Elimination algoritması ile işletilecek altmatrisinin (1,1) ögesidir. Örneğin, Örnek3.1 'deki A 'nın p_1 aşamasından sonra elde ettiği matrisin pivot 'u ise -3 'tir. Ayrıntılı bilgi için bakınız. Golub Gene H.(1996),s.110.

yok etmeksizin D üst üçgen şekline dönüştürülemez.

Teorem 3.1. 'de matrisin LU 'ya ayrıştırılabilmesi için bir koşul ortaya koyuldu, yani matrisinin köşegenindeki altmatrisler determinantlarının hepsi sıfır olmamalıdır. Başka deyişle bir $n \times n$ boyutlu matris X 'in k 'nci Gauss Eliminasyon aşamasındaki, yani $X_k = M_k M_{k-1} \dots M_1 X$ ($1 \leq k \leq n-1$) aşamasındaki pivot sıfıra eşit olmasa, X üst üçgen matris şekline dönüştürülebilir, yani LU ayrışımına dönüştürülebilir.

İlk örnekte rankı tam olan matrisin LU ayrışımını tartıştık. Rankı tam olmayan bir matris LU ayrışımına dönüştürülebilir mi ? Burada tartışalım.

$$C = \begin{bmatrix} 1 & 2 & 3 & 6 \\ 3 & 4 & 2 & 4 \\ 5 & 6 & 2 & 4 \\ 4 & 0 & 1 & 2 \end{bmatrix} \text{ olsun. } C \text{ matrisinin 3'üncü ve 4'üncü sütunu arasında}$$

lineer bağlantı vardır. C matrisinin Gauss transformasyonunu kullanarak C matrisinin LU ayrışımını şu şekilde bulabiliriz :

$$C = \begin{bmatrix} 1 & 2 & 3 & 6 \\ 3 & 4 & 2 & 4 \\ 5 & 6 & 2 & 4 \\ 4 & 0 & 1 & 2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & l_{14} \\ 3 & 1 & 0 & l_{24} \\ 5 & 2 & 1 & l_{34} \\ 4 & 4 & 17 & l_{44} \end{bmatrix}}_L \underbrace{\begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & -2 & -7 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_U$$

Yukarıda C matrisinin LU ayrışımını görebiliriz, bir tane U vardır, fakat L 'nin l_{14} , l_{24} , l_{34} ve l_{44} terimleri belirli olmadığı için L 'nin değeri sonsuz tane dir, yani C matrisinin sonsuz tane LU ayrışımı vardır. Fakat $n \times n$ boyutlu tekil olmayan kare matrisin LU ayrışımı tektir (unique).

Teorem 3.2. $n \times n$ boyutlu tekil olmayan matris A 'nın LU ayrışımı varsa, L ve U tektir. Ayrıca A matrisinin determinant $\det(A) = u_{11} u_{12} \dots u_{nn}$ 'dir. Buradaki u_{ii}

$(1 \leq i \leq n)$ U 'nun köşegenindeki öğesidir.

Burada tersten ispat yöntemi ile Teorem 3.2 'yi ispat edelim. Benzer ispat Golub Gene H. (1996) ⁸ 'de bulunuyor.

İspat : Tekil olmayan A matrisinin iki tane LU ayrışımı söz konusu olduğunu varsayalım. Yani $A = L_1 U_1$, aynı zamanda $A = L_2 U_2$ 'dir. Dolayısı ile $L_1 U_1 = L_2 U_2$ 'dir. L_1 , U_1 , L_2 ve U_2 'nin hepsi rankı tam olan üçgen matris olduğu için

$$L_2^{-1} L_1 = U_2 U_1^{-1}$$

olur. $L_2^{-1} L_1$ bir birim alt üçgen matris ve $U_2 U_1^{-1}$ bir üst üçgen matris olduğu için

$$L_2^{-1} L_1 = U_2 U_1^{-1} = I$$

olmalıdır ⁹. Dolayısı ile $L_2 = L_1$, $U_2 = U_1$ 'dir.

Ayrıca A 'nın LU ayrışımındaki L 'nin determinanı $\det(L) = I$ olduğu için $\det(A) = \det(LU) = \det(L)\det(U) = \det(U) = u_{11}u_{12}....u_{nn}$ 'dir.

3.1.2 Pivot Yöntemi ¹⁰

Ömek 3.2.'deki $D = \begin{bmatrix} 1 & 3 & 9 \\ 2 & 6 & 7 \\ 3 & 8 & 3 \end{bmatrix}$ matrisinin köşegenindeki bir altmatrisin $\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$

determinantı sıfır olduğu için D 'nin LU ayrışımının olmadığını Teorem 3.1. 'den

biliyoruz. $E = \begin{bmatrix} 0.0001 & 1 \\ 1 & 1 \end{bmatrix}$ olsun . E 'nin LU ayrışımı

$$E = LU = \begin{bmatrix} 1 & 0 \\ 10000 & 1 \end{bmatrix} \begin{bmatrix} 0.0001 & 1 \\ 0 & -9999 \end{bmatrix} \quad (3.4)$$

⁸ Ayrıntılı bilgi için bakınız. Golub Gene H. (1996),s.101

⁹ Üçgen matrisin ters ve çarpması ile ilgili teoremler için bakınız. Golub Gene H.(1996),s.93.

¹⁰ Konu ile ilgili ayrıntılar için bakınız. Golub Gene. H. (1996) , s.109.

olur. E 'nin LU ayrışımındaki L ve U 'nun terimleri çok büyük veya küçük olmasının sebebi, E 'nin ilk pivotu 0.0001'in çok küçük olmasıdır. Eğer bu pivot sıfıra eşitse, E matrisi LU 'ya da ayrıştırılamayacaktır.

D ile E matrisinin sorununu aşabilmek için pivot yöntemini uygulamamız gerekiyor. $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ olsun.

$$PE = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0.0001 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0.0001 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0.0001 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0.9999 \end{bmatrix} = LU \quad (3.5)$$

olur. Buradaki PE 'nin LU ayrışımının denklem (3.4) 'tekinden daha uygun olduğunu görebiliriz. Yani PE 'nin LU ayrışımında fazla aşırı değer yoktur. P 'ye permutasyon (permutation) matris denir ¹¹. Bir matris, permutasyon matrisle önden çarpılırsa, karşılık gelen bu matrisin satırları yer değiştirir veya arkadan çarpılırsa bu matrisin karşılık gelen sütunlar yer değiştirir.

Denklem (3.5) ' te P permutasyon matrisi ile E matrisinin satırları yer değiştirilerek uygun LU 'ya ayrıştırıldı. Yukarıda P permutasyon matrisi kullanarak E matrisin uygun LU ayrışımını bulduk. Bir örnekle daha pivot yöntemini inceleyelim.

Örnek 3.3 $H = \begin{bmatrix} 3 & 17 & 10 \\ 2 & 4 & -2 \\ 6 & 18 & -12 \end{bmatrix}$ olsun.

E matrisinin LU ayrışımından görebiliriz ki pivotta yer alan terim ne kadar büyükse o kadar iyidir, yani pivotun yer aldığı sütundaki terimler arasından

¹¹ Permutasyon matris le ilgili ayrıntılı bilgi için bakınız. Golub Gene H.(1996), s.109 - 110.

pivottaki terim mutlak değerce en büyüktür. Bu zaten pivot yönteminin esasıdır. Burada pivot yönteminden biri – kısmi pivot (partial pivot) yöntemi ¹² ile H matrisinin LU ayrışımını bulalım, yani H 'in pivotunun yer aldığı 1'nci sütunda pivotunun mutlak değerini en büyük yapalım.

$$P_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \text{ olsun. Dolayısı ile } P_1 H = \begin{bmatrix} 6 & 8 & -12 \\ 2 & 4 & -2 \\ 3 & 17 & 10 \end{bmatrix} \text{ olur. } P_1 H \text{ 'in 1'nci}$$

$$\text{Gauss transformasyonu } M_1 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{3} & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix} \text{ 'dir. } H \text{ matrisinin Gauss}$$

Eliminasyonunun birinci aşaması yani H matrisinin LU ayrışımındaki U ögesinin türetilmesindeki birinci aşaması aşağıdaki gibi gösterilerek

$$H_1 = M_1 P_1 H = \begin{bmatrix} 6 & 8 & -12 \\ 0 & -2 & 2 \\ 0 & 8 & 16 \end{bmatrix}$$

olur. H_1 'in pivotu $-2 < 8$ olduğu için H_1 'in 2 ve 3'üncü satırının yer değiştirilmesi gerekiyor, yani H_1 'e kısmi pivot yöntemini uygulaması da gerekiyor.

$$\text{Permutasyon matris } P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ olsun.}$$

$$P_2 H_1 = P_2 M_1 P_1 H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 6 & 8 & -12 \\ 0 & -2 & 2 \\ 0 & 8 & 16 \end{bmatrix} = \begin{bmatrix} 6 & 8 & -12 \\ 0 & 8 & 16 \\ 0 & -2 & 2 \end{bmatrix} \text{ 'dir. Dolayısı ile}$$

¹² Kısmi pivot yöntemi ile ilgili ayrıntılı bilgi için bakınız. Golub Gene H.(1996), s.110 – 113.

H matrisinin ikinci Gauss Transformasyonu $M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/4 & 1 \end{bmatrix}$ 'dir. Dolayısı ile H

'in Gauss Eliminasyonu ikinci aşaması

$$H_2 = M_2 P_2 H_1 = M_2 P_2 M_1 P_1 H = \begin{bmatrix} 6 & 8 & -12 \\ 0 & 8 & 16 \\ 0 & 0 & 6 \end{bmatrix} = U \quad (3.6)$$

olur. Yukarıdaki U aynı zamanda da $P_2 P_1 H$ matrisinin LU ayrışımının U ögesidir, bu kolayca ispat edilebilir.

$P_2 P_1 = P$ olsun. PH matrisinin LU ayrışımının U ögesini denklem (3.6)'daki gibi bulduk, PH matrisinin LU ayrışımının L ögesini de bulalım. Burada kullanacağımız algoritma Golub Gene H. (1996)'da bulunmaktadır ¹³.

L 'nin k 'inci sütunun notasyonu l_k olsun. l_k 'nin köşegeninin altındaki vektörün notasyonu g_k olsun ve $g_k = f_k (k+1:n)$ 'dir. $f_k = P_{n-1} \dots P_{k+2} P_{k+1} t^{(k)}$ 'dir. $t^{(k)}$ ise Gauss vektördür.

Bu algoritmayı uygulayarak PH matrisinin LU ayrışımının L ögesini bulalım.

$$P_1 H \text{ matrisinin Gauss vektörü } t^{(1)} = \begin{bmatrix} 0 \\ 1/3 \\ 1/2 \end{bmatrix} \text{ olduğu için}$$

$$f_1 = P_2 t^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1/3 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/2 \\ 1/3 \end{bmatrix}$$

$$\text{olur. Dolayısı ile } g_1 = f_1 (2:3) = \begin{bmatrix} 1/2 \\ 1/3 \end{bmatrix} \text{ 'dir ve } l_1 = \begin{bmatrix} 1 \\ g_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1/2 \\ 1/3 \end{bmatrix} \text{ 'dir. Keza,}$$

$g_2 = f_2(3:3) = t^{(2)} = -\frac{1}{4}$ 'dir. $t^{(2)}$ ise $P_2 H_1$ 'in Gauss vektörüdür. Dolayısı ile

$$l_2 = \begin{bmatrix} 0 \\ 1 \\ g_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -\frac{1}{4} \end{bmatrix} \text{ 'dir, } L \text{ ise } \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & -\frac{1}{4} & 1 \end{bmatrix} \text{ 'dir. } PH \text{ matrisinin } LU \text{ ayrışımı ise}$$

$$PH = LU = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & -\frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} 6 & 8 & -12 \\ 0 & 8 & 16 \\ 0 & 0 & 6 \end{bmatrix}$$

olur.

Örnek 3.2.'deki D matrisinin köşegeninde bir altmatris tekil olduğu için, yani D 'nin birinci Gauss Eliminasyonu uygulandıktan sonra elde ettiğimiz matrisin pivotu sıfır olduğu için D üst üçgen şekilli matrisine dönüştürülemiyor, yani D 'nin LU ayrışımı söz konusu değildir. Burada kısmi pivot yöntemi ile D 'nin LU ayrışımını sağlayalım.

$$P^* D = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 3 & 9 \\ 2 & 6 & 7 \\ 3 & 8 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 8 & 3 \\ 2 & 6 & 7 \\ 1 & 3 & 9 \end{bmatrix} \longrightarrow \begin{bmatrix} 3 & 8 & 3 \\ 0 & \frac{2}{3} & 5 \\ 0 & 0 & \frac{11}{12} \end{bmatrix} = U^*$$

$$L^* \text{ ise, } L^* = \begin{bmatrix} 1 & 0 & 0 \\ \frac{2}{3} & 1 & 0 \\ \frac{1}{3} & \frac{1}{2} & 1 \end{bmatrix} \text{ olarak bulunur. Dolayısı ile}$$

$$P^* D = L^* U^* = \begin{bmatrix} 1 & 0 & 0 \\ \frac{2}{3} & 1 & 0 \\ \frac{1}{3} & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 3 & 8 & 3 \\ 0 & \frac{2}{3} & 5 \\ 0 & 0 & \frac{11}{12} \end{bmatrix}$$

olur.

Kısmi pivot yönteminden başka bir pivot yöntemi daha vardır , complete pivot yöntemidir ¹³.

Complete pivot yöntemi ile bir matrisin her aşamadaki Gauss Eliminasyonu uygulandıktan sonra elde ettiğimiz matrisin köşegenindeki altmatrisinin mutlak değer en büyük olan terim pivot yerine geçer. Ayrıca Complete pivot yöntemi ile bir matrisin hem sütunları hem satırlarının yer değiştirmesini sağlayabiliriz.

Örneğin $K = \begin{bmatrix} 3 & 8 & 10 \\ 2 & 4 & -2 \\ 6 & 16 & -12 \end{bmatrix}$ olsun. K matrisinde mutlak değeri en büyük

olan terim 16 'dır. Burada permutasyon matris θ ve θ^* ¹⁴ 'le 16 terimini matris K 'nin pivot yerine çevirelim.

$$\theta K \theta^* = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} K \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 16 & 6 & -12 \\ 4 & 2 & 2 \\ 8 & 4 & 10 \end{bmatrix} \longrightarrow \begin{bmatrix} 16 & 6 & -12 \\ 0 & \frac{1}{2} & 1 \\ 0 & 1 & 16 \end{bmatrix} = K_1$$

K_1 'in köşegenindeki altmatris $\begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 16 \end{bmatrix}$ 'de mutlak değeri en büyük olan

terim 16 'dır. Burada 16 'yı permutasyon matris θ_1 ve θ_1^* 'le $\begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 16 \end{bmatrix}$

altmatrisinin pivot yerine çevirelim.

$$\theta_1 K_1 \theta_1^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} K_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 16 & -12 & 6 \\ 0 & 16 & 1 \\ 0 & 1 & \frac{1}{2} \end{bmatrix} \longrightarrow \begin{bmatrix} 16 & -12 & 6 \\ 0 & 16 & 1 \\ 0 & 0 & -\frac{7}{6} \end{bmatrix}$$

Dolayısı ile complete pivot üzerine K 'nin LU ayrışımının U ögesini aşağıdaki gibi bulduk. Yani,

¹³ Complete pivot yöntemi ile ilgili ayrıntılar için bakınız. Golub Gene H.(1996), s.117 - 119.

¹⁴ Buradaki θ ve θ^* ise ayrı ayrı olarak K matrisinin satırları ve sütunlarının yerini değiştiren permutasyon matristir.

$$\theta_1 \theta K \theta^* \theta_1^* = \begin{bmatrix} 16 & -12 & 6 \\ 0 & 16 & 1 \\ 0 & 0 & -7/16 \end{bmatrix} = U$$

Ayrıca Örnek 3.3. 'teki PH matrisinin LU ayrışımının L öğesinin türetilmesine benzer bir algoritma ile $\theta_1 \theta K \theta^* \theta_1^*$ 'nin LU ayrışımının L öğesini kolayca bulabiliriz.

Yukarıdaki örnek gibi complete pivot yöntemi ile herhangi bir matris LU 'ya ayrıştırılabilir.

3.2. Matrisin QR Ayrışımına Hesaplanmasına LU 'nun Katkısı

QR ayrışımının hesaplanmasında Gram – Schmidt, Householder, Given ve Fast Given ideal algoritmalarıdır ¹⁵. Ayrıca Fast Given 'in, QR ile LU ayrışımın arasındaki bağlantısını gösterebilmek için, bu kısımda Fast Given algoritmasını kullanarak bir örnek matrisin QR ayrışımını hesaplayalım.

Örnek 3.4. $A = \begin{bmatrix} 0 & -2 & 1 \\ 1 & 3 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 5 \end{bmatrix}$ olsun ¹⁶. A 'nın QR ayrışımı aşağıdaki gibi olur :

$$QR = \begin{bmatrix} 0 & -2/\sqrt{6} & -1/2 \\ 1/\sqrt{2} & 1/\sqrt{6} & -1/2 \\ 0 & 0 & 1/2 \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/2 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 2\sqrt{2} & 3\sqrt{2} \\ 0 & \sqrt{6} & -\sqrt{6} \\ 0 & 0 & 2 \end{bmatrix} \quad (3.7)$$

Aynı zamanda A 'nın QR ayrışımı şu şekilde de açıklanabilir :

¹⁵ Ayrıntılı bilgi için bakınız. Golub Gene H. (1996), s 223 – 228.

¹⁶ Bu örnekteki matris A ise, Pursell Lyle and Trimble S.Y., American Mathematical Monthly, Vol.98, No.6, s.544 'ten kaynaklanmaktadır.

$$A = Q^* R^* = \begin{bmatrix} 0 & -2 & -1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.8)$$

Buradaki Q^* normalleştirilmemiştir, fakat denklem (3.7) 'deki Q normalleştirilmiştir.

$[A^T A \mid A^T]$ 'e Gauss Eliminasyonun algoritmasını uygulayalım :

$$[A^T A \mid A^T] = \left[\begin{array}{ccc|ccc} 2 & 4 & 6 & 0 & 1 & 0 & 1 \\ 4 & 14 & 6 & -2 & 3 & 0 & 1 \\ 6 & 6 & 28 & 1 & 1 & 1 & 5 \end{array} \right] \longrightarrow \left[\begin{array}{ccc|ccc} 2 & 4 & 6 & 0 & 1 & 0 & 1 \\ 0 & 6 & -6 & -2 & 1 & 0 & -1 \\ 0 & 0 & 4 & -1 & -1 & 1 & 1 \end{array} \right]$$

$[A^T A \mid A^T]$ 'nin Gauss Eliminasyon algoritmasını uyguladıktan sonra elde

ettiğimiz matrisin sağ tarafı yani $\begin{bmatrix} 0 & 1 & 0 & 1 \\ -2 & 1 & 0 & -1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$, satır ortogonal olan bir matristir.

Ayrıca bu matrisin devriği ile denklem (3.8) 'deki Q^* aynı olur. Bu tesadüfi bir şey değildir. QR ile LU ayrışımın arasında zaten bir bağlantı vardır. Burada A örnek matrisi ile bu bağlantıyı daha somut bir şekilde gösterelim.

$A^T A$ 'nin LU ayrışımı

$$A^T A = LU = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 6 \\ 0 & 6 & -6 \\ 0 & 0 & 4 \end{bmatrix}$$

olur. $\hat{A} = A(L^{-1})^T$ olsun.

$$\hat{A} = A(L^{-1})^T = \begin{bmatrix} 0 & -2 & 1 \\ 1 & 3 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 5 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -1 & 1 \end{bmatrix}^{-1} \right)^T = \begin{bmatrix} 0 & -2 & -1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix}$$

olur. $\hat{A}$ ile denklem (3.8)'deki Q^* aynıdır. Yani $\hat{A}$, A 'nın QR ayrışımının Q ögesinin normalleşmemiş şeklidir. Ayrıca $D = \hat{A}^T \hat{A}$ olsun.

$$D = \hat{A}^T \hat{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

'dır.

Fast Given yönteminden biliyoruz ki, A 'nın QR ayrışımının Q ve R ögesi aşağıdaki gibi açıklanabilirler ¹⁷:

$$Q = \hat{A} D^{-\frac{1}{2}} = \begin{bmatrix} 0 & -2 & -1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 1/\sqrt{6} & 0 \\ 0 & 0 & 1/2 \end{bmatrix} = \begin{bmatrix} 0 & -2/\sqrt{6} & -1/2 \\ 1/\sqrt{2} & 1/\sqrt{6} & -1/2 \\ 0 & 0 & 1/2 \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/2 \end{bmatrix}$$

Ayrıca A 'nın QR ayrışımının R ögesi

$$R = D^{\frac{1}{2}} L^T = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{6} & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -1 & 1 \end{bmatrix}^T = \begin{bmatrix} \sqrt{2} & 2\sqrt{2} & 3\sqrt{2} \\ 0 & \sqrt{6} & -\sqrt{6} \\ 0 & 0 & 2 \end{bmatrix}$$

olur. Yukarıdaki Fast Given yöntemi ile elde ettiğimiz Q ve R ile denklem (3.7) 'deki A 'nın QR ayrışımının Q ve R aynı olduğunu görebiliriz.

¹⁷ Ayrıntılı bilgi için bakınız. Golub Gene H. (1996), s.228 ve Pursell Lyle and Trimble S.Y., American Mathematical Monthly, Vol.98, No.6, s.544 – 549.

SONUÇ

SVD algoritmasına daha önce gereken önemin verilmediğini söyleyen Gilbert Strang¹, matrislerin SVD algoritması ile ayrıştırılmasının öneminin altını çizmiştir. İlk bölümde SVD 'ye yer verildi. Herhangi bir A matrisi, SVD algoritması ile $U\theta V^T$ şeklinde ayrıştırılabilir diye ispat edildi. Burada U , A 'nın sütun ve V , A 'nın satır uzayını göstermektedir. Bağımsız değişkenler arasında çoklu doğrusallık, θ 'nın sıfıra yakın olan en küçük ögesi tarafından yansıtılmaktadır. Bu en küçük öge atılarak gereksiz bağımsız değişkenlerin modelden dışlanması sağlanmaktadır. Ayrıca bu bölümde, EKK yöntemine alternatif olarak regresyon analizine SVD yöntemi uygulanmıştır. Yine bu bölümde örneklerle SVD ve matris rotasyon kavramlarının yardımı ile $\hat{\beta}$ vektörü geometrik bakımdan incelenmiştir.

QR ayrışımı bir matrisin sütun uzayının ortonormalleştirilmesini sağlayan bir algoritmadır. Tezde Classic Gram - Schmidt yöntemi ile matrisin QR ayrışımını gerçekleştirilmektedir. Bu bölümde EKKY ile aynı teoriyi taban alan QR yöntemi regresyon analizine uygulanmıştır.

QR ayrışımı gerçekleştirmek için Gram - Schmidt ve Householder Triangularization yöntemlerine ilaveten Fast Given yöntemi de kullanılabilir. Bu bölümde bir örnek yardımıyla Fast Given yöntemi uygulanmış ve QR ile LU ayrışım arasındaki bağlantılar ortaya konulmuştur.

Özet olarak, bu çalışmanın EKKY 'ye alternatif olan SVD 'nin anlaşılmasında yararlı olacağı inancını taşımaktayız.

¹ Kalman Dan, A Singularly Valuable Decomposition: The SVD of A Matrix, The College Mathematics Journal, Vol.27, No.1, 1996, s.2

KAYNAKÇA

1. Ansley Graig F., Quick Prof of Some Regression Theorem via the QR Algorithm, **the American Statistician**, February 1985,39.No.1, p55 – 56.
2. Altan M. Emin, **Lineer Cebir**, 2. Baskı, Mimar Sinan Üniversitesi Yayınevi, 1997, İstanbul.
3. Bipschuts Seymoun, **Lineer Cebir**, 2 Baskı Türkçe Çevrisi, Nobel Yayın Dağıtım, Ankara,1991.
4. Boullion Thomas L. and Odell Patrick L. , **Generalized Inverse Matrics**, John Wiley & Sons Inc, 1971.
5. Campbell S.L.and Meyer C.D.,**Generalized Inverse of Linear Transformations**, Dover Publications, New York, 1991.
6. Chan Tony F., Deflated Decomposition of Solutions of Nearly Singular Systems, **SIAM Journal on Numerical Analysis**, Vol.24,No.4, p738 – 754.
7. Chan Tony F., On the Existence and Camputation of LU Factorization with Small Pivots, **Mathematis of Camputation**, Vol.42,No.166, p535 – 547.
8. Chu Moody T., Funderlic Robert E. and Golub Gene H., A Rank – One Reduction Formula and Its Application to Matrix Factorizations, **SIAM Review**, Vol.37, No.4, p512 – 530.
9. Cline A. K., An Elimination Method for the Solution of Linear Least Squares Problems, **SIAM Journal Numerical Analysis**, Vol.10, No.2, April 1973.
10. Damodav N. Gujarati, Çeviren Ümit Şenesen, **Temel Ekonometri**, Literatür Yayıncılık, İstanbul, 1999.
11. Davies P.T. and Tso Mark s., Procedure for Reduced Rank Regression, **Applied Statistics**, Vol.31, No.3 (1982), p244 – 255.
12. Demme James W. I, **Applied Numerical Linear Algebra** , SIAM, 1997, Philadelphia.
13. Eubank R. L. and Webster J. T., The Singular Value Decoposition as a Tool for Solving Estimability Problems, **the American Statistician**, February,1985, 39, No.1, p64 – 72.
14. Flanigan Frank and Pfiefer Richard, Gaussian Elimination in Integer Arithmetic: An Application of the LU Factorization, **The College Mathematics Journal**, Vol.24,No.1, January 1993, 67 – 71.
15. Golub Gene H., **Matrix Computations**, Third Edition, The Johns Hopkins University Pres, Baltimore 1996.

16. Green Paul E. and Carroll J. Douglas, **Mathematical Tools for Applied Multivariate Analysis**, Academic Press, New York, 1976.
17. Hern Thomas, Gaussin Elimination in Integer Arithmetic : An Application of the LU Factorization, **The College Mathematics Journal**, Vol.24, No.1, p67 – 71.
18. Hong Y. P. and Pan C.T., Rank – Revealing QR Factorizations and the Singular Value Decomposition, **Mathematics of Computation**, Vol.58, No.197, p213 - 232
19. Kalman Dan, A Singularly Valuble Decomposition: SVD of a Matrix, **the College Mathematics Journal**, Vol.27, No.1, p2 – 23.
20. Kraines David P. and Kraines Vivian Y., Elementary Row Operattions and LU Decomposition, **The College Mathematics Journal**, Vol.21, No.5, p418 – 419.
21. Loan Van Chales F., A General Matrix Eigenvalue Algorithm, **SIAM Journal on Numerical Analysis**, Vol.12, No.6, p819 – 834.
22. Mandel J., Use of the Singular Value Decomposition in Regression Analysis, **The American Statistician**, 36, p15-24.
23. Markham T. L., Factorizations of Nonegitive Matrics, **Proceedings of the American Mathematical Society**, Vol.32, No.1, p45 – 47.
24. Nelder John A., an Alternative Interpretation of Singular Value Decompostion in Regression, **the American Statistician**, February 1985, 39, No.1, p63 – 65.
25. Paige C. C. and Saunders M. A. , Towards a Generalized Singular Value Decomposition, **SIAM Journal on Numerical Analysis**, Vol.18, No.3, 398 – 405.
26. Parlett B.N., the QR Algorithm, **Computing in Science & Engineering**, January / February, 2000, p38 – 42.
27. Parlett B. N. and Poole W. G., A Geometric Theory for the QR, LU and Power Iterations, **SIAM Journal on Numerical Analysis**, Vol.10, No.2, p389 – 412.
28. Piziak R. and Odell P.L., Full Rank Factorization of Matrics, **Mathematics Magazine**, Vol.72, No.3, p193-201.
29. Pursell Lyle and Trimble S. Y., Gram – Schmidt Orthogonalization by Gauss Elimination, **the American Mathematical Monthly**, Vol.98, No.6, p544 -549.
30. Sınıksaran Enis, **İstatistiksel Yöntemler**, Filiz Kitabevi, İstanbul, 2001.
31. Stewart G. W., **Matrix Algorithms –Volume I: Basic Decompositions**, SIAM, 1998, Philadelphia.
32. Stewart G.W., On the Early History of the Singular Value Decompostion, **SIAM Review**, Vol.35, No.4, p551-566.

33. Vaccaro Richard J., **SVD and Signal Processing II**, Elsevier Science Publisher Amsterdam, 1991.
34. Varah J. M., On the Numerical Solution of ill – Conditional Linear Systems with Application to ill Posed Problems, **SIAM Journal on Numerical Analysis**, Vol.10, No.2, p257 – 267.
35. Wang Guorong, Wei Yimin and Qiao S., **Generalized Inverse: Theory and Computations**, Science Press, Beijing,2003.
36. Watkins David.S., Understanding the QR Algorithm, **SIAM Review**, Vol.24,No.4 p 427-440.