Computational Mathematics

Dr. Cahit Karakuş

Giriş

- Yapay Zeka, kökü Matematik olan bir ağacın gövdesi ve dallarıdır.
- Yapay Zeka alanında kariyer yapmak ve başarılı olmak istiyorsanız temel uygulamalı matematik çalışmanız gerekmektedir; sadece bilim kurgu hayranı olmak yeterli olmayacaktır.
- Bir Yapay Zeka kariyeri inşa edeceksin matematikle arkadaş olun, göreceksiniz dünyanız sallanmaya başlayacaktır.
- Bir matematikçi ve yapay zeka uzmanı olarak, yapay zeka ve matematik arasındaki büyülü bağlantıyı sizinle paylaşmak istiyorum.

Lecture Outline

- Linear algebra
  - Vectors
  - Matrices
  - Özdeğerler ve Özvektörler
  - Eigen decomposition
- Türev
- İntegral
- Differential calculus
- Optimization algorithms
- Simulation
- Probability
  - Random variables
  - Probability distributions
- Information theory

Yapay Zeka - Matematik

- Yapay zekanın birincil amacı, insan anlayışı için kabul edilebilir bir model oluşturmaktır. Ve bu modeller, matematiğin çeşitli dallarından gelen fikir ve stratejilerle hazırlanabilir.
- Kendi kendine giden arabaları düşünün; amaçları, video görüntülerindeki nesneleri ve insanları tanımaktır. Kameralardan ve algılayıcılardan akan veri nehrinden beslenen veri denizinin içinde yüzecek bir araba düşünün. Bu arabaların arkasında minimizasyon süreçleri ve geri yayılım şeklinde matematik vardır. Matematik, bilim adamlarının yüzlerce yıldır bilinen geleneksel yöntem ve teknikleri kullanarak zorlu derin soyut problemleri çözmelerine yardımcı olur.
- Makine öğrenimi mühendisi, veri bilimcisi veya robot bilimcisi olarak kariyer yapmak istiyorsanız, matematikte başarılı olmanız gerekir. Matematik, yapay zekada hayati önem taşıyan analitik düşünme becerileri geliştirir.
- Tüm önemli ilerlemelerin arkasında matematik vardır. Lineer cebir, matematik, oyun teorisi, olasılık, istatistik, gelişmiş lojistik regresyon ve Gradient Descent kavramlarının tümü, veri biliminin temel dayanaklarıdır.
- Yapay zekada matematiğin üç ana dalı lineer cebir, hesaplamalı matematik ve olasılıktır.

Lineer Cebir

- Lineer Cebir, yapay zeka uzmanlarının onsuz yaşayamamayacağı bir şey olan uygulamalı matematik alanıdır. Bu alanda uzmanlaşmadan asla iyi bir yapay zeka uzmanı olamazsınız. Skyler Speakman'ın dediği gibi, “Lineer Cebir, 21. yüzyılın matematiğidir.”
- Lineer Cebirin skaler, vektörler, tensörler, matrisler, kümeler ve diziler yeni fikirler üretmeye yardımcı olur, bu yüzden yapay zeka uzmanları ve araştırmacıları için mutlaka öğrenilmesi gereken bir alandır. Topoloji, oyun teorisi, grafik teorisi, fonksiyonlar, doğrusal dönüşümler, özdeğerler ve özvektörler kavramları ile veri ve modelleri soyutlayabilirler.

Büyüklükler

- Skaler büyüklükler: Sayı ve birim kullanılarak belirtilebilen büyüklüklere skaler büyüklük denir. Örneğin 2 kg, 20 m gibi büyüklükler skaler büyüklüklerdir. Gerçek veya doğal tek bir sayı olabilir. Kütle, elektriksel yük, sıcaklık ve bir ortamdaki bir noktanın elektrik potansiyeli skaler niceliklerdir. 3 boyutlu evrende iki nokta arasındaki mesafe skalerdir, fakat birinden diğerine olan yön ile beraber belirtildiğinde vektörel bir nicelik olur. Fizikte ölçümü yapılan başlıca iki büyüklük vardır; skaler ve vektörel büyüklüklerdir.
- Vektorel büyüklükler: Büyüklüğü, başlangıç noktası, yönü ve ekseni ya da düzlemi olan büyüklüklere vektörel büyüklük denir.

Key Vocabulary

Expression – a math sentence ***without*** equal signs
Equation – a math sentence ***with*** equal signs
Variable – a letter that is used to represent a number
Terms – the part of an expression that is being added together
Coefficient – the number part of the term that has a variable with it
Constant Term – a term that has a number but no variable
Like Terms – terms that have identical variables
Inverse Operation – the opposite operation used to solve an equation

What is the difference between a numeric expression and an algebraic or variable expression?

Numeric Expression | Algebraic Expression

-3 + 2 + 4 – 5 | -3x + 2y – 4z - 5

An **algebraic or variable expression** consists of three parts

- Variable
- Coefficient
- Constant

Matematikte, tensör,

- Matematikte, tensör, çok boyutlu verinin simgelenebildiği geometrik bir nesnedir. Skaler denilen yönsüz nicel büyüklükler, vektör denilen yönlü büyüklükler ve matris denilen iki boyutlu nesneler birer tensördür. Tensör, tüm bu nesnelerin genelleştirilmiş halidir ve çok boyutlu veri kümeleri için kullanılır. Nesnenin kaç boyutla ifade edildiğine de tensörün derecesi denilir. Bir skalerin derecesi sıfır, bir vektörün bir, bir matrisin ise ikidir. Tensörler üç ve üzeri dereceye sahip olabilir. Tensörler — N eksenli düzenli bir ızgara üzerinde düzenlenmiş sayıların bir N-D dizisi (N>2). Makine Öğrenimi, Derin Öğrenme ve Bilgisayarla Görmede Önemlidir.

Matematikte matris

- Matematikte matris, dikdörtgen bir sayılar tablosu veya daha genel bir açıklamayla, toplanabilir veya çarpılabilir soyut miktarlar tablosudur. Matrisler daha çok doğrusal denklemleri tanımlamak, doğrusal dönüşümlerde (lineer transformasyon) çarpanların takibi ve iki parametreye bağlı verilerin kaydedilmesi amacıyla kullanılırlar. Matrislerin toplanabilir, çıkartılabilir, çarpılabilir, bölünebilir ve ayrıştırılabilir olmaları, doğrusal cebir kuramının temel kavramı olmalarını sağlamıştır.

Matematiksel Hesaplama

- Diferansiyel hesap, Çok değişkenli hesap, İntegral hesap, Gradyan iniş yoluyla hata minimizasyonu ve optimizasyonu, Limitler, Gelişmiş lojistik regresyonlar, matematiksel modellemede kullanılan tüm kavramlardır.

- Matematik, parametrelerdeki, fonksiyonlardaki, hatalardaki ve yaklaşıklıklardaki değişikliklerle ilgilenir. Yapay Zekada çok boyutlu hesaplamalarda, sapma bulmada, yörünge öngörmede kullanılan matematiksel hesaplamada en önemli kavramlar (ayrıntılı olmamakla birlikte):
- • Türevler — kurallar (toplama, çarpım, zincir kuralı vb.), hiperbolik türevler (tanh, cosh vb.) ve kısmi türevler.
- • Vektör/Matris Hesabı — farklı türev operatörleri (Gradient, Jacobian, Hessian ve Laplacian)
- • Gradyan (eğim) Algoritmaları — yerel/global maksimumlar ve minimumlar, eyer noktaları, dışbükey fonksiyonlar, yığınlar ve mini yığınlar, stokastik gradyan inişi ve performans karşılaştırması.

İstatistikte, temel bileşen analizi

- İstatistikte, temel bileşen analizi (TBA), çok boyutlu uzaydaki bir verinin daha düşük boyutlu bir uzaya izdüşümünü, varyansı maksimize edecek şekilde bulma yöntemidir. Uzayda bir noktalar kümesi için, tüm noktalara ortalama uzaklığı en az olan "en uygun doğru" seçilir. Daha sonra bu doğruya dik olanlar arasından yine en uygun doğru seçilerek, bu adımlar, yeni bir boyutun varyansı belirli bir eşiğin altına inene kadar tekrarlanır. Bu sürecin sonunda elde edilen doğrular, bir doğrusal uzayın tabanlarını oluşturur. Bu taban vektörlerine temel bileşen denir. Verinin temel bileşenleri birbirinden bağımsız olur. Bu kavram bazen orijinal terimin kısaltması olan PCA (İngilizce: Principal component analysis) olarak da anılır. TBA'nın ana kullanım amaçları keşifsel veri analizi yapmak ve kestirimsel modeller oluşturmaktır.
-
- “Temel Bileşenler Analizi” olan PCA tanıma, sınıflandırma, görüntü sıkıştırma alanlarında kullanılan yararlı bir istatistiksel tekniktir. Temel amacı yüksek boyutlu verilerde en yüksek varyans ile veri setini tutmak ancak bunu yaparken boyut indirgemeyi sağlamak olan bir tekniktir. Fazla boyutlu verilerdeki genel özellikleri bularak boyut sayısının azaltılmasını, verinin sıkıştırılmasını sağlar. Boyut azalmasıyla bazı özelliklerin kaybedileceği kesindir; fakat amaçlanan, bu kaybolan özelliklerin popülasyon hakkında çok az bilgi içeriyor olmasıdır. Bu yöntem, yüksek korelasyonlu değişkenleri bir araya getirerek, verilerdeki en çok varyasyonu oluşturan “temel bileşenler” olarak adlandırılan daha az sayıda yapay değişken kümesi oluşturur.

İstatistik & Olasılık Kavramları

- Yapay zeka dünyasında çok fazla soyut problem var. Belirsizlik ve stokastikliği birçok biçimde yaşayabilirsiniz. Olasılık teorisi, belirsizlikle başa çıkmak için araçlar sunar. Bir olayın meydana gelme sıklığını analiz etmek için, bir olayın meydana gelme şansı olarak tanımlandığı için olasılık kavramları kullanılır.

- Bir Robot düşünelim. Bir robot yalnızca belirli bir süre ileri gidebilir, ancak belirli bir mesafeye gidemez. Bilim adamları robotu ilerletmek için programlamasında matematiği kullanıyor. Ayrık rastgele değişkenler, sürekli rastgele değişkenler, Bayes Formülü ve normalleştirme, diğer lineer cebir kavramlarıyla birlikte Robotik navigasyon ve harekette kullanılan bazı olasılık kavramlarıdır.

- Bu konu muhtemelen zamanınızın önemli bir bölümünü alacaktır. İyi haber: Bu kavramlar zor değil, bu yüzden bu kavramlarda ustalaşmamanız için hiçbir neden yok.
- Temel İstatistikler — Ortalama, medyan, mod, varyans, kovaryans vb.
- Olasılık — olaylarda (bağımlı ve bağımsız), örnek uzaylarda, koşullu olasılıkta temel kurallar.
- Rastgele değişkenler — sürekli ve ayrık, beklenti, varyans, dağılımlar (ortak ve koşullu).
- Bayes Teoremi — inançların geçerliliğini hesaplar. Bayesian yazılımı, makinelerin kalıpları tanımasına ve karar vermesine yardımcı olur.
- Maksimum Olabilirlik Tahmini (MLE) — parametre tahmini. Temel olasılık kavramları hakkında bilgi gerektirir (ortak olasılık ve olayların bağımsızlığı).
- Ortak Dağılımlar — binom, poisson, bernoulli, gauss, üstel.

Bilgi Teorisi Kavramları

- Bu konu, AI ve Derin Öğrenmeye önemli katkılarda bulunan ve henüz birçok kişi tarafından bilinmeyen önemli bir alandır. Bunu matematik, istatistik ve olasılığın bir karışımı olarak düşünün.
- Entropi — Shannon Entropisi olarak da adlandırılır. Bir deneydeki belirsizliği ölçmek için kullanılır.
- Çapraz Entropi — iki olasılık dağılımını karşılaştırır ve bize bunların ne kadar benzer olduğunu söyler.
- Kullback Leibler Divergence — iki olasılık dağılımının ne kadar benzer olduğunun başka bir ölçüsü.
- Viterbi Algoritması — Doğal Dil İşleme (NLP) ve Konuşmada yaygın olarak kullanılır.
- Enkoder-Dekoder — Makine Çevirisi RNN'lerinde ve diğer modellerde kullanılır.

Fourier Transform

- Fourier dönüşümü (Fourier Transform) sıklık (frekans)analizinde kullanılan, istatistik tabanlı, matematiksel bir işlemdir.

- Zaman domenindeki karışık sinyal yumaklarını ayrıştırır ve hangi frekansta ne şiddette (genlik) bir sıklık olduğunu gösterir. Kısaca sinyallerimizi zaman alanından frekans alanına geçirirken kullandığımız bir işlemdir.

- Fourier dönüşümü peryodik olarak tekrarlanmayan sinyalleri dikkate almaz. Karmaşık sinyaller içinde periyodik olanları belirleyip harmonik bileşenlerine ayırır.

Fourier Transform

Fourier Transform:

$$F(j\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt$$ Analysis

Inverse Fourier Transform:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega) e^{j\omega t} d\omega$$ Synthesis

Continuous-Time Fourier Transform:

$$F(j\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt \qquad f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\omega) e^{j\omega t} d\omega$$

Discrete-Time Fourier Transform(DTFT):

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n] e^{-j\omega n} \qquad x[n] = \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} X(e^{j\omega}) e^{-j\omega n}$$

- Zamanla genliği değişen analog sinyallerindeki frekansların belirlenmeside ve frekansa göre genliklerinin ayrıştırma Fourier dönüşümü.
- Frekansların bileşenlerinden zaman domeninde analog sinyal elde edilmesi sentez. Frekanslara göre birleştirme.
- Peryodik Sinyallarin Fourier Dönüşümü daha doğru hesaplanır.
- Bir analog sinyalin Fourier dönüşümü komplks sayılar içerir.

Laplace Dönüşümü

- Doğrusal adi diferansiyel denklemlerin çözümü için önemli bir analitik yöntemdir. Doğrusal olmayan ODE'lerde ilk önce doğrusallaştırılmalıdır. Laplace dönüşümleri önemli proses kontrol kavramları ve tekniklerinde önemli bir rol oynar.
- Sistemlerin, sadece bir frekans veya zaman alanı verileri yerine, sinusoidler ve üsteller gibi doğal bileşenler açısından analiz edilmesine yardımcı olur.
- Yoğun aktarım işlevini, kutuplar ve sıfırlar cinsinden (limit) tanımlanabildiği ve sezgisel olarak sistemin kararlılığını, aşılmasını veya gürültüsünü tahmin etmede yardımcı olan uygun bir alana (S-etki alanı) çevirir.
- Kutuplar: Paydayı sıfır yapan değişkenlerdir.
- Sıfırlar: Payı sıfır yapan değişkenlerdir.
- Fourier Dönüşümü, şifreli bir dönüşümü çok daha kolay bir çarpıma dönüştürürken, bir Laplace Dönüşümü, S-etki alanında basit bir polinom cebiri ile yoğun bir diferansiyel denklemin çözülmesine yardımcı olur.
- Yinelemeli filtrelerin kararlılığını analiz etmek için DSP'de temel olan Z-Dönüşümü, Laplace Dönüşümü'nün yakın akrabasıdır.
- Örnekler: Transfer fonksiyonları, Frekans tepkisi, Kontrol sistemi tasarımı, Kararlılık analizi

Optimizasyon (Parametreleri Eniyileme)

- Performansa göre akıllı modellerdeki parametreleri en iyilime olarak bilinir.
- Makine öğrenmesi ile yakın bağı vardır: birçok öğrenme probleminde, bir öğrenme seti örneğindeki işlevlerin en aza indirilmesine odaklanır.
- Her hangi bir dalda (örneğin bilgisayar ya da telefonda) performansı artırmak ve daha iyi verim alabilmek için yani sistemi daha iyi bir hale getirmek için yapılan işlemlerin tümüne optimize (Eniyileme) ya da optimize etmek denir.
- Yeni girdiye aşırı duyarlılık gösteren parametreler; kaotik davranış.
- Giriş değerinde en küçük değişim, çıkışta çok büyük sapmaya neden olursa?
- Bir sistemin işlevini tanımlayan çok sayıda değişkenleri olan bir denklemde, değişken değerlerinin belirlenmesidir. Değiken değerlerin minimize edilmesi.
- Bir değişkenle eniyileme kolay, onlarca değişken eniyileme çok zordur. Burda optimizasyon devreye girer. Optime – kabul edilebilir değerlenden kabul edilebilir bir sonuç elde edilir.

Simülasyon (Benzerini Matematiksel Programlama ile Oluşturma)

- Bir sistemi temsil edebilecek ya da benzeri olabilecek bir matematiksel model geliştirme işlemidir.
- Ucuş simülasyonu, anten tasarım HFSS
- Gerçek sistemin modelinin tasarımlanması ve bu model ile sistemin işletilmesi amacına yönelik olarak , sistemin davranışını anlayabilmek veya değişik stratejileri değerlendirebilmek için deneyler yürütülmesi sürecidir.

- Geliştirilen veya yeniden düzenlenen süreçleri tamamlamada ve deneme çalışmalarını yürütmede ve süreçlerin hata zamanlarını tahmin etmek için yapılan deneysel çalışmadır .Yeni sürecin değişikliklere gösterdiği olası reaksiyonları da anlayabiliriz.

Simülasyon

Simülasyon, **gerçek dünyayı taklit eden** matematiksel ve algoritma tabanlı dijital ortam oluşturma sürecidir.
Simülasyonlar, matematiksel modeller ve algoritmalar kullanarak gerçeklikten esinlenir. Bu sayede, farklı senaryoları deneyimlemek ve sonuçlarını tahmin etmek mümkün olur.
Simülasyonlar, eğitimden sağlık sektörüne, mühendislikten oyunlara kadar birçok alanda kullanılır.
Gerçek sistemin yapısı ve davranışını anlayabilmek için hesaplamalı matematik uygulamalar temelli mantıksal ve matematiksel modelleme ile deney yapma olanağı sağlayan bir yazılımsal bir sistemdir.

Bilgi Teorisi Kavramları

- Bu konu, AI ve Derin Öğrenmeye önemli katkılarda bulunan ve henüz birçok kişi tarafından bilinmeyen önemli bir alandır. Bunu matematik, istatistik ve olasılığın bir karışımı olarak düşünün.
- Entropi — Shannon Entropisi olarak da adlandırılır. Bir deneydeki belirsizliği ölçmek için kullanılır.
- Çapraz Entropi — iki olasılık dağılımını karşılaştırır ve bize bunların ne kadar benzer olduğunu söyler.
- Kullback Leibler Divergence — iki olasılık dağılımının ne kadar benzer olduğunun başka bir ölçüsü.
- Viterbi Algoritması — Doğal Dil İşleme (NLP) ve Konuşmada yaygın olarak kullanılır.
- Enkoder-Dekoder — Makine Çevirisi RNN'lerinde ve diğer modellerde kullanılır.

Lineer Cebir

22

Özdeğerler ve Özvektörler

- Özdeğerler (veya karakteristik değerler) ve özvektörleri (karakteristik özvektörler), fiziksel bir sistemin sahip olabileceği özel değerlerde nasıl davrandıklarını belirlemek için önemlidir.
- Bu değerler sisteme ait özel bir enerji, özel bir frekans değeri, dalgaların girişimi veya kuvvet dengesinin sağlandığı bir duruma ait olabilir.
- Özvektörler ise, fiziksel sistemin sahip olduğu özdeğerlerdeki (örneğin bir dalga) fonksiyonları olabilir. Özdeğerler ve özvektörler, diferensiyel denklemler içeren denklem sistemlerinin çözümlerinde, sınır-değer problemlerinde ortaya çıkabilir.
- Bu tür denklemlere, kuantum mekaniğinde elektriksel bir potansiyel içinde bulunan bir parçacığın enerjisini hesaplarken, elastik çarpışma problemlerinde, akışkanlar mekaniğinde, titreşim yapan cisimlerin hareketlerinde sıkça karşılaşılır.
- Titreşim yapan bir sistemin doğal frekansı ile dışarıdan uygulanan sürücü kuvvetin frekansı birbirineeşit veya yakın olması sistemin kararlılığı açısından veya malzemelerin elastik özelliklerinin incelendiği durumlarda, şekil bozukluklarının başladığı noktaların belirlenmesinde özfonksiyonların alacağı özdeğerler önemli olmaktadır.

Özvektörler ve Özdeğerler

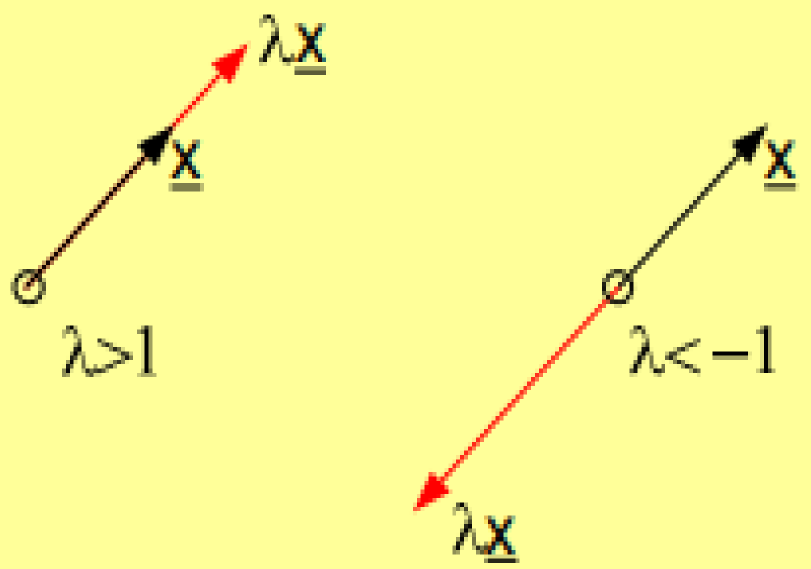
- Özvektörler ve Özdeğerler — özel vektörler ve bunlara karşılık gelen skaler nicelik. Mühendislikte sık sık karşılaşılan problemlerin öznesidir. Bir vektörün özdeğerleri o matrisin karakteristik polinomunun kökleridir. Özdeğerler, özvektörler ve özuzaylar, bir matrisin özellikleridir ve matris hakkında önemli bilgiler verir. Matrislerin çarpanlarına ayrılmasında kullanılabilirler. Uygulamalı matematik alanlarında olduğu kadar finans ve kuantum mekaniğinde de kullanılır. Bir vektör üzerine uygulanan matris o vektörün hem büyüklüğünü, hem de yönünü değiştirir. Buna rağmen, bir matris bazı belirli vektörler üzerine etkidiğinde onların sadece büyüklüğünü değiştirir, doğrultularını değiştirmez (ancak vektörün yönü ters çevrilebilir). Doğrultusu değişmeyen bu vektörler söz konusu matrisin özvektörleri olarak adlandırılır. Bir matris, bir özvektörü üzerine etkidiğinde onun büyüklüğünü bir çarpan kadar katlar. Bu çarpan pozitif ise vektörün yönü değişmeden kalır, negatif ise vektörün yönü tersine döner (her iki durumda da vektörün doğrultusu değişmez). Bu çarpana, söz konusu özvektöre ilişkin özdeğer denir. Bir özuzay, aynı özdeğere sahip tüm özvektörlerin oluşturduğu kümedir. Bu kavramlar matrisler, vektörler ve doğrusal dönüşümler üzerinde tanımlıdır.

Özdeğer, Özvektör

- Özdeğerler, bir matrisin orijinal yapısını görmek için kullanılan alternatif bir yoldur.
- Bazı vektörler bir **A** matrisi ile çarpıldıkları zaman yön değiştirir, bazıları ise değiştirmezler.
- **Bazı özel x vektörleri, Ax vektörü ile aynı yönde kalmaktadır. İşte bu vektörlere "özvektörler" denir.**
- Bir özvektörün **A** matrisi ile çarpımı olan A**x** vektörü, orijinal **x** vektörünün λϵR olmak üzere λ katıdır.

Özdeğer ve özvektörün geometrik yorumu

$\underline{A}\ \underline{x}=\lambda\underline{x}$ bağıntısından hesaplan λ özdeğeri ve $\underline{x}$ özvektörü şu şekilde yorumlanabilir: $\underline{A}$ matrisi $\underline{x}$ vektörünü λ kadar büyütmekte veya küçültmektedir. $\underline{x}$ vektörünün doğrultusu değişmemekte fakat yönü değişebilmektedir. λ pozitif ise $\underline{x}$ ve $\lambda\underline{x}$ aynı yönde, aksi hale ters yöndedirler.



Özdeğer ve özvektörlerin özellikleri

Özdeğer ve özvektörlerin bazı önemli özellikleri

1. Özdeğer problemi sadece kare matrisler için tanımlıdır.
2. $\underline{A}_{nxn}$ matrisinin daima n tane özdeğeri, $\lambda_1, \lambda_2, \ldots, \lambda_n$ vardır.
3. λ_i özdeğeri $\underline{A}$ matrisinin determinantını sıfır yapar.
4. Her λ_i özdeğerine karşılık gelen bir $\underline{x}_i$ özvektörü vardır. λ_i ve $\underline{x}_i$ çifti beraber $(\underline{A} - \lambda_i \underline{I})\underline{x}_i = \underline{0}$ bağıntısını sağlar.
5. Özdeğerler pozitif, negatif, sıfır gerçek sayıları olabildiği gibi sanal sayılar da olabilir.
6. Özvektörlerin elemanları gerçek ve sanal sayılardan oluşabilir.
7. Elemanları gerçek sayılardan oluşan $\underline{A}$ simetrik ($\underline{A}^T=\underline{A}$) ise, tüm özdeğerler de gerçek sayılardan oluşur. Simetrik matrisin özvektörleri ortogonaldır: $\underline{X}^T\underline{X}=\underline{I}$
8. $\underline{A}$ simetrik ($\underline{A}^T=\underline{A}$) ve pozitif tanımlı ise tüm özdeğerler de pozitiftir.
9. Bazı özdeğerler birbirine eşit olabilir. Fakat Eşit özdeğerlerin özvektörleri mutlaka farklıdır. Çünkü özvektörler doğrusal bağımsızdır. Örneğin

$\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}$ birim matrisinde $\lambda_1 = \lambda_2 = \lambda_3 = 1$ ve özvektörler $\underline{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cdot \ \underline{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \cdot \ \underline{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

dır. Görüldüğü gibi, özdeğerler birbirine eşit fakat özvektörler birbirinden farklıdır.

Özdeğer ve özvektörlerin özellikleri

10. $\underline{x}_i$ özvektörünün herhangi bir gerçek c sayısı ile çarpılması veya bölünmesi sonucunda elde edilen yeni vektör de bir özvektördür. Yanı yeni vektör $c\underline{x}_i$ ile de $(\underline{A}-\lambda_i \underline{I})(c\underline{x}_i)=\underline{0}$ sağlanır. Bu önemli özellik nedeniyle, istenirse, c herhangi bir gerçek sayı seçilebilir.
11. $\underline{A}$ ve $\underline{A}^T$ aynı özdeğerlere sahiptir, fakat özvektörleri genelde farklıdır.
12. $\underline{A}_{nxn}$ ve $\underline{B}_{nxn}$ kare matrisler olmak üzere $\underline{A}\,\underline{B}$ ve $\underline{B}\,\underline{A}$ matrisleri aynı özdeğerlere sahiptir.
13. $\underline{A}$ nın özdeğeri λ_i ise $\underline{A}^{-1}$ in özdeğeri $1/\lambda_i$ dir. $\lambda_i=0$ durumunda $1/\lambda_i$ tanımsızdır, bu ise $\underline{A}$ nın tekil ve $\underline{A}^{-1}$ in olmadığı anlamındadır.
14. $\underline{A}_{nxn}$ matrisinin izi özdeğerlerin toplamına eşittir:

    İz $\underline{A}=a_{11}+a_{22}+\ldots+a_{nn}=\lambda_1+\lambda_2+\ldots+\lambda_n$.
15. $\underline{A}_{nxn}$ matrisinin determinantı özdeğerlerin çarpımına eşittir:

    det $\underline{A}=\lambda_1\cdot\lambda_2\cdot\ldots\cdot\lambda_n$.

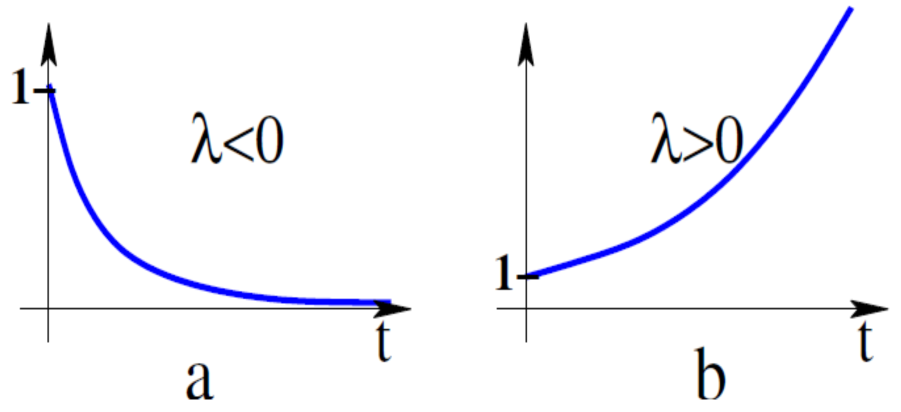
    Dolayısıyla, özdeğerlerden herhangi biri sıfırsa, $\lambda_i=0$, det $\underline{A}$=0 dır ve $\underline{A}^{-1}$ tanımsızdır.

Özdeğerlerden kararlı ya da kararsız durumların belirlenmesi

- λ_1 ve λ_2 özdeğerlerine bağlı olarak birkaç farklı denge türü ortaya çıkmaktadır. Bilindiği gibi λ_1 ve λ_2, genel ikinci dereceden bir denklem olan $|A - \lambda I| = 0$, karakteristik denklemin kökleridir. Bu nedenle kökler gerçek veya karmaşık sayılar olabilir ve aşağıdaki denge kararlılığı durumları ortaya çıkarır.

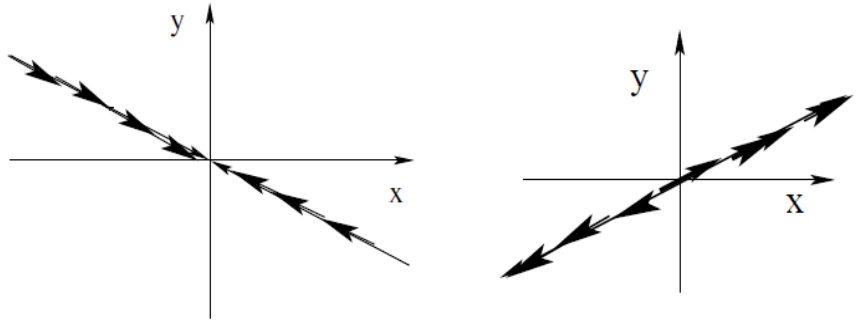
- Özdeğerler, çözüm denkleminde $e^{\lambda_1 t}$ ve $e^{\lambda_2 t}$ katkı verirler. Bu durumda üssel negatif özdeğeri olan fonksiyon $t \to \infty$ olurken çözüm eğrisine yakınsar. üssel pozitif özdeğeri olan fonksiyon $t \to \infty$ olurken çözüm eğrisinden ıraksar.

- $e^{\lambda t}$, fonksiyonunun iki ana davranış türü vardır.
- $\lambda<0$ iken t artığında $e^{\lambda t}$ kararlı bir noktaya yaklaşır.
- $\lambda>0$ iken t artığında $e^{\lambda t}$ kararsız sonsuza gider.



Özdeğerlerden kararlı ya da kararsız durumların belirlenmesi

- Bu durumda,
- $\lambda<0$ iken, aşağıdaki denklem kararlı bir durum belirleyecektir.
- $\lambda>0$ iken, aşağıdaki denklem kararsız bir durum belirleyecektir.
- $\begin{pmatrix} x \\ y \end{pmatrix} = C \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} e^{\lambda t}$



a   b

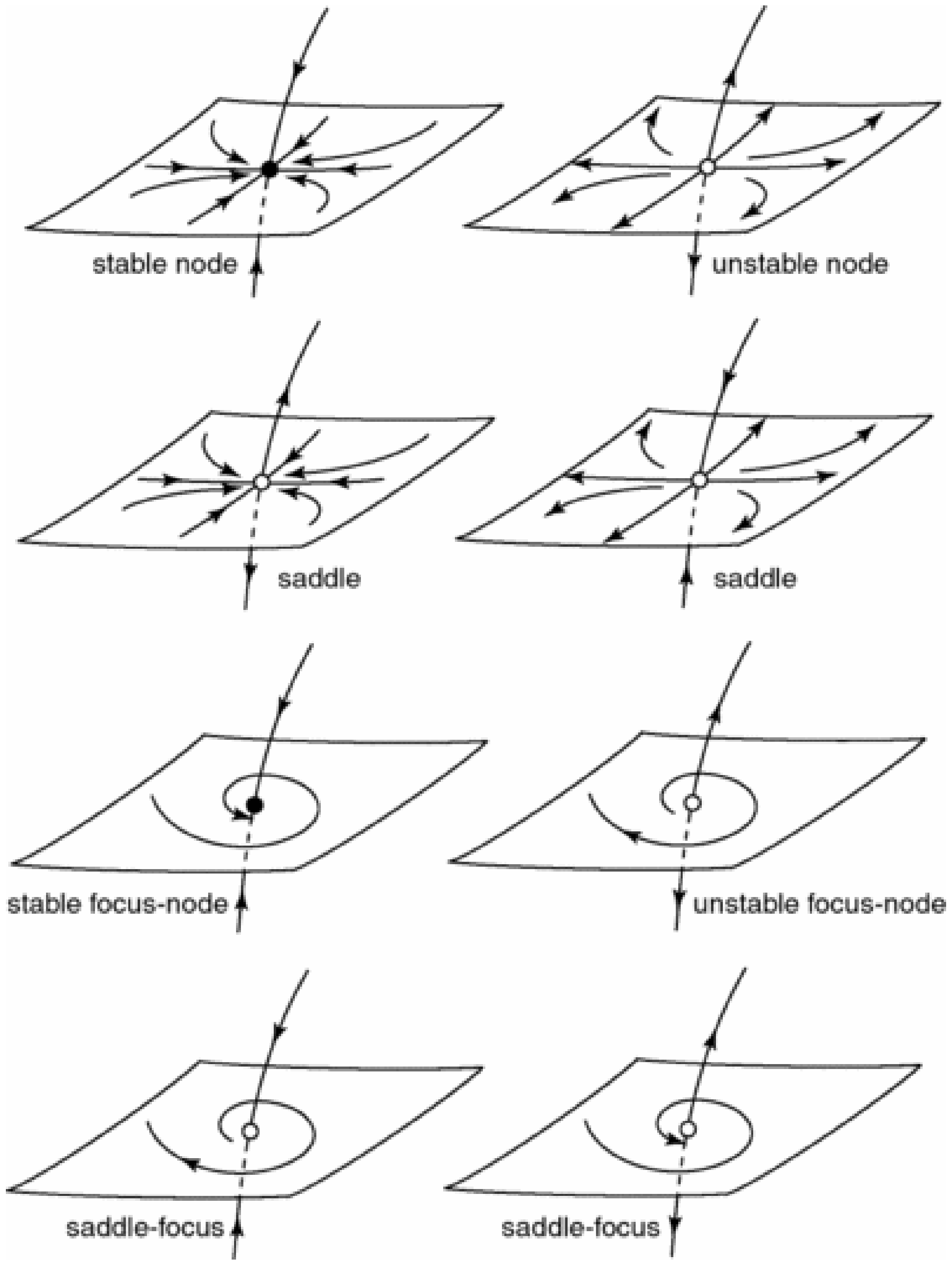
- Kararlı durumlar:
- Kararlı düğüm noktası: $\lambda_1<0$ ve $\lambda_2<0$; kökler reel.
- Kararlı spiral: $\lambda_{1,2} = \alpha \pm i\beta$; $\alpha<0$;

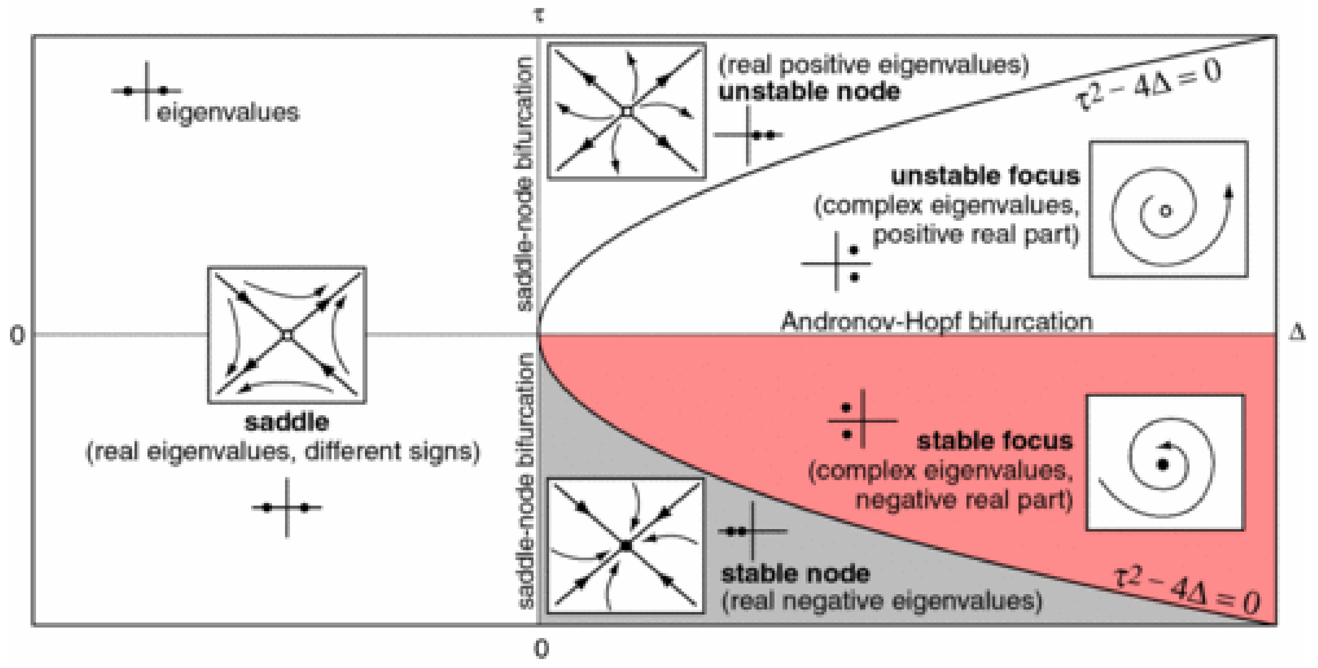
- Kararsız durumlar:
- Kararsız düğüm noktası: $\lambda_1>0$ ve $\lambda_2>0$; kökler reel
- Kararsız spiral: $\lambda_{1,2} = \alpha \pm i\beta$; $\alpha>0$
- Dairesel spiral: $\lambda_{1,2} = \pm i\beta$; $\alpha=0$;
- Eyer ya da denge noktası: $\lambda_1>0$, $\lambda_2<0$ ya da $\lambda_1<0$, $\lambda_2>0$; kökler reel

Denge – Dönüm - Eyer Noktaları

- Bir f fonksiyonunun konvekslikten konkavlığa veya konkavlıktan konveksliğe geçtiği ve fonksiyonunun sürekli olduğu noktaya dönüm (büküm) noktası adı verilir.
- Denge, değişmeyen bir sistem durumudur.
- İki boyutlu bir uzayda kararlı denge: düğüm ve odak
- Bir sistemin dinamikleri bir diferansiyel denklem veya bir diferansiyel denklem sistemi ile tanımlanıyorsa, o zaman denge noktası bir türevi (tüm türevleri) sıfıra ayarlayarak tahmin edilebilir.
- Özdeğerlerin hiçbirinin gerçek kısmı yoksa denge noktası hiperboliktir.
- Tüm özdeğerlerin negatif gerçek kısmı varsa, denge kararlı bir denklemdir.
- En az birinin pozitif bir gerçek kısmı varsa, denge kararsız bir düğümdür.
- En az bir özdeğerin negatif gerçek kısmı varsa ve en az birinin pozitif gerçek kısmı varsa, denge bir eyer noktasıdır.

Denge – Dönüm - Eyer Noktaları





Denge durum noktaları ve faz düzlemleri

- Lineer veya lineer olmayan tipten $\ddot{x} = f(t, x, \dot{x})$, denklemler teorisinde x(t) çoğu kez x-ekseninde hareket eden bir noktanın t anındaki yerini ve y(t)= $\dot{x}(t)$ de t anındaki hızını tanımlar. (x(t), y(t)) ikilisi, birlikte, sistemin t anındaki durumunu belirler.
-
- Sistemin davranışı, (x,y) -düzleminde (x(t), y(t)) noktasının geometrik yeri ile tarif edilebilir. Bu biçimde diferansiyel denklem ile ilişkilendirilen (x,y)-düzlemi faz düzlemi olarak adlandırılır. (x(t), y(t)) parametrik çözüm eğrisine yörünge ve onun görüntüsüne de orbit veya iz denir. Bir yörünge ile orbit arasındaki fark, yörüngenin çözüm eğrisinin oryantasyonunu veren t parametresi ile donatılmış olmasıdır.
-
- $\ddot{x} = f(t, x, \dot{x})$, genel denklemi için, denge noktaları x ekseninde bulunur ve tüm f(x, 0) = 0 çözümleri tarafından tanımlanır. (x , y) (y= $\dot{x}$) düzlemdeki faz yolları birinci mertebeden denklemin çözümleri yardımıyla belirlenir.

Denge durum noktaları ve faz düzlemleri

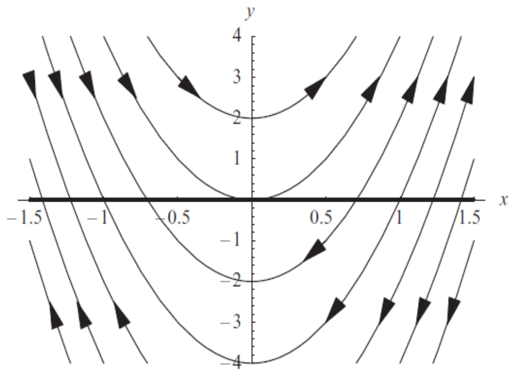
- $y(t) = \dot{x} = \frac{dx}{dt}$
-
- $\ddot{x} = \frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt} = y\frac{dy}{dx} = f(x, y)$
-
- $\frac{dy}{dx} = \frac{f(x,y)}{y}$
-
- $\frac{dy}{dx} = \frac{\ddot{x}}{\dot{x}}$
-
- x ve y eksenlerindeki ölçeklerin her zaman aynı olmadığı unutulmamalıdır. Genellikle eş zamanlı diferansiyel denklemler olarak muamele edilen f(x, y) çözümü olan (x (t), y (t)), t cinsinden parametrik olarak elde edilir.
-

Denge durum noktaları ve faz düzlemleri

- $\ddot{x} - 8\dot{x}x = 0$ diferansiyel f(x, y) = 8xy olur. f(x, 0) = 0 olduğundan, x eksenindeki her nokta bir denge noktasıdır. Faz yolları için diferansiyel denklem,
- $\ddot{x} = 8\dot{x}x$
- $\dot{x} = y$
- $\dot{y} = 8xy$
-
- $\frac{dy}{dx} = 8x,$ genel çözüm, $y=4x^2 + C$ dir.
- f(x, y) = 8xy denkleminin faz çözüm çizgileri aşağıdaki şekilde verilmiştir.

Matlab Yazılım kodu:

```
clear all
close all
x = linspace(-1.5,1.5,15);
y = linspace(-4,4,15);
[M1,M2]=size(y);
[X,Y] = meshgrid(x,y);
U = Y;
for i=1:M2
    for j=1:M2
   V(i,j)=-8*X(i,j)*Y(i,j);
    end
end
figure(1), quiver(X,Y,U,V,'r')
```

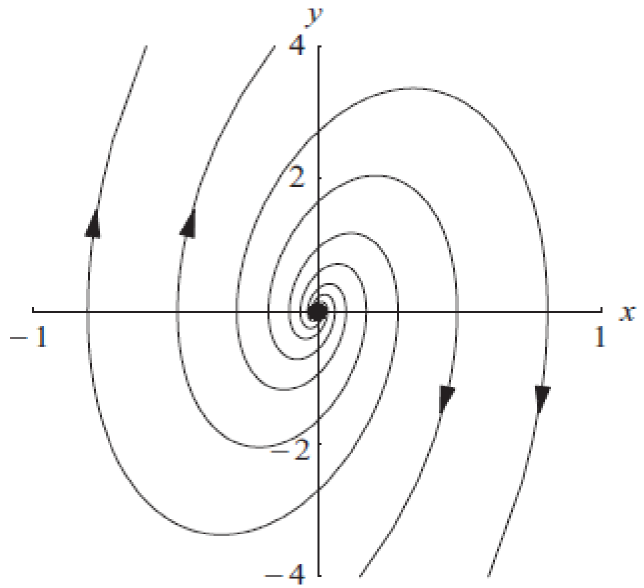


Denge durum noktaları ve faz düzlemleri

- $\ddot{x} - 4\dot{x} + 40x = 0$ diferansiyel f(x, y) = 4y-40x olur. f(x, 0) = 0 olduğundan, x eksenindeki her nokta bir denge noktasıdır. Faz yolları için diferansiyel denklem,
- f(x, y) =4y-40x denkleminin faz çözüm çizgileri aşağıdaki şekilde verilmiştir.

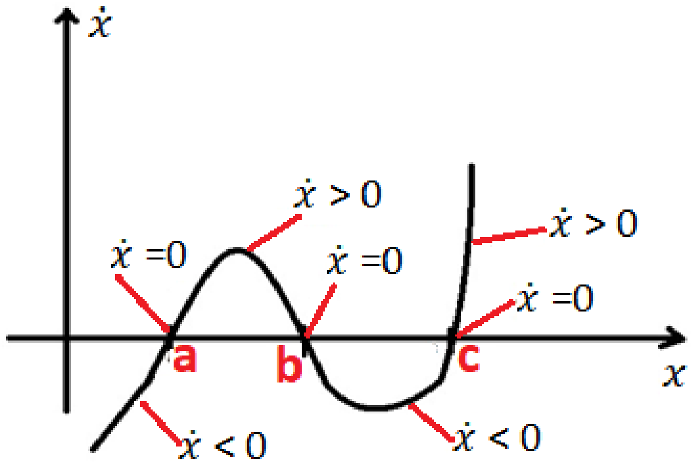
Matlab Yazılım Kodu:

```
clear all
close all
x = linspace(-1,1,20);
y = linspace(-4,4,20);
[M1,M2]=size(y);
[X,Y] = meshgrid(x,y);
U = Y;
for i=1:M2
   for j=1:M2
   F(i,j)=-40*X(i,j)+4*Y(i,j);
   end
end
figure(1), quiver(X,Y,U,F,'r')
```



Otonom Denklemler

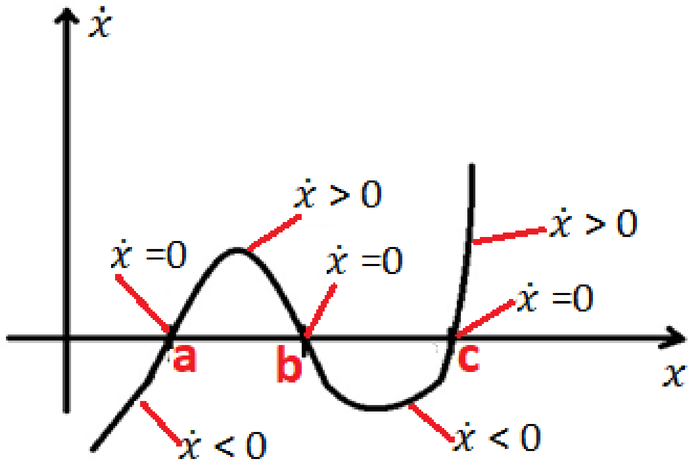
- Otonom denklemlerin genel formu, $\dot{x}$= f(x) dır. Bir diferansiyel denklemin otonom olabilmesi için türev eşitliğinin karşısında fonksiyonda sadece x ifadeleri olmasıdır. Başka bir değişken olmamasıdır. Bu formdaki denklemlere otonom difarensiyel denklemler adı verilir. Otonom diferansiyel denklemlerin davranışı genellikle ($\dot{x}$, x) grafiği çizilerek anlaşılmaya çalışılır. Davranıştan kastedilen çözüm eğrilerini tahmin etmektir. ($\dot{x}$, x) grafiği yorumlanarak çözüm eğrilerinin görünümü tahmin edilebilir. Grafik çizilerek çözüm eğrilerinin tahmin edilmesine faz düzlem analizi denir.



Öncelikle ($\dot{x}$, x) grafiği çizilir.

Otonom Denklemler

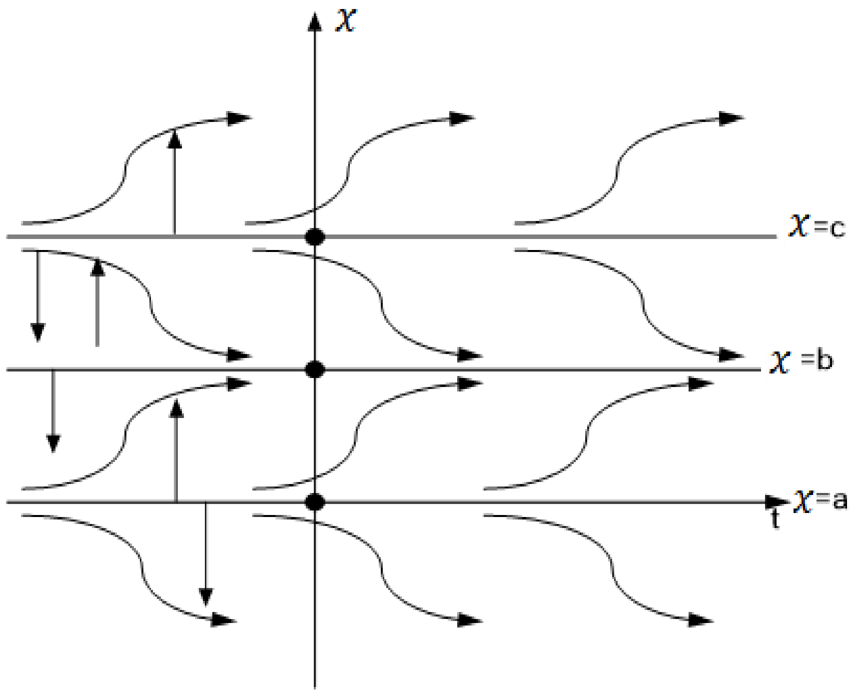
- Bir grafik çizildikten sonra yorumlar yapılır; x =a, x =b, x =c değerleri $\dot{x}$ =f(x) fonksiyonunun kökleridir.
- x <a ise $\dot{x}$ <0
- x =a ise $\dot{x}$ =0
- a< x <b ise $\dot{x}$ >0
- x =b ise $\dot{x}$ =0
- b< x <c ise $\dot{x}$ <0
- x =c ise $\dot{x}$ =0
- x >c ise $\dot{x}$ >0



Öncelikle ($\dot{x}$, x) grafiği çizilir.

Otonom Denklemler

- Grafiğe bakılarak ve elde edilen bilgilerden yararlanılarak çözüm eğrilerinin şekli tahmin edilebilir.$\dot{x}$=0 ise çözüm eğrileri eğimi sıfır olan düz birer çizgi olacaktır. Bu bilgiler kullanılarak basitçe çözüm eğrileri çizilebilir. Burada x'in bağlı olduğu değişken olarak genellikle t kullanılır. Yukarıda faz düzlem analizinde $\dot{x}$-x grafiği çizilip yorumlanabilir. Aşağıda ise çözüm eğrilerini gösteren x-t grafiği çizilecektir. Şöyleki; $\dot{x}$>0 ise çözüm eğrileri yukarı doğru meyilli olacaktır. Dolayısıyla buna çizilen teğet yani türev pozitif olmalıdır.Teğetin eğimi aşağıdaki şekildeki gibi pozitiftir. $\dot{x}$<0 ise çözüm eğrileri aşağı meyilli olacaktır.



Otonom Denklemler

- x=a , x=b ve x=c noktalarında türevimiz “0” a eşit olduğu için eğrilerimizin eğimi sıfır olan düz çizgiler olmak zorundadır. Düz olarak çizdiğimiz çözüm eğrilerine denge çözümleri adı verilir. Dikkat edin hiçbir çözüm eğrisi birbirine değmemelidir. Denge çözümleri ile de kesişmez.

- Grafikte görülen çözüm eğrilerinin şekilleri böylelikle tahmin edilir. Faz düzlem analizi (x’ , x) grafiğini inceleyerek bu sonuçlara vardık. Diğer çözümler denge çözümlerine (x’=0 olan çizgilere) yaklaşmaya çalışır. Sanki diğer çözüm eğrileri bir şekilde dengesini arıyor ve oraya doğru yaklaşmaya çalışıyor. Bu yüzden denge çözümleri olarak adlandırılmıştır. Bu denge çözümleri örneğimiz için; x=a, x=b ve x=c olmak üzere üç tanedir.

Denge çözümlerinin 3 çeşidi vardır:

- Kararlı denge çözümüleri
- Kararsız denge çözümleri
- Yarı kararlı denge çözümleri

Otonom Denklemler

-
- Bunları anlamak için tekrar grafiğe bakılırsa, x=c denge çözümüne bakılırsa altında ve üstünde kalan çözüm eğrilerine göre yorumlanır. t sonsuza giderken altındaki ve üstündeki iki çözüm eğrisi de x=c denge çözümünden uzaklaşmaktadır. Dolayısıyla “x=c” ye kararsız denge çözümü denir. Kararsız dengede, çözümden biraz dışarı çıkıldığında çözümler dengeden uzaklaşır.

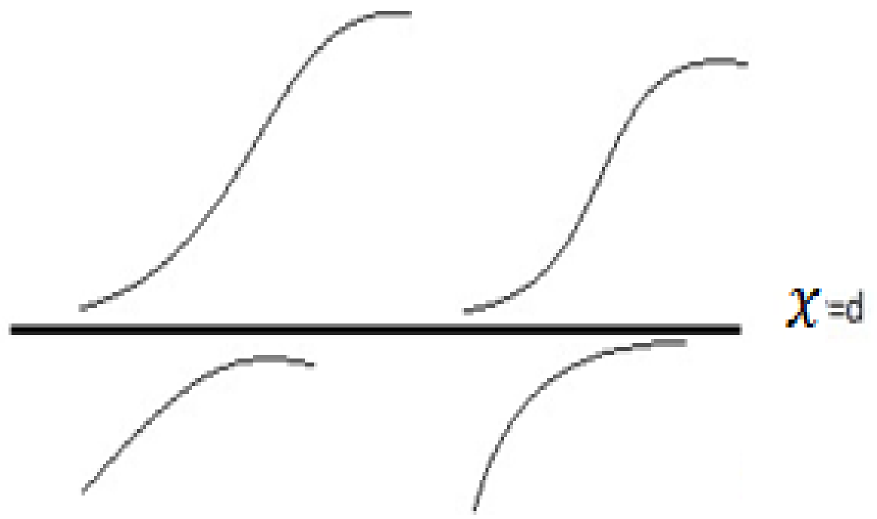
- x=b denge çözümüne bakarsak, t sonsuza giderken çözümden biraz dışarı çıkıldığında alttaki ve üsttteki iki çözüm eğrisi de x=b denge çözümüne yaklaşmaktadır. Dolayısıyla “x=b” ye kararlı denge çözümü denir. Kararlı denge çözümünde, çözümden biraz dışarı çıkıldığında çözümler dengeye yaklaşma eğilimi gösterir.

- x=a denge çözümüne bakarsak, t sonsuza giderken altında ve üstünde kalan iki çözüm eğrisi de “x=a” çözümünden uzaklaşmaktadır. Dolayısıyla “x=a” ya kararsız denge çözümü denir.

Otonom Denklemler

- Yarı kararlı denge çözümününü göstermek gerekirse "x=d" de denge çözümümüz var diyelim.

- t sonsuza doğru giderken "x=d" doğrusundan yukarı doğru yaklaştığında "x=d" doğrusundan uzaklaşırken. Aşağıdaki çözüm eğrisine yaklaşıldığında "x=d" doğrusuna yakınlaşmaktadır. Yarı kararlı denge çözümü= çözüm eğrilerinden biri çözüm doğrusuna yaklaşırken diğeri uzaklaşıyorsa buna yarı kararlı denge çözümü denir.



Differential Calculus

42

Türevin Yorumu

Birinci Türev

- Birinci ve ikinci türevlerinin verdiği bilgilerden f’(x) veya df/dx olarak yazılan f(x) fonksiyonunun ilk türevi, x noktasındaki teğet çizgisi fonksiyonun eğimidir.
- Grafik olmayan terimlerle ifade etmek gerekirse, ilk türev bize bir fonksiyonun nasıl arttığını veya azaldığını ve ne kadar artacağını veya azalacağını söyler.
- Pozitif eğim bize x arttıkça f(x)'nin de arttığını söyler. Negatif eğim bize x arttıkça f(x)’in azaldığını söyler. Sıfır eğim bize özel bir şey söylemez: fonksiyon o noktada artar, ne azalır veya yerel maksimumda veya yerel minimumda olabilir.

Türevler açısından bu bilgileri yazarken şunu görüyoruz:

-
- $\frac{df(p)}{dx} > 0,$ ise f(x), x = p’de artan bir fonksiyondur.
- $\frac{df(p)}{dx} < 0,$ ise f(x), x = p’de azalan bir fonksiyondur.
- $\frac{df(p)}{dx} = 0,$ ise o zaman x = p, f (x)'in kritik noktası olarak adlandırılır ve x(p)’deki f (x) 'nin davranışı hakkında yorum yapılamaz.

Türevin Yorumu

İkinci Türev

- Bir fonksiyonun ikinci türevi, f’’ (x) veya $\frac{d^2 f}{dx^2}$ olarak yazılır. İlk türev bize fonksiyonun arttığını veya azaldığını söylese de, ikinci türev,
- x = p'de $\frac{d^2 f(p)}{dx^2}$ >0 ise, f(x), x = p'de yukarı doğru kavislidir.
- x = p'de $\frac{d^2 f(p)}{dx^2}$ <0 ise, f(x), x = p'de aşağı doğru kavislidir.
- x = p'de $\frac{d^2 f(p)}{dx^2}$ =0 ise, o zaman f(x) 'in x = p'deki davranışı hakkında bir yorum yapamıyoruz.
-
- Birinci türevin anlamından x, f(x) fonksiyonunun kritik bir noktası olduğunda, o noktada fonksiyonun davranışı hakkında bir yorum yapabilmek için, x’in bölgesel maksimum veya bölgesel minimum olduğunu öğrenmek için genellikle işlevin ikinci türevi kullanılır.

Yönlü türev ve Gradyan

- Bir skaler alanın **yön türevi** (gradyan) artımın en çok olduğu yere doğru yönelmiş bir vektör alanını verir ve büyüklüğü değişimin en büyük değerine eşittir.
- Bir scaler field'in maksimum artış gösterdiği yönü ve hızı belirten bir vektör verir.
- Bir noktada hesaplandığında, o noktada fonksiyonun maksimum artış gösterdiği yönü gösteren vektördür.

Gradient

- Differential Calculus
  - We can concatenate partial derivatives of a multivariate function with respect to all its input variables to obtain the ***gradient*** vector of the function
  - The gradient of the multivariate function $f(\mathbf{x})$ with respect to the *n*-dimensional input vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$, is a vector of *n* partial derivatives

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n}\right]^T$$

  - When there is no ambiguity, the notations $\nabla f(\mathbf{x})$ or $\nabla_{\mathbf{x}} f$ are often used for the gradient instead of $\nabla_{\mathbf{x}} f(\mathbf{x})$
    - The symbol for the gradient is the Greek letter ∇ (pronounced "nabla"), although $\nabla_{\mathbf{x}} f(\mathbf{x})$ is more often it is pronounced "gradient of *f* with respect to **x**"
  - In ML, the gradient descent algorithm relies on the opposite direction of the gradient of the loss function $\mathcal{L}$ with respect to the model parameters θ ($\nabla_\theta \mathcal{L}$) for minimizing the loss function
    - Adversarial examples can be created by adding perturbation in the direction of the gradient of the loss $\mathcal{L}$ with respect to input examples x ($\nabla_x \mathcal{L}$) for maximizing the loss function

- (Kısmi) türevler bize bir değişkenin diğerini ne kadar etkilediğini söyler.

Evaluating the Gradient

As an example, given the function $f(x, y) = 3x^2y - 2x$ and the point (4, -3), the gradient can be calculated as:

$$[6xy - 2 \quad 3x^2]$$

Plugging in the values of x and y at (4, -3) gives

$$[-74 \quad 48]$$

which is the value of the gradient at that point.

Bu noktadaki teğet düzlemin eğimi x yönünde -74, y yönünde +48 olacaktır.
Gradyan vektörünün yönü her zaman fonksiyonun en dik artış yönünü gösterecektir. Ve büyüklüğü bize düzlemin o yöndeki eğimini verecektir.

Örnek

Let $f(x,y) = x^2 y$. (a) Find $\nabla f(3,2)$. (b) Find the derivative of f in the direction of (1,2) at the point (3,2).

Solution: (a) The gradient is just the vector of partial derivatives. The partial derivatives of f at the point $(x,y) = (3,2)$ are:

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= 2xy & \qquad \frac{\partial f}{\partial y}(x,y) &= x^2 \\ \frac{\partial f}{\partial x}(3,2) &= 12 & \qquad \frac{\partial f}{\partial y}(3,2) &= 9 \end{aligned}$$

Therefore, the gradient is

$$\nabla f(3,2) = 12\mathbf{i} + 9\mathbf{j} = (12, 9).$$

(b) Let $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j}$ be a unit vector. The directional derivative at (3,2) in the direction of $\mathbf{u}$ is

$$\begin{aligned} D_{\mathbf{u}} f(3,2) &= \nabla f(3,2) \cdot \mathbf{u} \\ &= (12\mathbf{i} + 9\mathbf{j}) \cdot (u_1\mathbf{i} + u_2\mathbf{j}) \\ &= 12u_1 + 9u_2. \end{aligned} \tag{1}$$

To find the directional derivative in the direction of the vector (1,2), we need to find a unit vector in the direction of the vector (1,2). We simply divide by the magnitude of $(1,2)$.

$$\mathbf{u} = \frac{(1,2)}{\|(1,2)\|} = \frac{(1,2)}{\sqrt{1^2+2^2}} = \frac{(1,2)}{\sqrt{5}} = (1/\sqrt{5}, 2/\sqrt{5}).$$

Plugging this expression for $\mathbf{u} = (u_1, u_2)$ into equation (1) for the directional derivative, and we find that the directional derivative at the point $(3,2)$ in the direction of $(1,2)$ is

$$\begin{aligned} D_{\mathbf{u}} f(3,2) &= 12u_1 + 9u_2 \\ &= \frac{12}{\sqrt{5}} + \frac{18}{\sqrt{5}} = \frac{30}{\sqrt{5}}. \end{aligned}$$

Hessian Matrix

- Differential Calculus
  - To calculate the second-order partial derivatives of multivariate functions, we need to calculate the derivatives for all combination of input variables
  - That is, for a function $f(\mathbf{x})$ with an n-dimensional input vector $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, there are n^2 second partial derivatives for any choice of *i* and *j*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)$$

  - The second partial derivatives are assembled in a matrix called the ***Hessian***

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

  - Computing and storing the Hessian matrix for functions with high-dimensional inputs can be computationally prohibitive
    - E.g., the loss function for a ResNet50 model with approximately 23 million parameters, has a Hessian of $23\ \mathrm{M} \times 23\ \mathrm{M} = 529\ \mathrm{T}$ (trillion) parameters

What Is A Hessian Matrix?

The Hessian matrix is a matrix of second order partial derivatives. Suppose we have a function f of n variables, i.e.,

$$f : R^n \rightarrow R$$

The Hessian of f is given by the following matrix on the left. The Hessian for a function of two variables is also shown below on the right.

$$f : R^n \rightarrow R$$

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

For f(x,y):

$$f : R^2 \rightarrow R$$

$$H_{(f(x,y))} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix}$$

Hessian a function of n variables (left). Hessian of f(x,y) (right)

We already know from our tutorial on gradient vectors that the gradient is a vector of first order partial derivatives. The Hessian is similarly, a matrix of second order partial derivatives formed from all pairs of variables in the domain of f.

What Is The Discriminant?

The **determinant** of the Hessian is also called the discriminant of f. For a two variable function f(x, y), it is given by:

$$\begin{vmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{vmatrix} = f_{xx} f_{yy} - f_{xy}^2$$

Discriminant of f(x, y)

- **What Do The Hessian And Discriminant Signify?**
- The Hessian and the corresponding discriminant are used to determine the local extreme points of a function. Evaluating them helps in the understanding of a function of several variables. Here are some important rules for a point (a,b) where the discriminant is D(a, b):
- The function f has a **local minimum** if $f_{xx}(a, b) > 0$ and the discriminant $D(a,b) > 0$
- The function f has a **local maximum** if $f_{xx}(a, b) < 0$ and the discriminant $D(a,b) > 0$
- The function f has a saddle point if $D(a, b) < 0$
- We cannot draw any conclusions if $D(a, b) = 0$ and need more tests

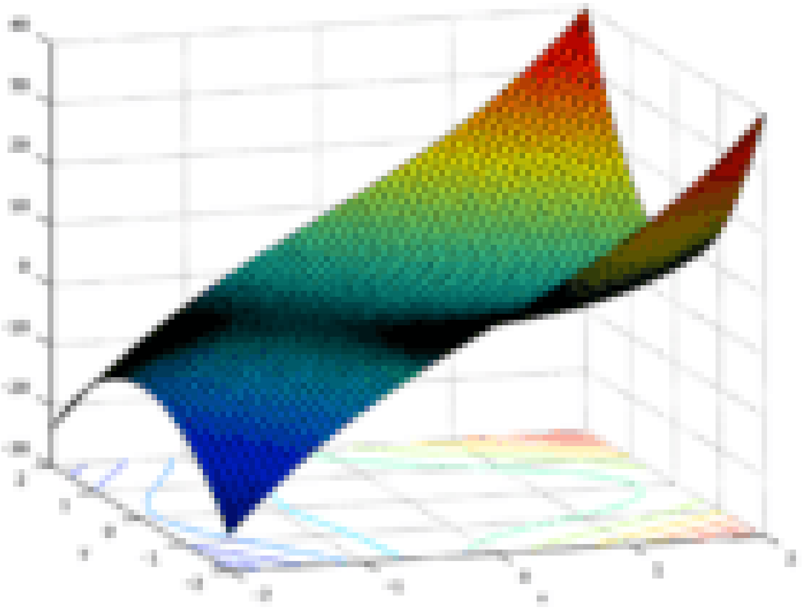
Example: g(x, y)

For the function g(x,y):

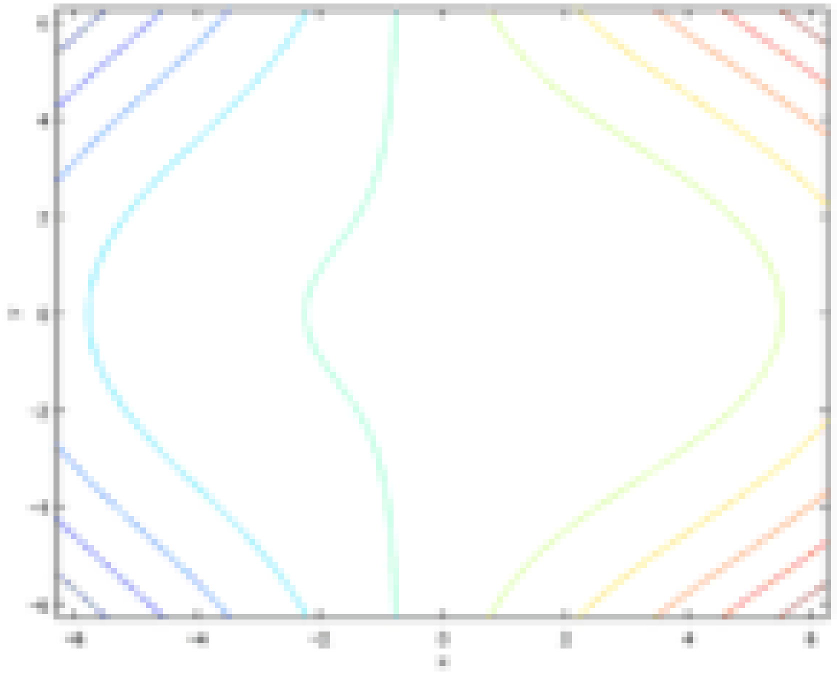
1. We cannot draw any conclusions for the point (0, 0)
2. f_xx(1, 0) = 6 > 0 and D_g(1, 0) = 60 > 0, hence (1, 0) is a local minimum
3. The point (0,1) is a saddle point as D_g(0, 1) < 0
4. f_xx(-1,0) = -6 < 0 and D_g(-1, 0) = 12 > 0, hence (-1, 0) is a local maximum

The figure below shows a graph of the function g(x, y) and its corresponding contours.

Graph of g(x,y) and contours of g(x,y)



Graph of $g(x, y) = x^3 + 2y^2 + 3xy^2$



Contours of $g(x, y) = x^3 + 2y^2 + 3xy^2$

Why Is The Hessian Matrix Important In Machine Learning?

- The Hessian matrix plays an important role in many machine learning algorithms, which involve optimizing a given function. While it may be expensive to compute, it holds some key information about the function being optimized. It can help determine the saddle points, and the local extremum of a function. It is used extensively in training neural networks and deep learning architectures.
- **Extensions**
- This section lists some ideas for extending the tutorial that you may wish to explore.
- Optimization
- Eigen values of the Hessian matrix
- Inverse of Hessian matrix and neural network training

Jacobian Matrix

- Differential Calculus
  - The concept of derivatives can be further generalized to vector-valued functions (or, vector fields) $f: \mathbb{R}^n \to \mathbb{R}^m$
  - For an *n*-dimensional input vector $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n$, the vector of functions is given as
$$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})]^T \in \mathbb{R}^m$$
  - The matrix of first-order partial derivates of the vector-valued function $\mathbf{f}(\mathbf{x})$ is an $m \times n$ matrix called a ***Jacobian***
$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix}$$
    - For example, in robotics a robot Jacobian matrix gives the partial derivatives of the translational and angular velocities of the robot end-effector with respect to the joints (i.e., axes) velocities

Jacobian Matrix

- Differential Calculus

- Jacobian matrisi, tüm birinci dereceden kısmi türevlerinin matrisidir. Bu matris kare olduğunda, yani fonksiyon, çıktısının vektör bileşenlerinin sayısı kadar girdi olarak aynı sayıda değişken aldığında, determinantına Jacobian determinantı denir.
- Jacobian matrisi, fonksiyonun çoklu boyutlara göre eğimini verir. Bir x değişkenine göre türev bize x boyutu boyunca eğimi verecektir.

Example of the Jacobian matrix

- Having seen the meaning of the Jacobian matrix, we are going to see step by step how to compute the Jacobian matrix of a multivariable function.
- Find the Jacobian matrix at the point (1,2) of the following function:

$f(x,y) = (x^4 + 3y^2x\ ,\ 5y^2 - 2xy + 1)$

First of all, we calculate all the first-order partial derivatives of the function:

$$\frac{\partial f_1}{\partial x} = 4x^3 + 3y^2$$

$$\frac{\partial f_1}{\partial y} = 6yx$$

$$\frac{\partial f_2}{\partial x} = -2y$$

$$\frac{\partial f_2}{\partial y} = 10y - 2x$$

Now we apply the formula of the Jacobian matrix. In this case, the function has two variables and two vector components, so the Jacobian matrix will be a 2×2 square matrix:

$$J_f(x,y) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} \end{pmatrix} = \begin{pmatrix} 4x^3 + 3y^2 & 6yx \\ -2y & 10y - 2x \end{pmatrix}$$

Once we have found the expression of the Jacobian matrix, we evaluate it at point (1,2):

$$J_f(1,2) = \begin{pmatrix} 4 \cdot 1^3 + 3 \cdot 2^2 & 6 \cdot 2 \cdot 1 \\ -2 \cdot 2 & 10 \cdot 2 - 2 \cdot 1 \end{pmatrix}$$

And finally, we perform the operations:

$$J_f(1,2) = \begin{pmatrix} 16 & 12 \\ -4 & 18 \end{pmatrix}$$

$$\begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix} = \begin{bmatrix} x + \sin(y) \\ y + \sin(x) \end{bmatrix}$$

$$\begin{bmatrix} \partial f_1/\partial x & \partial f_1/\partial y \\ \partial f_2/\partial x & \partial f_2/\partial y \end{bmatrix}$$

Verilen nokta etrafında uzayın ne kadar genişleyip ne kadar sıkıştığını belireyebiliriz.

$$\begin{bmatrix} 1 & \cos(y) \\ \cos(x) & 1 \end{bmatrix}$$

$$\det\left(\begin{bmatrix} 1 & \cos(y) \\ \cos(x) & 1 \end{bmatrix}\right)$$

$$x = -2$$
$$y = 1$$

$$1 \cdot 1 - \underbrace{\cos(x)}_{\approx -0,42}\,\underbrace{\cos(y)}_{\approx 0,54}$$

$$1 - (-0,227)$$

$$1,227$$

Alanlar (-2,1) noktası etrafında 1.227 değeri ile ölçeklendirilebilir. Alan 1.227 ölçek katı genişliyor. Fazla bir genişleme söz konusu değil. (0,1) noktası etrafında 0.46 değeri bulunur. Küçülme söz konusudur.

Kritik noktaların bulunması

TEOREM : $z = f(x, y)$ fonksiyonu $(x_{0,} y_0)$ noktasında ikinci mertebeden sürekli kısmi türevlere sahip bir fonksiyon ve

$$f_x(x_{0,} y_0) = f_y(x_{0,} y_0) = 0$$

olsun. Bu takdirde $\Delta = f_{xx}(x_{0,} y_0).f_{yy}(x_{0,} y_0) - f_{xy}^2(x_{0,} y_0)$ olmak üzere;

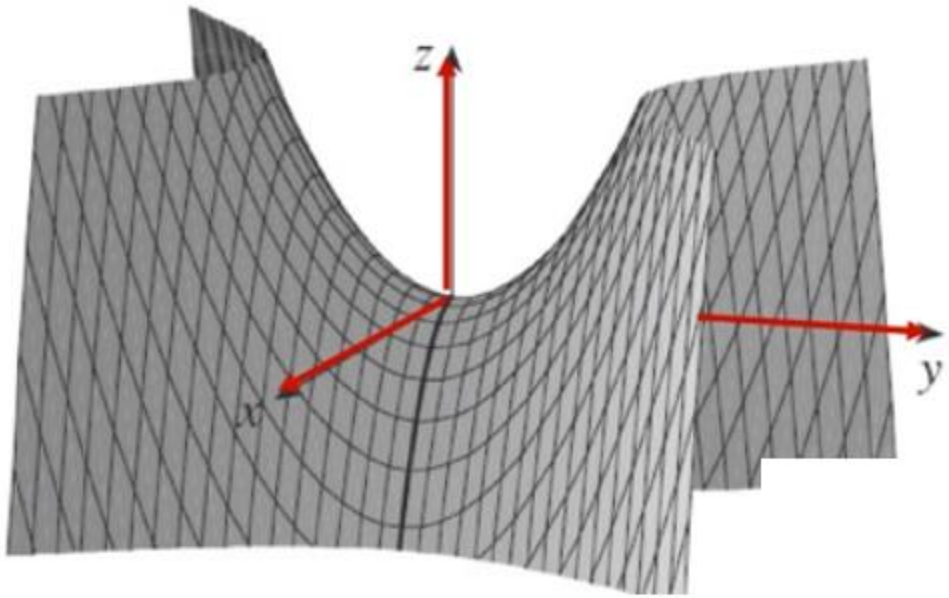
i . $\Delta > 0$ ve $f_{xx}(x_{0,} y_0) > 0$ ise $(x_{0,} y_0)$ noktası yerel minimum noktasıdır.

ii . $\Delta > 0$ ve $f_{xx}(x_{0,} y_0) < 0$ ise $(x_{0,} y_0)$ noktası yerel maksimum noktasıdır.

iii . $\Delta < 0$ ise $(x_{0,} y_0)$ noktası eyer noktasıdır.

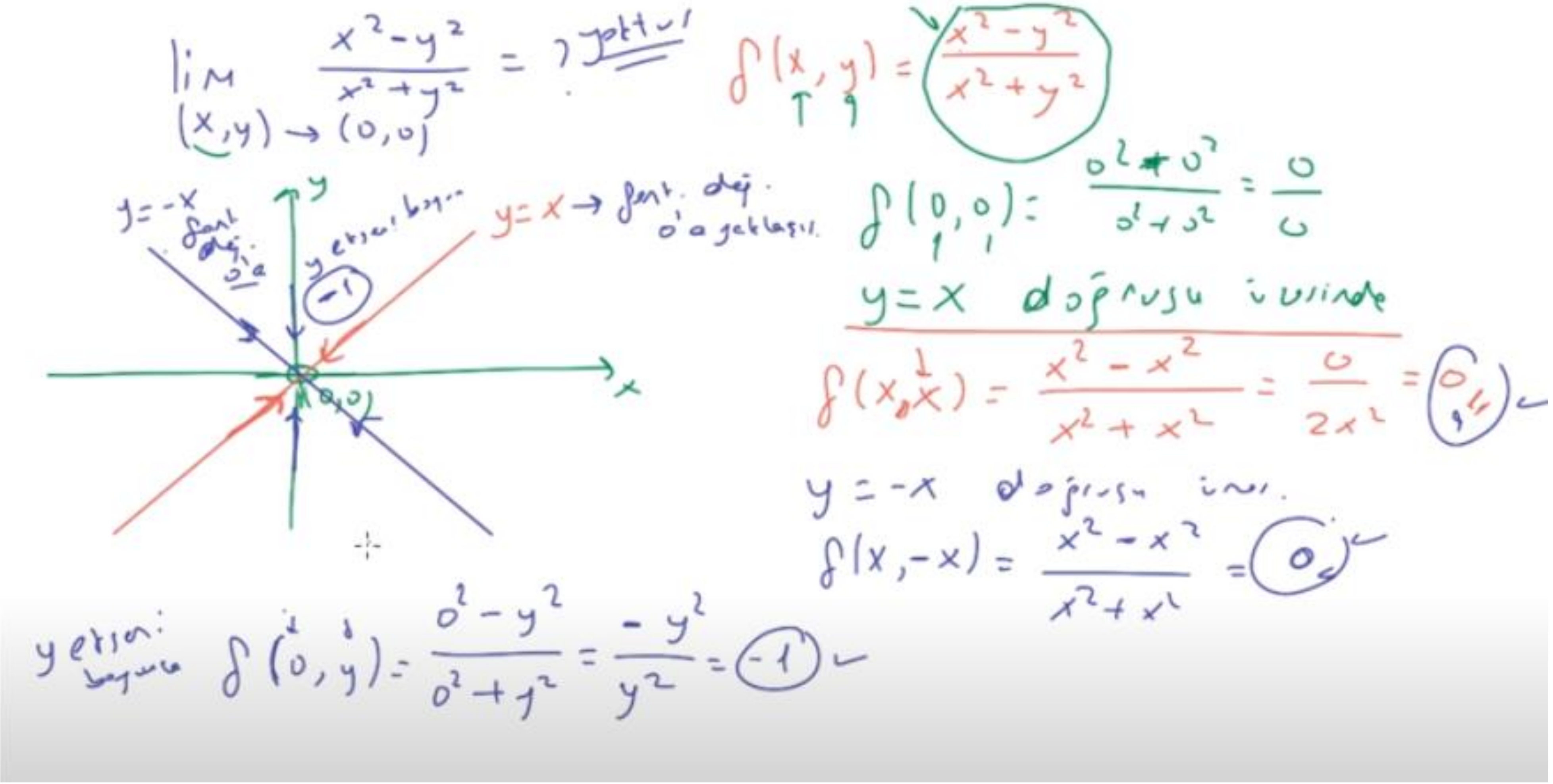
Kritik noktaların bulunması

ÖRNEK : $z = f(x,y) = y^2 - \frac{1}{4}x^2$ denklemi ile tanımlanan f fonksiyonunun hiçbir maksimumu ve minimumu olmadığını gösterelim; $z_x(x,y) = -\frac{1}{2}x$, $z_y(x,y) = 2y$ kısmi türevleri sıfır olmalıdır. Buradan $(x, y) = (0,0)$ bulunur. O halde $(0,0,0)$ noktası yüzeyimizin maksimum veya minimum olabilecek tek noktasıdır. Ancak yüzeyin şeklini incelersek bu olanaksızdır.



(0,0) noktası eyer noktasıdır.

Çok değişkenli fonksiyonlarda limit yönlü yaklaşım



Farklı yönlerden yaklaşırken aynı değer verirse limit vardır. Farklı değerler verirse limit yoktur.

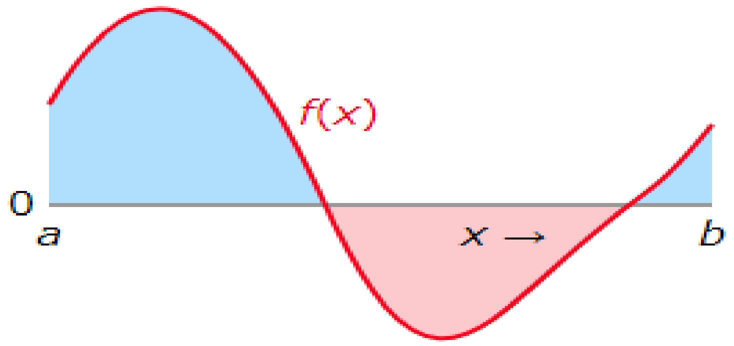
Integral Calculus

- Integral Calculus

- For a function $f(x)$ defined on the domain $[a, b]$, the definite ***integral*** of the function is denoted

$$\int_a^b f(x)dx$$

- Geometric interpretation of the integral is the area between the horizontal axis and the graph of $f(x)$ between the points *a* and *b*
  - In this figure, the integral is the sum of blue areas (where $f(x) > 0$) minus the pink area (where $f(x) < 0$)



Olasılık

63

Variance

- Probability

- ***Variance*** gives the measure of how much the values of the function $f(X)$ deviate from the expected value as we sample values of X from $P(X)$

$$\mathrm{Var}\big(f(X)\big) = \mathbb{E}[(f(X) - \mathbb{E}[f(X)])^2]$$

- When the variance is low, the values of $f(X)$ cluster near the expected value
- Variance is commonly denoted with σ^2
  - The above equation is similar to a function $f(X_i) = X_i - \mu$
  - We have $\sigma^2 = \sum_i P(X_i) \cdot (X_i - \mu)^2$
  - This is similar to the formula for calculating the variance of a sample of observations: $\sigma^2 = \frac{1}{N-1} \sum_i (X_i - \mu)^2$
- The square root of the variance is the ***standard deviation***
  - Denoted $\sigma = \sqrt{\mathrm{Var}(X)}$

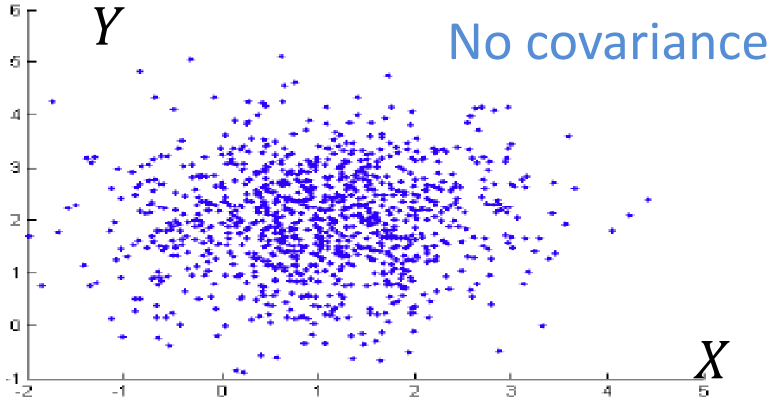
Covariance

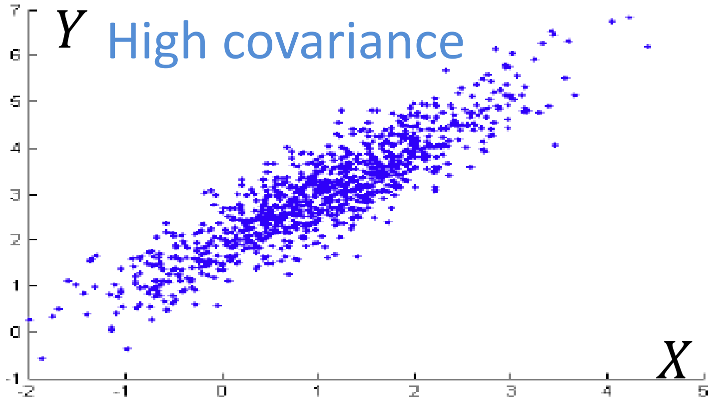
- Probability

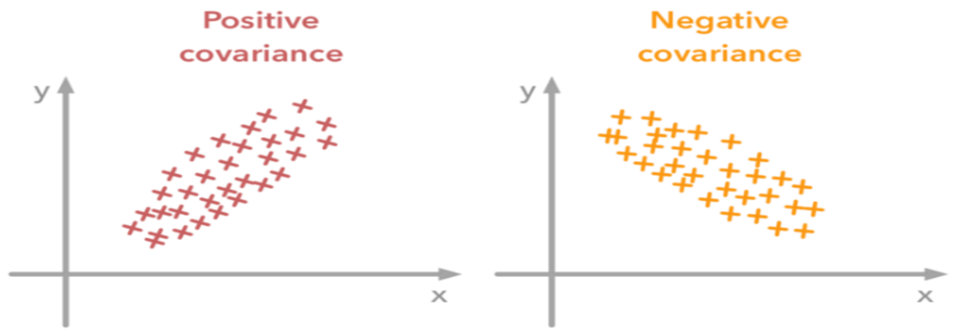
- ***Covariance*** gives the measure of how much two random variables are linearly related to each other

$$\mathrm{Cov}\big(f(X), g(Y)\big) = \mathbb{E}[(f(X) - \mathbb{E}[f(X)])(g(Y) - \mathbb{E}[g(Y)])]$$

- If $f(X_i) = X_i - \mu_X$ and $g(Y_i) = Y_i - \mu_Y$
  - Then, the covariance is: $\mathrm{Cov}(X, Y) = \sum_i P(X_i, Y_i) \cdot (X_i - \mu_X) \cdot (Y_i - \mu_Y)$
  - Compare to covariance of actual samples: $\mathrm{Cov}(X, Y) = \frac{1}{N-1} \sum_i (Y_i - \mu_X)(Y_i - \mu_Y)$
- The covariance measures the tendency for X and Y to deviate from their means in same (or opposite) directions at same time







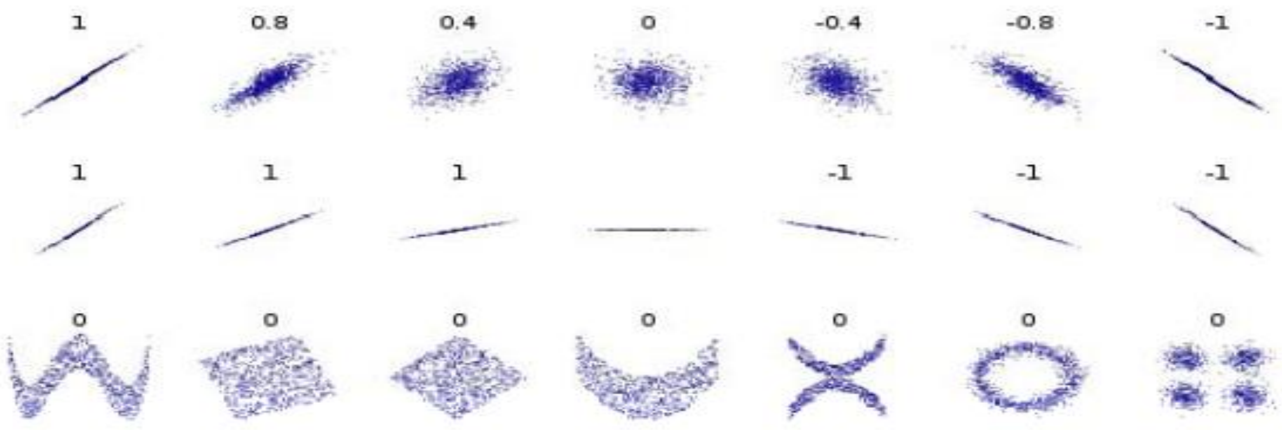
Correlation

- Probability

- ***Correlation coefficient*** is the covariance normalized by the standard deviations of the two variables

$$\mathrm{corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

  - It is also called Pearson's correlation coefficient and it is denoted $\rho(X,Y)$
  - The values are in the interval $[-1,1]$
  - It only reflects linear dependence between variables, and it does not measure non-linear dependencies between the variables



Linear dependence with noise

Linear dependence without noise

Various nonlinear dependencies

Covariance Matrix

- Probability

- ***Covariance matrix*** of a multivariate random variable $\mathbf{X}$ with states $\mathbf{x} \in \mathbb{R}^{\boldsymbol{n}}$ is an $n \times n$ matrix, such that

$$\mathrm{Cov}(\mathbf{X})_{i,j} = \mathrm{Cov}\left(\mathbf{x}_i, \mathbf{x}_j\right)$$

- I.e.,

$$\mathrm{Cov}(\mathbf{X}) = \begin{bmatrix} \mathrm{Cov}(\mathbf{x}_1, \mathbf{x}_{\mathbf{1}}) & \mathrm{Cov}(\mathbf{x}_1, \mathbf{x}_{\mathbf{2}}) & \cdots & \mathrm{Cov}(\mathbf{x}_1, \mathbf{x}_{\boldsymbol{n}}) \\ \mathrm{Cov}(\mathbf{x}_2, \mathbf{x}_{\mathbf{1}}) & \ddots & & \mathrm{Cov}(\mathbf{x}_2, \mathbf{x}_{\boldsymbol{n}}) \\ \vdots & & & \vdots \\ \mathrm{Cov}(\mathbf{x}_n, \mathbf{x}_{\mathbf{1}}) & \mathrm{Cov}(\mathbf{x}_n, \mathbf{x}_{\mathbf{2}}) & \cdots & \mathrm{Cov}(\mathbf{x}_{\boldsymbol{n}}, \mathbf{x}_{\boldsymbol{n}}) \end{bmatrix}$$

- The diagonal elements of the covariance matrix are the variances of the elements of the vector

$$\mathrm{Cov}(\mathbf{x}_i, \mathbf{x}_i) = \mathrm{Var}(\mathbf{x}_i)$$

- Also note that the covariance matrix is symmetric, since $\mathrm{Cov}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \mathrm{Cov}\left(\mathbf{x}_j, \mathbf{x}_i\right)$

References

1. A. Zhang, Z. C. Lipton, M. Li, A. J. Smola, *Dive into Deep Learning*, https://d2l.ai, 2020.
2. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2017.
3. M. P. Deisenroth, A. A. Faisal, C. S. Ong, *Mathematics for Machine Learning*, Cambridge University Press, 2020.
4. Jeff Howbert — Machine Learning Math Essentials presentation
5. Brian Keng – Manifolds: A Gentle Introduction blog
6. Martin J. Osborne – Mathematical Methods for Economic Theory (link)