



Penulis lahir di Pati, 17 Juli 1969. Pendidikan dasarnya diselesaikan di Pati (SD,SMPN 2 PATI, SMAN 1 Pati). Gelar Sarjana diperoleh dari Prodi Matematika UGM, demikian juga gelar Magisternya. Adapun gelar Ph.D diperoleh dari UMT (Universitas Malaysia Terengganu)

Penulis meniti karier sebagai dosen sejak tahun 2000 di Prodi Matematika UAD. Pernah menjabat Kaprodi dan Wakil Dekan, kini ia menjadi Kepala Laboratorium Matematika serta pembina kegiatan mahasiswa di Pusat Studi Data Sains.

Penulis 2 lahir di Raja Basa Lama, 10 Desember 1999. Pendidikan dasarnya diselesaikan di SDN 10 Girimulya Bengkulu Utara, SMP di SMP N 1 Giri Mulya Bengkulu Utara, SMA N 1 Bengkulu Utara. Gelar sarjana (S.Si) diperoleh dari Prodi Matematika UAD, demikian juga gelar masternya.

Penulis meniti karier sebagai Mahasiswa sejak tahun 2019 di Prodi Matematika UAD. Penulis juga pernah menjadi asisten praktikum mata kuliah Algoritma Pemrograman dan sebagai Instructor di Pusat Studi Data Sains (PSDS) UAD dari 2022 Hingga 2023.



Penulis 3 lahir di Sleman, 6 Juli 2000. Pendidikan dasarnya diselesaikan di Sleman (SD,SMPN 1 Berbah, SMAN 1 Prambanan). Gelar Sarjana diperoleh dari Prodi Matematika Fakultas Sains dan Teknologi Terapan Universitas Ahmad Dahlan.

Penulis meniti karier sebagai Mahasiswa sejak tahun 2019 di Prodi Matematika UAD. Penulis juga pernah menjadi asisten praktikum mata kuliah Metode Optimasi, Metode Statistika, dan Paket Komputasi. Penulis juga pernah menjadi Instruktur di Pusat Studi Data Sains.



Jalan Pasir Putih No. 16 Kelurahan
Mekarjaya, Kecamatan Rancasari
Kota Bandung

ISBN: 978-623-448-791-6



Prof. H. Sugiyarto, M.Si., Ph.D.

Pengantar Metode Optimasi

Pengantar Metode Optimasi

Prof. H. Sugiyarto, M.Si., Ph.D.

Anggi Setyawan, S.Si.

Galang Suryaputra, S.Si.



Pengantar Metode Optimasi

Prof. H. Sugiyarto, M.Si., Ph.D.

Anggi Setyawan, S.Si.

Galang Suryaputra, S.Si.

PENGANTAR METODE OPTIMASI

Penyusun : Prof. H. Sugiyarto, M.Si., Ph.D.
Layout dan Desain Cover : Anggi Setyawan, S.Si.
Galang Suryaputra, S.Si.

Cetakan : Februari 2024
Ukuran : 21,59 x 27,9 cm
Hal. vi + 197
ISBN : 978-623-448-791-6

Diterbitkan Oleh

Perkumpulan Rumah Cemerlang Indonesia

ANGGOTA IKAPI JAWA BARAT

Pondok Karisma Residence Jalan Rafflesia VI D.151

Panglayungan, Cipedes Tasikmalaya – 085223186009

Website : www.rcipress.rcipublisher.org

E-mail : rumahcemerlangindonesia@gmail.com

Hak cipta dilindungi undang-undang
Dilarang memperbanyak buku ini dalam bentuk dan dengan
cara apapun tanpa izin tertulis dari penulis dan penerbit
Undang-undang No.19 Tahun 2002 Tentang
Hak Cipta Pasal 72

Kata Pengantar

Bismillahirrahmanirrahiim

Puji syukur Alhamdulillah penyusun panjatkan kepada Allah SWT atas berkat rahmat dan hidayah-Nya sehingga penyusun dapat menyelesaikan Buku dengan judul “Pengantar Metode Optimasi”. Buku ini digunakan sebagai salah satu pegangan untuk mata kuliah Pemodelan Matematika.

Penyusun menyadari bahwa buku ini masih banyak kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi kesempurnaan buku ini. Akhir kata, penyusun berharap semoga buku ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan khususnya pada bidang Matematika

Yogyakarta, 2023

Penyusun

Daftar isi

Contents

PENGANTAR METODE OPTIMASI.....	1
Kata Pengantar	iv
Daftar isi	v
BAB 1	8
Metode Pembuktian dan Beberapa Notasi	8
1.1 Metode Pembuktian.....	8
1.2 Notasi	11
BAB 2 Ruang Vektor dan Matriks.....	14
2.1 Ruang Vektor	14
2.2 Tingkatan Matriks	19
2.3 Persamaan Linear	24
2.4 Perkalian Titik dan Aturannya	27
BAB 3 Transformasi	36
3.1 Transformasi Linear	36
3.2 Nilai Eigen dan Vektor Eigen.....	37
3.3 Ortogonal Proyeksi.....	40
3.4 Bentuk Kuadrat	41
3.5 Norma Matriks	46
BAB 4 Konsep Geometri	58
4.1 Ruang Garis.....	58
4.2 Hyperplanes dan Varietas linear	58
4.3 Himpunan Cembung.....	61
4.4 Lingkungan.....	63
4.5 Polytopes dan Polyhedra	65
BAB 5 Elemen Kalkulus.....	69
5.1 Baris dan Limit.....	69
5.2 Differensibilitas	78
5.3 Matriks Derivatif	79

5.4	Aturan Diferensiasi.....	82
5.5	Tingkatan Himpunan dan Gradien	84
5.6	Deret Taylor	88
BAB 6 Konsep Matematika dalam Optimasi.....		99
6.1	Konsep Dasar Optimasi.....	99
6.2	Vektor Gradien $f\mathbf{x}$	100
6.3	Matriks Hessian dari $f\mathbf{x}$	100
6.4	Kondisi Peminimum Lokal	104
BAB 7 Optimasi Tanpa Kendala		116
7.1	Fungsi Univariat	116
7.2	Fungsi Multivariat	120
BAB 8 Metode Pencarian Satu Dimensi.....		123
8.1	Pengantar Plot Grafik <i>Non Linear</i>	123
8.2	Metode <i>Golden Section Search</i>	124
8.3	Barisan Fibonacci	126
8.4	Metode Biseksi/Metode Bagi Dua	128
8.5	Metode Newton	129
BAB 9 Algoritma Machine Learning		133
9.1	Data Mining	133
9.1.1	Hierarchy Clustering	133
9.1.2	K-Means Clustering	134
9.1.3	Jarak metrik	135
9.2	Data Mining untuk Big Data	136
9.3	Artificial Neural Networks (Jaringan Syaraf Tiruan).....	140
9.3.1	Model Neuron	141
9.3.2	Neural Networks	142
9.3.3	Algoritma Back Propagation	145
9.3.4	Fungsi Kerugian ANN	146
9.3.5	Turunan Gradien Stokastik	147
9.3.6	Boltzmann Terbatas	148
9.4	<i>Support Vector Machines (SVM)</i>	149

BAB 10	156
Masalah Optimasi dengan Kendala Berupa Persamaan.....	156
10.1 Metode Pengali Lagrange.....	156
10.2 Masalah Fungsi Penalty.....	167
BAB 11	176
11.1 Masalah Optimasi dengan Kendala Berupa Pertidaksamaan Kondisi Karush-Kuhn-Tucker	176
Daftar Pustaka.....	182
Simulasi Menggunakan Python	187

BAB 1

Metode Pembuktian dan Beberapa

Notasi

1.1 Metode Pembuktian

Perhatikan dua pernyataan “ A ” dan “ B ”, yang mana keduanya dapat bernilai benar dan salah. Sebagai contoh, misalkan “ A ” adalah pernyataan “John adalah mahasiswa teknik”, dan “ B ” adalah pernyataan “John mengambil mata kuliah optimisasi.” Kedua pernyataan dapat digabungkan menjadi pernyataan lain seperti, “ A ” dan “ B ” atau “ A ” atau “ B ”. Pada contoh, “ A ” dan “ B ” berarti “John adalah mahasiswa teknik, dan dia mengambil mata kuliah optimisasi.” Kita juga dapat membentuk pernyataan seperti “*bukan A*”, “*bukan B*”, “*bukan (A dan B)*”, dan seterusnya. Untuk contoh, “*bukan A*” berarti “John bukan mahasiswa teknik.” Kebenaran atau kesalahan dari pernyataan gabungan tergantung pada kebenaran atau kesalahan pernyataan asli, “ A dan B .” Hubungan ini dinyatakan dengan tabel kebenaran; lihat **Tabel 1.1 dan 1.2**.

Dari **Tabel 1.1 dan 1.2**, mudah untuk dilihat bahwa pernyataan “*bukan (A dan B)*” adalah setara dengan “(*bukan A*) atau (*bukan B*)” (lihat Latihan 1.3). Ini disebut hukum *DeMorgan*. Dalam membuktikan pernyataan, akan lebih mudah untuk mengungkapkan pernyataan gabungan dengan kondisional, seperti “ A menyiratkan B ,” yang kami nyatakan sebagai “ $A \Rightarrow B$.” Kondisional “ $A \Rightarrow B$ ” hanyalah pernyataan gabungan “(*bukan A*) atau B ,” dan sering juga dibaca “ A hanya jika B ,” atau “jika A maka B ”, atau “ A cukup untuk B ”, atau “ B diperlukan untuk A ”.

Kita dapat menggabungkan dua pernyataan bersyarat untuk membentuk pernyataan bikondisional dari bentuk " $A \Leftrightarrow B$," yang berarti " $(A \Rightarrow B)$ dan $(B \Rightarrow A)$ ". Pernyataan " $A \Leftrightarrow B$ " berbunyi "A jika dan hanya jika B", atau "A setara dengan B," atau "A diperlukan dan cukup untuk B." Tabel kebenaran untuk pernyataan bersyarat dan bikondisional diberikan pada **Tabel 1.3**.

Tabel 1.1 Tabel Kebenaran untuk " A dan B " dan " A atau B "

A	B	A dan B	A atau B
F	F	F	F
F	T	F	T
T	F	F	T
T	T	T	T

Tabel 1.2 Tabel Kebenaran untuk " $\text{bukan } A$ "

A	Bukan A
F	T
T	F

Tabel 1.3 Tabel Kebenaran untuk Kondisional dan Bikondisional

A	B	$A \Rightarrow B$	A	$A \Leftrightarrow B$
F	F	T	T	T
F	T	T	F	F
T	F	F	T	F
T	T	T	T	T

Sangat mudah untuk memverifikasi, dengan menggunakan tabel kebenaran, bahwa pernyataan " $A \Rightarrow B$ " setara dengan pernyataan " $(\text{bukan } B) \Rightarrow (\text{bukan } A)$ ". Yang terakhir disebut kontraposisif dari yang pertama.

Jika kita mengambil kontraposisif dengan Hukum DeMorgan, kita memperoleh pernyataan bahwa "*bukan(A atau B)*" sama dengan "*(bukan A) dan (bukan B)*".

Sebagian besar pernyataan yang diselesaikan memiliki bentuk " $A \Rightarrow B$ ". Untuk membuktikan pernyataan seperti itu, kita dapat menggunakan salah satu dari tiga teknik berbeda berikut:

1. Metode langsung
2. Bukti dengan kontraposisi
3. Buktikan dengan kontradiksi atau *reductio ad absurdum*.

Dalam kasus metode langsung, kita mulai dengan " A ", kemudian menyimpulkan rantai berbagai konsekuensi untuk diakhiri dengan " B ".

Metode yang berguna untuk membuktikan pernyataan adalah pembuktian dengan kontraposisi, berdasarkan ekivalensi pernyataan " $A \Rightarrow B$ " dan " $(bukan B) \Rightarrow (bukan A)$ ". Kita mulai dengan "*bukan B*", kemudian simpulkan berbagai konsekuensi untuk diakhiri dengan "*bukan A*" sebagai kesimpulan. Metode pembuktian lain yang kita gunakan adalah pembuktian dengan kontradiksi, berdasarkan ekivalensi pernyataan " $A \Rightarrow B$ " dan "*bukan (A dan (bukan B))*." Di sini kita mulai dengan "*A dan (bukan B)*" dan turunkan kontradiksi.

Metode pembuktian lain yang kita gunakan adalah pembuktian dengan kontradiksi, berdasarkan ekivalensi pernyataan " $A \Rightarrow B$ " dan "*bukan (A dan (bukan B))*." Di sini kita mulai dengan "*A dan (bukan B)*" dan turunkan kontradiksi. Terkadang, kami menggunakan prinsip induksi untuk membuktikan pernyataan. Prinsip ini dapat dinyatakan sebagai berikut. Asumsikan bahwa properti tertentu dari bilangan bulat positif memenuhi kondisi berikut:

- Bilangan 1 memiliki sifat ini
- Jika bilangan n memiliki sifat ini, maka bilangan $n + 1$ juga memilikinya.

Prinsip induksi menyatakan bahwa di bawah asumsi ini setiap bilangan bulat positif memiliki sifat.

Prinsip induksi mudah dipahami dengan menggunakan argumen intuitif berikut. Jika angka 1 memiliki sifat yang diberikan maka kondisi kedua menyiratkan bahwa angka 2 memiliki

properti. Tetapi, sekali lagi, kondisi kedua menyiratkan bahwa angka 3 memiliki sifat ini, dan seterusnya. Prinsip induksi adalah pernyataan formal dari penalaran intuitif ini.

1.2 Notasi

Jika X adalah himpunan, maka dapat dituliskan $x \in X$ yang berarti x adalah anggota X . Jika x bukan anggota himpunan X , ditulis $x \notin X$. Kami juga menggunakan "notasi kurung kurawal" untuk himpunan, dengan menuliskan beberapa elemen pertama dari suatu himpunan diikuti oleh tiga titik. Misalnya, $\{x_1, x_2, x_3, \dots\}$ adalah himpunan yang memuat elemen x_1, x_2, x_3 , dan seterusnya. Atau, kita dapat secara eksplisit menunjukkan hukum pembentukan. Misalnya, $\{x : x \in \mathbb{R}, x > 5\}$ berbunyi "kumpulan semua x sedemikian sehingga x adalah bilangan real dan x lebih besar dari 5". Tanda titik dua setelah x berbunyi "sedemikian sehingga". Notasi alternatif untuk himpunan yang sama adalah $\{x \in \mathbb{R} : x > 5\}$.

Jika X dan Y himpunan, maka dapat dituliskan $X \subset Y$ yang berarti bahwa setiap elemen X juga sebuah elemen dari Y . Dalam hal ini, kita katakan bahwa X adalah himpunan bagian dari Y . Jika X dan Y adalah himpunan, maka kita tunjukkan dengan $X \setminus Y$ (" X dikurangi Y ") himpunan semua titik di X yang tidak ada di Y . Perhatikan bahwa $X \setminus Y$ adalah himpunan bagian dari X . Notasi $f : X \rightarrow Y$ berarti "f adalah fungsi dari himpunan X ke himpunan Y " Simbol $:=$ menunjukkan tugas aritmatika. Jadi, pernyataan dalam bentuk $x := y$ berarti " x menjadi y ." Simbol \triangleq berarti "sama dengan menurut definisi."

Latihan Soal:

1.1 Buatlah tabel kebenaran untuk pernyataan " $(\text{bukan } B) \Rightarrow (\text{bukan } A)$ ", dan gunakan untuk menunjukkan bahwa pernyataan ini setara dengan pernyataan " $A \Rightarrow B$."

Penyelesaian:

A	B	Bukan A	Bukan B	AB	(bukan B) (bukan A)
F	F	T	T	T	T
F	T	T	F	T	T
T	F	F	T	F	F
T	T	F	F	T	T

1.2 Buatlah tabel kebenaran untuk pernyataan “bukan (A dan (bukan B))”, dan gunakan untuk tunjukkan bahwa pernyataan ini ekuivalen dengan pernyataan " $A \Rightarrow B$."

Penyelesaian:

A	B	Bukan A	Bukan B	AB	Bukan(A dan (bukan B))
F	F	T	T	T	T
F	T	T	F	T	T
T	F	F	T	F	F
T	T	F	F	T	T

1.3 Buktikan Hukum *De Morgan* dengan membuat tabel kebenaran yang sesuai.

Penyelesaian:

A	B	Bukan (A dan B)	Bukan A	Bukan B	(bukan A) atau (bukan B)
F	F	T	T	T	T
F	T	T	T	F	T
T	F	T	F	T	T
T	T	F	F	F	F

1.4 Buktikan bahwa untuk setiap pernyataan A dan B, kita memiliki " $A \Leftrightarrow (A \text{ dan } B) \text{ atau } (A \text{ dan (bukan B))$." Ini berguna karena memungkinkan kita untuk membuktikan pernyataan A dengan membuktikan keduanya kasus terpisah "(A dan B)," dan "(A dan (bukan B))." Misalnya, untuk membuktikan bahwa $|x| \geq x$ untuk setiap $x \in \mathbb{R}$, kita buktikan secara terpisah kasus " $|x| > x$ dan $x > 0$," dan " $|x| > x$ dan $x < 0$." Membuktikan kedua kasus ternyata lebih mudah daripada membuktikan langsung pernyataan $|x| \geq x$ (lihat Bagian 2.4 dan Latihan 2.4)

Penyelesaian:

A	B	A dan B	A dan (bukan B)	(A dan B) atau (A dan (bukan B))
F	F	F	F	F
F	T	F	F	F

T	F	F	T	T
T	T	T	F	T

1.5 (Latihan ini diambil dari [17, hlm. 80-81]) Misalkan Anda diberikan empat kartu, disusun berjajar. Setiap kartu memiliki huruf di satu sisi dan angka di sisi lain. Di sisi kartu yang terlihat dicetak simbol: Tentukan kartu mana yang harus Anda balikkan untuk memutuskan apakah aturan berikut ini benar atau salah: "Jika ada vokal di satu sisi kartu, maka ada angka genap sisi lain."

Penyelesaian:

Kartu yang harus balikkan adalah 3 dan A. Kartu yang tersisa tidak relevan dengan kebenaran atau kepalsuan aturan. Kartu dengan S tidak relevan karena S bukan vokal. Kartu dengan 8 tidak relevan karena aturan tidak mengatakan bahwa jika sebuah kartu memiliki nomor genap di satu sisi, maka ia memiliki vokal di sisi lain. Membalik kartu A secara langsung memverifikasi aturan, sementara membalik kartu 3 memverifikasi kontraposisi.

BAB 2

RUANG VEKTOR DAN MATRIKS

2.1 Ruang Vektor

Kita definisikan bahwa n -vektor kolom menjadi n -array dinotasikan oleh :

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Bilangan a_i disebut sebagai komponen ke- i dari vektor a . Himpunan bilangan real dinotasikan dengan \mathbb{R} , dan himpunan dari n -vektor kolom dengan komponennya adalah bilangan real dinotasikan dengan \mathbb{R}^n . Kita sebut \mathbb{R}^n sebagai ruang vektor berdimensi n . Kita biasa menyatakan elemen \mathbb{R}^n dengan menggunakan huruf kecil yang dicetak tebal, misalnya, \mathbf{x} . Komponen dari $\mathbf{x} \in \mathbb{R}^n$ dinotasikan dengan x_1, \dots, x_n .

Kita definisikan n -vektor baris sebagai berikut :

$$[a_1, a_2, \dots, a_n]$$

Transpos dari vektor kolom yang diberikan a adalah vektor panah dengan elemen yang sesuai, dilambangkan dengan a^T . Misalnya, jika

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Maka :

$$\mathbf{a}^T = [a_1, a_2, \dots, a_n]$$

Sedemikian sehingga secara ekuivalen, kita dapat menuliskannya $\mathbf{a} = [a_1, a_2, \dots, a_n]^T$. Dua vektor misalnya $\mathbf{a} = [a_1, a_2, \dots, a_n]^T$ dan $\mathbf{b} = [b_1, b_2, \dots, b_n]^T$ adalah ekuivalen jika $a_i = b_i$, dengan $i = 1, 2, \dots, n$. Jumlah dari vektor \mathbf{a} dan vektor \mathbf{b} , dinotasikan dengan $\mathbf{a} + \mathbf{b}$, sedemikian sehingga vektornya dapat diuraikan menjadi

$$\mathbf{a} + \mathbf{b} = [a_1 + b_1, a_2 + b_2, \dots, a_n + b_n]^T$$

Operasi penjumlahan vektor memiliki sifat-sifat sebagai berikut :

1) Komutatif :

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$$

2) Asosiatif :

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}).$$

3) Vektor Nol

$$\mathbf{0} = [0, 0, \dots, 0]^T$$

Sedemikian sehingga

$$\mathbf{a} + \mathbf{0} = \mathbf{0} + \mathbf{a} = \mathbf{a}$$

Vektor

$$[a_1 - b_1, a_2 - b_2, \dots, a_n - b_n]^T$$

Disebut selisih antara \mathbf{a} dan \mathbf{b} , dan dinotasikan dengan $\mathbf{a} - \mathbf{b}$. Vektor $\mathbf{0} - \mathbf{b}$ dapat dinotasikan dengan $-\mathbf{b}$. Perhatikan bahwa

$$\mathbf{b} + (\mathbf{a} - \mathbf{b}) = \mathbf{a} - (-\mathbf{b}) = \mathbf{b} \quad -(\mathbf{a} - \mathbf{b}) = \mathbf{b} - \mathbf{a}$$

Vektor $\mathbf{b} - \mathbf{a}$ adalah solusi tunggal dari persamaan vector

$$\mathbf{a} + \mathbf{x} = \mathbf{b}$$

Sesungguhnya, misalkan $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ adalah solusi untuk $\mathbf{a} + \mathbf{x} = \mathbf{b}$. Maka

$$\mathbf{a}_1 + \mathbf{x}_1 = \mathbf{b}_1 \quad \mathbf{a}_2 + \mathbf{x}_2 = \mathbf{b}_2 \quad \vdots \quad \mathbf{a}_n + \mathbf{x}_n = \mathbf{b}_n$$

Dengan demikian

$$\mathbf{x} = \mathbf{b} - \mathbf{a}$$

Kita definisikan operasi perkalian dari vektor $\mathbf{a} \in \mathbb{R}^n$ terhadap skalar $\alpha \in \mathbb{R}$ sebagai berikut

$$\alpha \mathbf{a} = [\alpha a_1, \alpha a_2, \dots, \alpha a_n]^T$$

Operasi tersebut mengikuti sifat-sifat berikut :

1) Distributif : untuk setiap skalar α dan β ,

$$\alpha(\mathbf{a} + \mathbf{b}) = \alpha\mathbf{a} + \alpha\mathbf{b}$$

$$(\alpha + \beta)\mathbf{a} = \alpha\mathbf{a} + \beta\mathbf{a}$$

2) Asosiatif:

$$\alpha(\beta\mathbf{a}) = (\alpha\beta)\mathbf{a}$$

3) Skalar 1 memenuhi:

$$1\mathbf{a} = \mathbf{a}$$

4) Setiap skalar α memenuhi

$$\alpha\mathbf{0} = \mathbf{0}$$

5) Skalar 0 memenuhi :

$$0\mathbf{a} = \mathbf{0}$$

6) Skalar -1 memenuhi:

$$(-1)\mathbf{a} = -\mathbf{a}$$

Ingat bahwa $\alpha\mathbf{a} = \mathbf{0}$ jika dan hanya jika $\alpha = \mathbf{0}$ atau $\mathbf{a} = \mathbf{0}$. Untuk itu, perhatikan bahwa $\alpha\mathbf{a} = \mathbf{0}$ adalah ekuivalen dengan $\alpha a_1 = \alpha a_2 = \dots = \alpha a_n = 0$. Jika $\alpha = 0$ atau $\mathbf{a} = \mathbf{0}$, maka $\alpha\mathbf{a} = \mathbf{0}$. Jika $\mathbf{a} \neq \mathbf{0}$, maka setidaknya satu dari komponen tersebut $a_k \neq 0$. Untuk komponen ini, $\alpha a_k = 0$, dan karenanya kita harus memiliki $\alpha = 0$. Argumen serupa juga dapat diterapkan pada kasus saat $\alpha \neq 0$.

Sebuah himpunan vektor $\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$ adalah dikatakan bebas linear jika persamaan

$$\alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \dots + \alpha_k\mathbf{a}_k = \mathbf{0}$$

Menunjukkan bahwa seluruh koefisien α_i , dengan $i = 1, \dots, k$ adalah sama dengan nol. Himpunan vektor $\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$ bergantung linier jika tidak bebas linier.

Perhatikan bahwa himpunan yang tersusun dari vektor tunggal $\mathbf{0}$ adalah tidak bebas linear, jika $\alpha \neq 0$ maka $\alpha\mathbf{0} = \mathbf{0}$. Sesungguhnya, setiap himpunan dari vektor yang memuat $\mathbf{0}$ adalah tidak bebas linear. Sebuah himpunan yang tersusun dari vektor tunggal tak nol $\mathbf{a} \neq \mathbf{0}$ adalah bebas linear saat $\alpha\mathbf{a} = \mathbf{0}$ yang berarti $\alpha = \mathbf{0}$.

Sebuah vektor \mathbf{a} dapat dikatakan menjadi kombinasi linear dari vektor $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$ jika terdapat skalar $\alpha_1, \dots, \alpha_k$ sedemikian sehingga

$$\mathbf{a} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \cdots + \alpha_k \mathbf{a}_k$$

Proposisi 2.1 Sebuah himpunan dari vektor $\{a_1, a_2, \dots, a_k\}$ adalah tidak bebas linear jika dan hanya jika salah satu vektor dari himpunan adalah kombinasi linear.

Bukti. Jika $\{a_1, a_2, \dots, a_k\}$ adalah tidak bebas linear maka

$$\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \cdots + \alpha_k \mathbf{a}_k = \mathbf{0}$$

Dimana setidaknya terdapat satu dari skalar $\alpha_i \neq 0$, sehingga

$$a_i = -\frac{\alpha_1}{\alpha_i} a_1 - \frac{\alpha_2}{\alpha_i} a_2 - \cdots - \frac{\alpha_k}{\alpha_i} a_k$$

Misalkan bahwa

$$a_1 = \alpha_2 a_2 + \alpha_3 a_3 + \cdots + \alpha_k a_k$$

Maka

$$(-1)a_1 + \alpha_2 a_2 + \cdots + \alpha_k a_k = 0$$

Oleh karena skalar pertama adalah bukan nol, himpunan vektor $\{a_1, a_2, \dots, a_k\}$ adalah tidak bebas linear. Argumen yang sama juga berlaku jika a_i , dengan $i = 2, \dots, k$, adalah kombinasi linear.

Himpunan bagian v dari \mathbb{R}^n dikatakan sub-ruang dari \mathbb{R}^n jika v adalah tertutup pada operasi penjumlahan vektor dan perkalian skalar. Maka dari itu, jika a dan b adalah vektor pada v , maka vektor $a + b$ dan αa juga pada v untuk setiap skalar α .

Setiap sub-ruang memuat nol vektor 0 , jika a adalah elemen dari sub-ruang, maka $(-1)a = -a$. Oleh karenanya, $a - a = 0$ juga bagian dari sub-ruang.

Misalkan a_1, a_2, \dots, a_k menjadi vektor yang berubah-ubah di \mathbb{R}^n . Himpunan dari seluruh kombinasi linearnya disebut span dari a_1, a_2, \dots, a_k dan dinotasikan sebagai berikut :

$$\text{span}[a_1, a_2, \dots, a_k] = \left\{ \sum_{i=1}^k \alpha_i a_i : \alpha_1, \dots, \alpha_k \in \mathbb{R} \right\}$$

Diberikan sebuah vektor a , sub-ruang $\text{span}[a]$ memuat vektor αa , dimana α adalah bilangan real yang berubah-ubah ($\alpha \in \mathbb{R}$). Perhatikan juga bahwa jika a adalah kombinasi linear dari a_1, a_2, \dots, a_k maka

$$\text{span}[a_1, a_2, \dots, a_k, a] = \text{span}[a_1, a_2, \dots, a_k]$$

Span dari setiap himpunan vektor adalah sub-ruang.

Diberikan sub-ruang v , setiap himpunan vektor bebas linear $\{a_1, a_2, \dots, a_k\} \subset v$ sedemikian sehingga $v = \text{span}[a_1, a_2, \dots, a_k]$ disebut sebagai basis dari sub-ruang v . Semua basis dari sub-ruang v memuat vektor bilangan yang sama. Bilangan ini disebut dimensi dari v , dinotasikan dengan $\dim v$.

Proposisi 2.2 Jika $\{a_1, a_2, \dots, a_k\}$ adalah sebuah basis dari v , maka setiap vektor a dari v direpresentasikan tunggal sebagai

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_k a_k$$

dimana $\alpha_i \in \mathbb{R}$, dengan $i = 1, 2, \dots, k$.

Bukti. Untuk membuktikan representasi tunggal dari a berkenaan dengan basis vektor, asumsikan bahwa

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_k a_k$$

Dan

$$a = \beta_1 a_1 + \beta_2 a_2 + \dots + \beta_k a_k$$

Sekarang kita tunjukkan bahwa $\alpha_i = \beta_i$, dengan $i = 1, \dots, k$. Kita memiliki

$$\alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_k a_k = \beta_1 a_1 + \beta_2 a_2 + \dots + \beta_k a_k$$

Atau

$$(\alpha_1 - \beta_1)a_1 + (\alpha_2 - \beta_2)a_2 + \dots + (\alpha_k - \beta_k)a_k = 0$$

Karena himpunan $\{a_i : i = 1, 2, \dots, k\}$ adalah bebas linear, $\alpha_1 - \beta_1 = \alpha_2 - \beta_2 = \dots = \alpha_k - \beta_k = 0$, maka, $\alpha_i = \beta_i$, dengan $i = 1, \dots, k$.

Misalkan kita memiliki sebuah basis $\{a_1, a_2, \dots, a_k\}$ dari v dan sebuah vektor $a \in v$ maka

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_k a_k$$

Koefisien α_i , dengan $i = 1, \dots, k$, disebut koordinat dari a dengan memperhatikan basis $\{a_1, a_2, \dots, a_k\}$.

Basis natural untuk \mathbb{R}^n adalah himpunan vector

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \dots, e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

Alasan untuk menyebut vektor-vektor tersebut sebagai basis natural adalah

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 e_1 + x_2 e_2 + \dots + x_n e_n$$

Kita juga dapat mendefinisikan ruang vektor kompleks. Untuk itu, misalkan \mathbb{C} menunjukkan himpunan bilangan kompleks, dan \mathbb{C}^n adalah himpunan n-vektor kolom dengan komponen bilangan kompleks. Hal ini dapat di verifikasi sendiri, himpunan \mathbb{C}^n memiliki sifat yang mirip dengan \mathbb{R}^n , dimana skalar dapat diambil dari bilangan kompleks

2.2 Tingkatan Matriks

Matriks adalah susunan bilangan berbentuk persegi ataupun persegi panjang, biasanya dilambangkan dengan huruf kapital yang tebal misalnya **A**. Suatu matriks dengan m baris dan n kolom disebut dengan matriks $m \times n$, dan dinotasikan sebagai berikut :

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \vdots & a_{2n} \\ \vdots & \dots & \ddots & \dots \\ a_{m1} & a_{m2} & \vdots & a_{mn} \end{bmatrix}$$

Mari kita notasikan kolom ke k dari matriks A dengan a_k , maka dari itu

$$a_k = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix}$$

Jumlah maksimal kolom bebas linear dari matriks A disebut tingkat dari matriks A , dinotasikan dengan $rank A$. Perhatikan bahwa tingkat A adalah dimensi dari $span[a_1, \dots, a_n]$.

Proposisi 2.3 Tingkat dari matriks A adalah invarian pada operasi berikut ini :

- 1) Perkalian kolom matriks A dengan skalar tak nol
 - 2) Pertukaran kolom
 - 3) Penambahan ke kolom tertentu merupakan kombinasi linear dari kolom lainnya
-

Bukti.

- 1) Misalkan $b_k = \alpha_k a_k$, dimana $\alpha_k \neq 0$, $k = 1, \dots, n$, dan misalkan $B = [b_1, b_2, \dots, b_n]$.
Terlihat jelas bahwa

$$span[a_1, a_2, \dots, a_n] = span[b_1, b_2, \dots, b_n]$$

sehingga

$$rank A = rank B$$

- 2) Banyaknya vektor bebas linear tidak bergantung pada ordonya
- 3) Misalkan

$$b_1 = a_1 + c_2 a_2 + \dots + c_n a_n$$

$$b_2 = a_2$$

\vdots

$$b_n = a_n$$

Jadi, untuk setiap $\alpha_1, \dots, \alpha_n$

$$\alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_n a_n = \alpha_1 a_1 + (\alpha_2 + \alpha_1 c_2) a_2 + \dots + (\alpha_n + \alpha_1 c_n) a_n$$

dan karena,

$$span[b_1, b_2, \dots, b_n] \subset span[a_1, a_2, \dots, a_n]$$

Dilain sisi

$$\begin{aligned}a_1 &= b_1 - c_2 b_2 - \dots - c_n b_n \\a_2 &= b_2 \\&\vdots \\a_n &= b_n\end{aligned}$$

Karena,

$$\text{span}[a_1, a_2, \dots, a_n] \subset \text{span}[b_1, b_2, \dots, b_n]$$

Oleh karena itu,

$$\text{rank } A = \text{rank } B$$

Berkaitan dengan setiap persegi ($n \times n$) matriks A adalah skalar yang disebut determinan dari matriks A , dinotasikan dengan $\det A$ atau $|A|$. Determinan matriks persegi adalah fungsi dari kolom-kolomnya dan memiliki sifat-sifat berikut :

- 1) Determinan matriks $A = [a_1, a_2, \dots, a_n]$ adalah fungsi linear setiap kolom, maka dari itu

$$\begin{aligned}\det[a_1, \dots, a_{k-1}, \alpha a_k^{(1)} + \beta a_k^{(2)}, a_{k+1}, \dots, a_n] &= \alpha \det[a_1, \dots, a_{k-1}, a_k^{(1)}, a_{k+1}, \dots, a_n] \\&+ \beta \det[a_1, \dots, a_{k-1}, a_k^{(2)}, a_{k+1}, \dots, a_n]\end{aligned}$$

Untuk setiap $\alpha, \beta \in \mathbb{R}, a_k^{(1)}, a_k^{(2)} \in \mathbb{R}^n$

- 2) Jika untuk beberapa k kita memiliki $a_k = a_{k+1}$, maka

$$\det A = \det[a_1, \dots, a_k, a_{k+1}, \dots, a_n] = \det[a_1, \dots, a_k, a_k, \dots, a_n] = 0$$

- 3) Misalkan

$$I_n = [e_1, e_2, \dots, e_n] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Dimana $\{e_1, \dots, e_n\}$ adalah basis natural untuk \mathbb{R}^n . Kemudian,

$$\det I_n = 1$$

Perhatikan bahwa jika $\alpha = \beta = 0$ pada sifat 1, maka

$$\det[a_1, \dots, a_{k-1}, 0, a_{k+1}, \dots, a_n] = 0$$

Jadi, jika salah satu kolom adalah 0, maka determinannya sama dengan nol.

Determinan tidak mengubah nilainya jika kita menambahkan ke kolom-kolom lain dikalikan dengan skalar. Ini mengikuti dari sifat 1 dan sifat 2 seperti yang ditunjukkan di bawah ini :

$$\begin{aligned} \det[a_1, \dots, a_{k-1}, a_k + \alpha a_j, a_{k+1}, \dots, a_j] &= \det[a_1, \dots, a_{k-1}, a_k, a_{k+1}, \dots, a_j] + \\ &\quad \alpha \det[a_1, \dots, a_{k-1}, a_j, \dots, a_n] \\ &= \det[a_1, \dots, a_n] \end{aligned}$$

Namun, determinan berubah tandanya jika kita menukar kolom. Untuk menunjukkan sifatnya perhatikan bahwa

$$\begin{aligned} &= \det[a_1, \dots, a_{k-1}, a_k, a_{k-1}, \dots, a_n] = \det[a_1, \dots, a_k + a_{k+1}, a_{k-1}, \dots, a_n] \\ &= \det[a_1, \dots, a_k + a_{k+1}, a_{k-1} - (a_k + a_{k+1}), \dots, a_n] \\ &= \det[a_1, \dots, a_k + a_{k+1}, -a_k, \dots, a_n] \\ &= -\det[a_1, \dots, a_k + a_{k+1}, a_k, \dots, a_n] \\ &= -(\det[a_1, \dots, a_k, a_k, \dots, a_n] + \det[a_1, \dots, a_{k+1}, a_k, \dots, a_n]) \\ &= -\det[a_1, \dots, a_{k+1}, a_k, \dots, a_n] \end{aligned}$$

Proposisi 2.4 Jika $m \times n$ ($m \geq n$) matriks A memiliki minor orde ke- n bukan nol, maka kolom-kolom dari A adalah bebas linear, yakni, $rank A = n$

Bukti. Misalkan A memiliki minor orde ke- n bukan nol. Kita asumsikan bahwa minor orde ke- n yang bersesuaian dengan n baris pertama A adalah bukan nol.

Misalkan x_i , dengan $i = 1, \dots, n$, adalah skalar sedemikian sehingga

$$x_1 a_1 + x_2 a_2 + \dots + x_n a_n = 0$$

Persamaan vektor di atas ekuivalen dengan himpunan persamaan m berikut :

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = 0$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = 0$$

$$\begin{aligned}
& \vdots \\
& a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = 0 \\
& \vdots \\
& a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0
\end{aligned}$$

Untuk $i = 1, \dots, n$, misalkan

$$\tilde{a}_i = \begin{bmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{bmatrix}$$

Maka, $x_1\tilde{a}_1 + \cdots + x_n\tilde{a}_n = 0$

Minor orde ke- n adalah $\det[\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_n]$ diasumsikan tak nol. Dari sifat-sifat determinan maka kolom $\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_n$ adalah bebas linear. Karena itu, semua $x_i = 0$, dengan $i = 1, \dots, n$. Karenanya, kolom a_1, a_2, \dots, a_n adalah bebas linear.

Dari persamaan di atas, jika ada minor bukan nol, maka kolom terkait dengan minor bukan nol ini adalah bebas linear.

Jika suatu matriks A memiliki minor orde ke- r $|M|$ dengan sifat (i) $|M| \neq 0$ dan (ii) setiap minor dari A yang dibentuk dengan menambahkan baris dan kolom A ke M adalah nol, maka

$$rank A = r$$

Jadi, tingkat suatu matriks adalah sama dengan orde tertinggi dari minor tak nol.

Matriks nonsingular adalah matriks persegi yang determinannya adalah tak nol

Misalkan A adalah matriks persegi $n \times n$. Maka, A nonsingular jika dan hanya jika

Ada $n \times n$ matriks B lain sedemikian sehingga

$$AB = BA = I_n$$

Dimana I_n menotasikan atau menunjukkan identitas matriks $\times n$:

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Kita sebut matriks B diatas sebagai invers dari A , dan dituliskan $B = A^{-1}$.

Diberikan matriks $m \times n$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \vdots & a_{m2} \\ \vdots & \cdots & \ddots & \cdots \\ a_{m1} & a_{2n} & \vdots & a_{mn} \end{bmatrix}$$

Transpose dari A dinotasikan dengan A^T , matriks $n \times m$ adalah

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \vdots & a_{m2} \\ \vdots & \cdots & \ddots & \cdots \\ a_{1n} & a_{2n} & \vdots & a_{mn} \end{bmatrix}$$

Jelas bahwa kolom matriks A adalah baris matriks A^T , demikian juga berlaku sebaliknya.

Matriks A adalah simetris jika $A = A^T$

2.3 Persamaan Linear

Misalkan kita memiliki m persamaan dalam n yang tidak diketahui bentuknya

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m$$

Kita dapat merepresentasikan himpunan persamaan di atas sebagai persamaan vektor

$$x_1a_1 + x_2a_2 + \cdots + x_na_n = b$$

Dimana

$$a_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Berkaitan dengan sistem persamaan diatas, berikut ini adalah matriksnya

$$A = [a_1, a_2, \dots, a_n]$$

Dan matriks yang diperbesar

$$[A \ b] = [a_1, a_2, \dots, a_n, b]$$

Kita juga dapat merepresentasikan sistem persamaan di atas sebagai

$$Ax = b$$

Dimana

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Teorema : Sistem persamaan $Ax = b$ memiliki solusi jika dan hanya jika

$$\text{rank } A = \text{rank}[A \ b]$$

Pembuktian : Misalkan sistem $Ax = b$ memiliki solusi. Karena itu, b adalah kolom kombinasi linear dari A , yakni, terdapat x_1, \dots, x_n sedemikian sehingga

$$x_1 a_1 + x_2 a_2 + \dots + x_n a_n = b$$

Oleh karena itu b termasuk $\text{span}[a_1, \dots, a_n]$ dan karenanya

$$\begin{aligned} \text{rank } A &= \dim \text{span} [a_1, \dots, a_n] \\ &= \dim \text{span} [a_1, \dots, a_n, b] \\ &= \text{rank} [A \ b] \end{aligned}$$

Misalkan $\text{rank } A = \text{rank}[A \ b] = r$. Jadi, kita memiliki r kolom bebas linear dari A . Misalkan a_1, a_2, \dots, a_r menjadi kolom ini. Oleh karena itu, a_1, a_2, \dots, a_r juga merupakan kolom-kolom bebas linear dari matriks $[A \ b]$. Karena $\text{rank}[A \ b] = r$, kolom yang tersisa dari $[A \ b]$ dapat dinyatakan sebagai kombinasi linier dari a_1, a_2, \dots, a_r . Secara khusus, b dapat dinyatakan sebagai kombinasi linear dari kolom-kolom tersebut. Oleh karena itu, terdapat x_1, \dots, x_n sedemikian sehingga $x_1 a_1 + x_2 a_2 + \dots + x_n a_n = b$

Simbol $\mathbb{R}^{m \times n}$ menunjukkan bahwa himpunan matriks $m \times n$ yang elemen-elemennya bilangan real

Teorema 2.1 Diberikan persamaan $Ax = b$, dimana $A \in \mathbb{R}^{m \times n}$, dan $\text{rank } A = m$. Sebuah solusi untuk $Ax = b$ dapat diperoleh dengan menentukan nilai yang berubah-ubah untuk $n - m$ variabel

Bukti

Kita memiliki $\text{rank } A = m$, dan kemudian kita dapat menemukan m kolom-kolom bebas linear dari A . Misalkan a_1, a_2, \dots, a_m menjadi kolom.

Tulis ulang persamaan $Ax = b$ seperti

$$x_1 a_1 + x_2 a_2 + \dots + x_m a_m = b - x_{m+1} a_{m+1} - \dots - x_n a_n$$

Tentukan $x_{m+1}, x_{m+2}, \dots, x_n$ nilai yang berubah-ubah, katakan

$$x_{m+1} = d_{m+1}, x_{m+2} = d_{m+2}, \dots, x_n = d_n$$

Dan misalkan

$$B = [a_1, a_2, \dots, a_m] \in \mathbb{R}^{m \times n}$$

Perhatikan bahwa $\det B \neq 0$. Kita dapat merepresentasikan sistem persamaan diatas sebagai

$$B = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = [b - d_{m+1} a_{m+1} - \dots - d_n a_n]$$

Matriks B adalah tidak singular, dan kemudian kita dapat menyelesaikan untuk $[x_1, x_2, \dots, x_m]^T$.
Secara khusus,

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = B^{-1}[b - d_{m+1}a_{m+1} - \dots - d_n a_n]$$

2.4 Perkalian Titik dan Aturannya

Nilai mutlak dari bilangan real a , dinotasikan $|a|$, didefinisikan sebagai berikut :

$$|a| = \begin{cases} a, & \text{if } a \geq 0 \\ -a, & \text{if } a < 0 \end{cases}$$

Formula berikut berlaku :

- 1) $|a| = |-a|$;
- 2) $-|a| \leq a \leq |a|$;
- 3) $|a + b| \leq |a| + |b|$;
- 4) $||a| - |b|| \leq |a - b| \leq |a| + |b|$;
- 5) $|ab| = |a||b|$;
- 6) $|a| \leq c$ dan $|b| \leq d$ berarti $|a + b| \leq c + d$;
- 7) Pertidaksamaan $|a| < b$ adalah ekuivalen dengan $-b < a < b$

Untuk $x, y \in \mathbb{R}^n$, kita definisikan Perkalian Titik Euclid

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i = x^T y$$

Perkalian titik adalah fungsi bernilai real $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ yang memiliki sifat-sifat berikut ini :

- 1) Positivity : $\langle x, x \rangle \geq 0$, $\langle x, x \rangle = 0$ jika dan hanya jika $x = 0$;
- 2) Simetri : $\langle x, y \rangle = \langle y, x \rangle$;
- 3) Additivity : $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$;
- 4) Homogeneity : $\langle rx, y \rangle = r\langle x, y \rangle$, $\forall r \in \mathbb{R}$

Sifat dari additivity dan homogeneity pada vektor kedua juga berlaku, yaitu

$$\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$$

$$\langle x, ry \rangle = r\langle x, y \rangle, \quad \forall r \in \mathbb{R}$$

Sifat diatas dapat ditunjukkan menggunakan sifat 2 sampai 4, yakni

$$\begin{aligned}\langle x, y + z \rangle &= \langle y + z, x \rangle \\ &= \langle y, x \rangle + \langle z, x \rangle \\ &= \langle x, y \rangle + \langle x, z \rangle\end{aligned}$$

Dan

$$\langle x, ry \rangle = r\langle x, y \rangle = r\langle y, x \rangle = r\langle x, y \rangle$$

Ini memungkinkan untuk mendefinisikan fungsi bernilai real lainnya pada $\mathbb{R}^n \times \mathbb{R}^n$ yang memenuhi sifat 1 sampai 4 diatas. Banyak hasil yang melibatkan perkalian titik Euclidean juga berlaku untuk perkalian titik lainnya

Vektor x dan y dikatakan orthogonal jika $\langle x, y \rangle = 0$

Norma Euclidean dari vektor x didefinisikan sebagai berikut :

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

Teorema 2.3 Cauchy-Schwarz Inequality

Untuk setiap dua vektor x dan y di \mathbb{R}^n , pertidaksamaan Cauchy-Schwarz adalah berlaku sebagai berikut :

$$|\langle x, y \rangle| \leq ||x|| ||y||$$

Selanjutnya, persamaan berlaku jika dan hanya jika $x = \alpha y$ untuk beberapa $\alpha \in \mathbb{R}$

Bukti

Pertama kita asumsikan bahwa x dan y adalah vektor satuan, yakni $||x|| = ||y|| = 1$. Maka,

$$\begin{aligned}
0 &\leq \|x - y\|^2 = \langle x - y, x - y \rangle \\
&= \|x\|^2 - 2\langle x, y \rangle + \|y\|^2 \\
&= 2 - 2\langle x, y \rangle
\end{aligned}$$

Atau

$$\langle x, y \rangle \leq 1$$

Dengan persamaan berlaku jika dan hanya jika $x = y$

Selanjutnya, dengan asumsi bahwa baik x maupun y adalah nol (karena pertidaksamaan berlaku jika salah satunya adalah nol), kita mengubah x dan y dengan vektor satuan $x/\|x\|$ dan $y/\|y\|$. Kemudian, terapkan sifat 4 untuk mendapatkan

$$\langle x, y \rangle \leq \|x\| \|y\|$$

Sekarang kita ubah x dengan $-x$ dan kemudian, terapkan lagi sifat 4 untuk mendapatkan

$$-\langle x, y \rangle \leq \|x\| \|y\|$$

Dua pertidaksamaan terakhir menunjukkan pertidaksamaan nilai mutlak. Persamaan berlaku jika dan hanya jika $x/\|x\| = \pm y/\|y\|$, dengan $x = \alpha y$ untuk beberapa $\alpha \in \mathbb{R}$.

Norma Euclidean dari sebuah vektor $\|x\|$ memiliki sifat-sifat berikut :

- 1) Positivity : $\|x\| \geq 0$, $\|x\| = 0$ jika dan hanya jika $x = 0$;
- 2) Homogeneity : $\|rx\| = |r|\|x\|$, $r \in \mathbb{R}$;
- 3) Triangle Inequality : $\|x + y\| \leq \|x\| + \|y\|$

Ketaksamaan segitiga dapat dibuktikan dengan menggunakan pertidaksamaan Cauchy-Schwarz sebagai berikut

$$\|x + y\|^2 = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2$$

Dengan menggunakan pertidaksamaan Cauchy-Schwarz,

$$\begin{aligned} \|x + y\|^2 &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2 \end{aligned}$$

Dan kemudian

$$\|x + y\| \leq \|x\| + \|y\|$$

Perhatikan bahwa jika x dan y adalah orthogonal, yaitu, $\langle x, y \rangle = 0$, maka

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2$$

Yang mana hal tersebut adalah Teorema Pythagoras untuk \mathbb{R}^n

Norma Euclidean adalah contoh dari norma vektor umum, yang merupakan fungsi apapun yang memenuhi tiga sifat yaitu positivity, homogeneity, dan ketaksamaan segitiga. Contoh lain dari norma vektor pada \mathbb{R}^n termasuk 1-norma, yang didefinisikan oleh $\|x\|_1 = |x_1| + \dots + |x_n|$, dan ∞ -norma, yang didefinisikan oleh $\|x\| = \max_i |x_i|$. Norma Euclidean sering disebut sebagai 2-norma, dan dilambangkan $\|x\|_2$. Norma di atas adalah kasus khusus dari p-norma, diberikan oleh

$$\|x\|_p = \begin{cases} (|x_1|^p + \dots + |x_n|^p)^{1/p}, & \text{jika } 1 \leq p < \infty \\ \max(|x_1|, \dots, |x_n|), & \text{jika } p = \infty \end{cases}$$

Kita dapat menggunakan norma untuk mendefinisikan pengertian fungsi kontinu, sebagai berikut. Suatu fungsi $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ adalah kontinu di x jika untuk setiap $\varepsilon > 0$, terdapat $\delta > 0$ sedemikian sehingga $\|y - x\| < \delta \Rightarrow \|f(y) - f(x)\| < \varepsilon$. Jika fungsi f adalah kontinu pada setiap titik di \mathbb{R}^n , kita sebut bahwa hal itu kontinu di \mathbb{R}^n . Perhatikan bahwa $f = [f_1, \dots, f_m]^T$ adalah kontinu jika dan hanya jika setiap komponen f_i , dengan $i = 1, \dots, m$, adalah kontinu

Untuk vektor ruang kompleks \mathbb{C}^n , kita definisikan perkalian titik $\langle x, y \rangle$ menjadi

$$\sum_{i=1}^n x_i \bar{y}_i$$

Dimana bar diatas y_i menunjukkan konjugasi kompleks. Perkalian titik pada \mathbb{C}^n adalah fungsi bernilai kompleks yang memiliki sifat-sifat berikut ini :

- 1) $\langle x, x \rangle \geq 0$, $\langle x, y \rangle = 0$ jika dan hanya jika $x = 0$
- 2) $\langle x, y \rangle = \overline{\langle y, x \rangle}$;
- 3) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- 4) $\langle rx, y \rangle = r\langle x, y \rangle$, dimana $r \in \mathbb{C}$

Dari sifat 1 sampai 4, kita dapat menyimpulkan sifat-sifat lain, seperti

$$\langle x, r_1y + r_2z \rangle = \bar{r}_1\langle x, y \rangle + \bar{r}_2\langle x, z \rangle$$

Dimana $r_1, r_2 \in \mathbb{C}$. Untuk \mathbb{C}^n , norma vektor dapat didefinisikan oleh $\|x\|^2 = \langle x, x \rangle$

Contoh soal:

2.1. Diberikan $A \in \mathbb{R}^{m \times n}$ dan $\text{rank } A = m$. Tunjukkan bahwa $m \leq n$.

Penyelesaian:

Dengan menggunakan kontradiksi. Dimisalkan $n < m$. Maka, jumlah kolom A adalah n. Karena rank A jumlah maksimum kolom bebas linier, maka rank A tidak bisa lebih besar dari n < m. Jadi kontradiksi dengan asumsi bahwa rank A = m.

2.2. Buktikan bahwa sistem $Ax = b$, $A \in \mathbb{R}^{m \times n}$, memiliki solusi yang unik jika dan hanya jika $\text{rank } A = \text{rank } [A \ b] = n$.

Penyelesaian:

Terdapat solusi dalam teorema 2.1, $\text{rank } A = \text{rank } [A: b]$. Jadi, masih harus membuktikan bahwa $\text{rank } A = n$. Untuk itu, dimisalkan $\text{rank } A < n$ (Perhatikan bahwa tidak mungkin $\text{rank } A > n$ selama A hanya kolom n). Oleh karena itu, terdapat $y \in \mathbb{R}^n, y \neq 0$, sedemikian sehingga $Ay = 0$ (karena kolom A bebas linier, dan Ay linier kombinasi kolom A. Dimisalkan x menjadi solusi $Ax = b$. Maka jelas bahwa $x + y \neq x$ juga merupakan solusi. Dimana ini bertentangan dengan solusi yang unik. Oleh karena itu, $\text{rank } A = n$.

Dengan teorema 2.1, terdapat solusi. Jadikan x dan y solusi, dimana $Ax = b$ dan $Ay = b$. Didapatkan $A(x - y) = 0$. Saat $\text{rank } A = n$ dan A memiliki kolom n , maka $x - y = 0$ dan oleh karena itu $x = y$ menunjukkan bahwa memiliki solusi yang unik.

2.3. Diketahui bahwa $k \geq n + 1$, maka vector $a_1, a_2, \dots, a_k \in \mathbb{R}^n$ tidak bebas linear, terdapat skalar $\alpha_1, \dots, \alpha_k$ sehingga setidaknya satu $\alpha_i \neq 0$ dan $\sum_{i=1}^k \alpha_i a_i = 0$.

Petunjuk : memperkenalkan vector $\bar{a}_i = [1, a_i^T]^T \in \mathbb{R}^{n+1}, i = 1, \dots, k$, dan menggunakan fakta bahwa terdapat $n + 2$ vektor di \mathbb{R}^{n+1} yang tidak bebas linear.

Penyelesaian:

Anggap vector $\bar{a}_i = [1, a_i^T]^T \in \mathbb{R}^{n+1}, i = 1, \dots, k$. Karena $k \geq n + 2$, maka vector $\bar{a}_1, \dots, \bar{a}_k$ menjadi bebas linear di \mathbb{R}^{n+1} . Oleh karena itu. Terdapat $\alpha_1, \dots, \alpha_k$, yang tidak semua nol, sehingga

$$\sum_{i=1}^k \alpha_i \bar{a}_i = 0.$$

Komponen pertama dari persamaan vector di atas adalah $\sum_{i=1}^k \alpha_i = 0$, sedangkan n komponen terakhir memiliki bentuk $\sum_{i=1}^k \alpha_i a_i = 0$, melengkapi pembuktian.

2.4. Buktikan tujuh sifat nilai absolut dari bilangan real.

a) Penerapan dari definisi $|-a|$:

$$|-a| = \begin{cases} -a & \text{jika } -a > 0 \\ 0 & \text{jika } -a = 0 \\ -(-a) & \text{jika } -a < 0 \end{cases}$$

$$= \begin{cases} -a & \text{jika } a < 0 \\ 0 & \text{jika } a = 0 \\ a & \text{jika } a > 0 \end{cases}$$

$$= |a|$$

b) Jika $a \geq 0$, maka $|a| = a$. Jika $a < 0$, maka $|a| = -a > 0 > a$. Sehingga $|a| \geq a$. Disisi lain, $|-a| \geq -a$ (dengan hal di atas). Maka, $a \geq -|-a| = -|a|$ (dengan properti 1).

c) Terdapat empat kasus yang perlu dipertimbangkan.

- Pertama, jika $a, b \geq 0$, maka $a + b \geq 0$. Sehingga, $|a + b| = a + b = |a| + |b|$.
- Kedua, jika $a, b \leq 0$, maka $a + b \leq 0$. Sehingga, $|a + b| = -(a + b) = -a - b = |a| + |b|$.
- Ketiga, jika $a \geq 0$ dan $b \leq 0$, maka terdapat dua kondisi yang lebih lanjut:
Jika $a + b \geq 0$, maka $|a + b| = a + b \leq |a| + |b|$.
Jika $a + b \leq 0$, maka $|a + b| = -a - b \leq |a| + |b|$.
- Keempat, $a \leq 0$ dan $b \geq 0$, itu identik dengan kasus ketiga, dengan a dan b ditukar.

d) Tunjukkan $|a - b| \leq |a| + |b|$. Sehingga

$$|a - b| = |a + (-b)| \leq |a| + |-b| = |a| + |b|$$

Untuk menunjukkan $||a| - |b|| \leq |a - b|$, perhatikan bahwa $|a| = |a - b + b| \leq |a - b| + |b|$, yang berarti $|a| - |b| \leq |a - b|$. Disisi lain, dari pernyataan di atas kita memiliki $|b| - |a| \leq |b - a| = |a - b|$ dengan property 1. Jadi, $||a| - |b|| \leq |a - b|$.

- e) Memiliki empat kasus. Pertama, jika $a, b \geq 0$, memiliki $ab \geq 0$ maka $|ab| = ab = |a||b|$. Kedua, jika $a, b \leq 0$, memiliki $ab \geq 0$ maka $|ab| = ab = (-a)(-b) = |a||b|$. Ketiga, jika $a \leq 0, b \geq 0$, memiliki $ab \leq 0$ maka $|ab| = -ab = a(-b) = |a||b|$. Keempat, $a \geq 0$ dan $b \leq 0$, itu identik dengan kasus ketiga, dengan a dan b ditukar.
- f) Mempunyai

$$|a - b| \leq |a| + |b|$$

$$\leq c + d.$$

- g) Pada properti 2, $-a \leq |a|$ dan $a \leq |a|$. Sehingga, $|a| < b$ berarti $-a \leq |a| < b$ dan $a \leq |a| < b$.

Jika $a \geq 0$, maka $|a| = a < b$. Jika $a < 0$, maka $|a| = -a < b$.

Untuk kasus saat “<” diganti dengan “≤”, itu hanya mengulangi bukti yang di atas dengan mengganti “<” menjadi “≤”.

- h) Ini hanya negasi properti 7, menerapkan hukum DeMorgan’s.

2.5. Pertimbangkanlah fungsi $\langle ., . \rangle_2: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$, dengan definisi $\langle x, y \rangle_2 = 2x_1y_1 + 3x_2y_1 + 3x_1y_2 + 5x_2y_2$, dimana $x = [x_1, x_2]^T$ dan $y = [y_1, y_2]^T$. Tunjukkan bahwa $\langle ., . \rangle_2$ memenuhi

kondisi 1 sampai 4 untuk produk inner. *Catatan : ini merupakan kasus khusus dari latihan 3.14.*

Penyelesaian:

Perhatikan bahwa $\langle x, y \rangle_2$ menunjukkan

$$\langle x, y \rangle_2 = x^T \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix} y = (Qx)^T(Qy) = x^T Q^2$$

Dimana,

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

Perhatikan bahwa matriks $Q = Q^T$ merupakan non-singular.

1) $\langle x, x \rangle_2 = (Qx)^T(Qx) = \|Qx\|^2 \geq 0$, dan

$$\begin{aligned} \langle x, x \rangle_2 = 0 &\Leftrightarrow \|Qx\|^2 = 0 \\ &\Leftrightarrow Qx = 0 \\ &\Leftrightarrow x = 0 \end{aligned}$$

dimana Q adalah non-singular.

2) $\langle x, y \rangle_2 = (Qx)^T(Qy) = (Qy)^T(Qx) = \langle y, x \rangle_2$.

3) $\begin{aligned} \langle x + y, z \rangle_2 &= (x + y)^T Q^2 z \\ &= x^T Q^2 z + y^T Q^2 z \\ &= \langle x, z \rangle_2 + \langle y, z \rangle_2. \end{aligned}$

4) $\langle rx, y \rangle_2 = (rx)^T Q^2 y = rx^T Q^2 y = r \langle x, y \rangle_2$.

2.6. Tunjukkan bahwa terdapat dua vector $x, y \in \mathbb{R}^n$, $|||x|| - ||y||| \leq \|x - y\|$.

Petunjuk : Tulis $x = (x - y) + y$, dan gunakan ketidaksetaraan segitiga. Lakukanlah hal yang sama untuk y.

Penyelesaian:

$\|x\| = \|(x - y) + y\| \leq \|x - y\| + \|y\|$ dengan ketidaksetaraan segitiga. Maka, $\|x\| - \|y\| \leq \|x - y\|$. Disisi lain, dari pernyataan di atas $\|y\| - \|x\| \leq \|y - x\| = \|x - y\|$.

$y\|$. Menggabungkan dua ketidaksetaraan tersebut, sehingga diperoleh $|\|x\| - \|y\|| \leq \|x - y\|$.

2.7. Gunakan latihan 2.6 untuk menunjukkan norma $\|\cdot\|$ adalah fungsi kontinu seragam., yaitu untuk semua $\varepsilon > 0$, *terdapat* $\delta > 0$ sedemikian sehingga jika $\|x - y\| < \delta$, *maka* $\|x\| - \|y\| < \varepsilon$.

Penyelesaian:

Misalkan $\varepsilon > 0$ *diberikan*. Set $\delta = \varepsilon$. Oleh karena itu, jika $\|x\| - \|y\| \leq \delta$, *maka dengan pembahasan latihan 2.6*, $|\|x\| - \|y\|| \leq \|x - y\| < \delta = \varepsilon$.

BAB 3

Transformasi

3.1 Transformasi Linear

Sebuah fungsi $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ disebut transformasi linear jika

- 1) $\mathcal{L}(ax) = a\mathcal{L}(x) \forall x \in \mathbb{R}^n$ dan $a \in \mathbb{R}$
- 2) $\mathcal{L}(x + y) = \mathcal{L}(x) + \mathcal{L}(y) \forall x, y \in \mathbb{R}^n$

Jika kita menentukan basis dari \mathbb{R}^n dan \mathbb{R}^m , maka transformasi linear \mathcal{L} bisa dinyatakan dalam sebuah matriks. Secara khusus, $A \in \mathbb{R}^{n+m}$ ada sehingga representasi tersebut berlaku. Misalkan $x \in \mathbb{R}^n$ adalah sebuah vector dan x' adalah representasi dari x terhadap basis yang diberikan untuk \mathbb{R}^n . Jika $y = \mathcal{L}(x)$ dan y' adalah representasi dari y terhadap basis yang diberikan untuk \mathbb{R}^m , maka

$$y' = Ax'$$

A disebut representasi matriks dari \mathcal{L} terhadap basis yang diberikan untuk \mathbb{R}^n dan \mathbb{R}^m . Dalam kasus khusus dimana kita mengasumsikan basis alami dari \mathbb{R}^n dan \mathbb{R}^m representasi matriks A memenuhi

$$\mathcal{L}(x) = Ax$$

Diberikan $\{e_1, e_2, \dots, e_n\}$ dan $\{e'_1, e'_2, \dots, e'_n\}$ menjadi basis dua untuk \mathbb{R}^n . Definisikan matriks

$$T = \{e'_1, e'_2, \dots, e'_n\}^{-1} \{e_1, e_2, \dots, e_n\}$$

T disebut transformasi matriks dari $\{e_1, e_2, \dots, e_n\}$ ke $\{e'_1, e'_2, \dots, e'_n\}$. Jelas bahwa

$$[e_1, e_2, \dots, e_n] = [e'_1, e'_2, \dots, e'_n] T$$

Yaitu, kolom ke- i dari T disebut vector koordinat dari e_i terhadap basis $\{e'_1, e'_2, \dots, e'_n\}$.

Tentukan vector di \mathbb{R}^n dan misalkan x adalah kolom dari vektor terhadap $[e_1, e_2, \dots, e_n]$ dan x' adalah koordinat dari vector yang sama terhadap $[e'_1, e'_2, \dots, e'_n]$. Maka, akan ditunjukkan bahwa $x' = Tx$.

Perhatikan transformasi linear :

$$\mathcal{L}: \mathbb{R}^n \rightarrow \mathbb{R}^n$$

Misalkan, A adalah representasi terhadap $\{e_1, e_2, \dots, e_n\}$ dan B adalah representasi terhadap $\{e'_1, e'_2, \dots, e'_n\}$. Misalkan $y = Ax$ dan $y' = Bx'$. Oleh karena itu, $y' = Ty = TAx = Bx' = BTx$ dan karenanya $TA = BT$ atau $A = T^{-1}BT$.

Dua buah matriks $n \times n$ A dan B sebangun jika terdapat matriks tak tunggal T sehingga $A = T^{-1}BT$. Kesimpulannya, matriks serupa sesuai dengan transformasi linear yang sama dengan basis yang berbeda.

3.2 Nilai Eigen dan Vektor Eigen

Diberikan A sebuah matriks persegi $n \times n$. A scalar λ dan vector v tak nol memenuhi persamaan $Av = \lambda v$, masing – masing dikatakan sebagai nilai eigen dan vector eigen dari A . Agar λ menjadi nilai eigen, maka perlu matriks $\lambda I - A$ menjadi singular, yaitu

$$\det[\lambda I - A] = 0$$

Dimana, I adalah identitas matriks $n \times n$. Ini mengarah ke persamaan polynomial orde- n .

$$\det[\lambda I - A] = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0 = 0$$

Polynomial $\det[\lambda I - A]$ disebut karakteristik polynomial dari matriks A dan persamaan diatas disebut persamaan karakteristik. Menurut teorema dasar aljabar, persamaan karakteristik harus memiliki n (mungkin tidak berbeda) akar yang merupakan nilai eigen dari A . Teorema berikut menyatakan bahwa jika A memiliki n nilai eigen yang berbeda, maka ia juga memiliki n vektor eigen yang bebas linier.

Teorema 3.1. Misalkan persamaan karakteristik $\det[\lambda I - A] = 0$ memiliki n akar yang berbeda yaitu $\lambda_1, \lambda_2, \dots, \lambda_n$. Maka, terdapat n vektor bebas linear v_1, v_2, \dots, v_n sehingga:

$$Av_i = \lambda_i v_i, \quad i = 1, 2, \dots, n$$

Bukti: Karena $\det[\lambda I - A] = 0$, $i = 1, 2, \dots, n$ terdapat v_i tak nol, $i = 1, \dots, n$. Sehingga, $Av_i = \lambda_i v_i$, $i = 1, \dots, n$. Sekarang, buktikan bahwa linear bebas dari $\{v_1, v_2, \dots, v_n\}$. Misalkan c_1, \dots, c_n adalah scalar, sehingga

$$\sum_{i=1}^n c_i v_i = 0$$

Perhatikan matriks berikut

$$Z = (\lambda_2 I - A)(\lambda_3 I - A) \dots (\lambda_n I - A)$$

Pertama, tunjukkan bahwa $c_1 = 0$. perhatikan

$$\begin{aligned} Zv_n &= (\lambda_2 I - A)(\lambda_3 I - A) \dots (\lambda_{n-1} I - A)(\lambda_n I - A)v_n \\ &= (\lambda_2 I - A)(\lambda_3 I - A) \dots (\lambda_{n-1} I - A)(\lambda_n v_n - Av_n) \\ &= 0 \end{aligned}$$

Ketika $\lambda_n v_n - Av_n = 0$. Dengan mengulang argument diatas kita punya

$$Zv_k = 0, \quad k = 2, 3, \dots, n$$

Tetapi

$$\begin{aligned} Zv_1 &= (\lambda_2 I - A)(\lambda_3 I - A) \dots (\lambda_{n-1} I - A)(\lambda_n I - A)v_1 \\ &= (\lambda_2 I - A)(\lambda_3 I - A) \dots (\lambda_{n-1} v_1 - Av_1)(\lambda_n - \lambda_1) : \\ &= (\lambda_2 I - A)(\lambda_3 I - A)v_1 \dots (\lambda_{n-1} - \lambda_1)(\lambda_n - \lambda_1) \\ &= (\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1) \dots (\lambda_{n-1} - \lambda_1)(\lambda_n - \lambda_1)v_1 \end{aligned}$$

Dari persamaan diatas, diperoleh

$$Z \left(\sum_{i=1}^n c_i v_i \right) = \sum_{i=1}^n c_i Zv_i = c_i Zv_i = c_i (\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1) \dots (\lambda_n - \lambda_1)v_1 = 0$$

Karena, λ_i berbeda, maka $c_1 = 0$.

Dengan menggunakan argumen yang sama, tunjukkan bahwa semua c_i harus hilang dan oleh karena itu himpunan vector eigen $\{v_1, v_2, \dots, v_n\}$ bebas linear.

Perhatikan basis yang dibentuk oleh himpunan vektor eigen yang bebas linear $\{v_1, v_2, \dots, v_n\}$. Dengan memperhatikan basis ini, matriks A adalah diagonal yaitu jika a_{ij} adalah elemen ke $-(i, j)$ dari A . Maka $a_{ij} = 0 \forall i \neq j$. Misal

$$T = [v_1, v_2, \dots, v_n]$$

Maka

$$\begin{aligned}
T^{-1}AT &= T^{-1}A(v_1, v_2, \dots, v_n) = T^{-1}[Av_1, Av_2, \dots, Av_n] = T^{-1}[\lambda_1 v_1, \lambda_2 v_2, \dots, \lambda_n v_n] \\
&= T^{-1}T \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \\ & & & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \\ & & & \lambda_n \end{bmatrix}
\end{aligned}$$

Karena, $T^{-1}T = I$.

Teorema 3.2. Semua nilai eigen matriks simetris adalah nyata.

Bukti: Misalkan, $Ax = \lambda x$, dimana $x \neq 0$. Ambil inner produk dari Ax dengan hasil x .

$$\langle Ax, x \rangle = \langle \lambda x, x \rangle = \lambda \langle x, x \rangle$$

Dengan kata lain,

$$\langle Ax, x \rangle = \langle x, A^T x \rangle = \langle x, Ax \rangle = \langle x, \lambda x \rangle = \bar{\lambda} \langle x, x \rangle.$$

Definisi diatas adalah inner produk pada C^n . Ingat bahwa $\langle x, x \rangle$ nyata dan $\langle x, x \rangle > 0$. Karena

$$\lambda \langle x, x \rangle = \bar{\lambda} \langle x, x \rangle$$

Dan

$$(\lambda - \bar{\lambda}) \langle x, x \rangle = 0$$

Karena $\langle x, x \rangle > 0$,

$$\lambda = \bar{\lambda}$$

Jadi, λ nyata.

Teorema 3.3 Setiap matriks $n \times n$ simetris nyata memiliki himpunan n vektor eigen yang saling orthogonal.

Bukti:

Buktikan hasil untuk kasus ketika nilai eigen n berbeda. Misalkan $Av_1 = \lambda_1 v_1, Av_2 = \lambda_2 v_2$ dimana $\lambda_1 \neq \lambda_2$. Karena $A = A^T$, maka

$$\langle Av_1, v_2 \rangle = \langle v_1, A^T v_2 \rangle = \langle v_1, Av_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle$$

Oleh karena itu

$$\lambda_1 \langle v_1, v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle$$

Karena $\lambda_1 \neq \lambda_2$, maka

$$\langle v_1, v_2 \rangle = 0$$

Jika A simetris, maka himpunan vektor eigen membentuk basis orthogonal untuk \mathbb{R}^n . Jika basis $\{v_1, v_2, \dots, v_n\}$ dinormalisasi maka setiap anggota memiliki norma kesatuan

$$T = [v_1, v_2, \dots, v_n]$$

Kita punya

$$T^T T = I$$
$$T^T = T^{-1}$$

Setiap kebalikan transpose matriks disebut orthogonal matriks.

3.3 Ortogonal Proyeksi

Subruang \mathcal{V} dari \mathbb{R}^n adalah himpunan bagian yang tertutup dibawah operasi penjumlahan vektor dan perkalian scalar. Dengan kata lain, \mathcal{V} adalah subruang dari \mathbb{R}^n jika $x_1, x_2 \in \mathcal{V} \Rightarrow \alpha x_1 + \beta x_2 \in \mathcal{V} \forall \alpha, \beta \in \mathbb{R}$. Selanjutnya, dimensi subruang \mathcal{V} sama dengan jumlah maksimum vektor bebas linear di \mathcal{V} . Jika \mathcal{V} adalah subruang \mathbb{R}^n maka komplemen orthogonal \mathcal{V} dinotasikan \mathcal{V}^\perp , terdiri dari semua vektor yang orthogonal untuk setiap vektor di \mathcal{V} .

$$\mathcal{V}^\perp = \{x: v^T x = 0 \forall v \in \mathcal{V}\}$$

pada pembahasan selanjutnya, misalkan $A \in \mathbb{R}^{n \times m}$, range dari A dinotasikan

$$\mathcal{R}(A) \triangleq \{Ax: x \in \mathbb{R}^m\}$$

Dan ruang kosong atau kernel dari A dinotasikan

$$\mathcal{N}(A) \triangleq \{x \in \mathbb{R}^m: Ax = 0\}$$

Teorema 3.4. Misalkan A adalah matriks. Lalu, $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$ dan $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$

Bukti: Misalkan $x \in \mathcal{R}(A)^\perp$. Maka

$$y^T (A^T x) = (Ay)^T x = 0 \forall y$$

Jadi, $A^T x = 0$, karena $x \in \mathcal{N}(A^T)$. Ini mengimplikasikan bahwa

$$\mathcal{R}(A)^\perp \subset \mathcal{N}(A^T)$$

Jika $x \in \mathcal{N}(A^T)$, maka

$$(Ay)^T x = y^T (A^T x) = 0 \forall y$$

Sehingga $x \in \mathcal{R}(A)^\perp$ dan oleh karena itu

$$\mathcal{N}(A^T) \subset \mathcal{R}(A)^\perp$$

Jadi

$$\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$$

Teorema 3.5. Sebuah matriks P adalah orthogonal proyeksi (pada subruang $\mathcal{V} = \mathcal{R}(P)$) jika dan hanya jika $P^2 = P = P^T$

Bukti: Misalkan P adalah orthogonal proyeksi pada subruang $\mathcal{V} = \mathcal{R}(P)$. Maka

$$\mathcal{R}(I - P) \subset \mathcal{R}(P)^\perp$$

$$\mathcal{R}(I - P) \subset \mathcal{N}(P^T)$$

Oleh karena itu

$$P^T(I - P)y = 0, \forall y$$

Dengan mengimplikasikan bahwa

$$P^T(I - P) = 0$$

Dimana, 0 adalah matriks dengan semua elemennya nol. Oleh karena itu $P^T = P^T P$ dengan demikian $P = P^T = P^2$.

Misalkan $P^2 = P = P^T$. Untuk setiap x , kita punya

$$(Py)^T(I - P)x = y^T P^T(I - P)x = y^T(P(I - P)x) = 0 \forall y$$

Jadi

$$(I - P)x \in \mathcal{R}(P)^\perp$$

Yang berarti bahwa P adalah orthogonal proyeksi.

3.4 Bentuk Kuadrat

Bentuk kuadrat $f : \mathbb{R}^n \rightarrow \mathbb{R}$ adalah sebuah fungsi

$$f(x) = x^T Q x$$

Dimana Q adalah matriks nyata $n \times n$. Tanpa menghilangkan generalisasi dalam asumsi Q simetris, sehingga $Q = Q^T$. Jika, matriks Q tidak simetris, maka ubah dengan matriks yang simetris

$$Q_0 = Q_0^T = \frac{1}{2}(Q + Q^T)$$

Catatan bahwa

$$x^T Q x = x^T Q_0 x = x^T \left(\frac{1}{2} Q + \frac{1}{2} Q^T \right) x$$

Sebuah bentuk kuadrat $x^T Q x$, $Q = Q^T$ dikatakan menjadi definit positif jika $x^T Q x > 0$ untuk semua vektor x tak nol. Semidefinite positif jika $x^T Q x \geq 0$ untuk setiap x . Dengan demikian, kita definisikan bentuk kuadrat menjadi definit negative atau semidefinite negative jika $x^T Q x < 0$ untuk setiap vektor x tak nol, atau $x^T Q x \leq 0$ untuk setiap masing-masing x .

Ingat bahwa minor dari matriks Q adalah determinan dari matriks yang diperoleh dengan menghilangkan baris dan kolom secara berurutan dari Q . Artinya, minor utama adalah

$$\det = \begin{bmatrix} q_{i_1 i_1} & q_{i_1 i_2} \cdots & q_{i_1 i_p} \\ q_{i_2 i_1} & q_{i_2 i_2} \cdots & q_{i_2 i_p} \\ \vdots & \vdots & \vdots \\ q_{i_p i_1} & q_{i_p i_2} \cdots & q_{i_p i_p} \end{bmatrix} \quad 1 \leq i_1 < \cdots < i_p \leq n, \quad p = 1, 2, \dots, n$$

$$\Delta_1 = q_{11}, \Delta_2 = \det \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}, \Delta_3 = \det \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}, \dots \Delta_n = \det Q$$

Teorema 3.6 Kriteria Sylvester Sebuah bentuk kuadrat $x^T Q x, Q = Q^T$ adalah definit positif jika dan hanya jika minor utama dari Q positif.

Bukti. Kunci pembuktian kriteria Sylvester adalah fakta bahwa sebuah bentuk kuadrat yang dimana minor utama tak nol dapat dinyatakan dalam beberapa basis sebagai jumlah kuadrat

$$\frac{\Delta_0}{\Delta_1} \tilde{x}_1^2 + \frac{\Delta_1}{\Delta_2} \tilde{x}_2^2 + \cdots + \frac{\Delta_{n-1}}{\Delta_n} \tilde{x}_n^2$$

Dimana \tilde{x} adalah koordinat dari vektor x pada basis baru, $\Delta_0 \triangleq 1$ dan $\Delta_1, \dots, \Delta_n$ adalah minor utama dari Q .

Perhatikan bentuk kuadrat $f(x) = x^T Q x$ dimana $Q = Q^T$. Misalkan $\{e_1, e_2, \dots, e_n\}$ adalah basis natural dari \mathbb{R}^n dan misalkan

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$$

Adalah sebuah vektor di \mathbb{R}^n . Misal $\{v_1, v_2, \dots, v_n\}$ adalah basis lain dari \mathbb{R}^n . Maka, vektor x direpresentasikan dalam basis baru yaitu \tilde{x} , dimana

$$x = [v_1, v_2, \dots, v_n] \tilde{x} \triangleq V \tilde{x}$$

Dengan demikian, bentuk kuadratnya dapat ditulis

$$x^T Q x = \tilde{x}^T V^T Q V \tilde{x} = \tilde{x}^T \tilde{Q} \tilde{x}$$

Dimana

$$\tilde{Q} = V^T Q V = \begin{bmatrix} \tilde{q}_{11} & \cdots & \tilde{q}_{1n} \\ \vdots & \ddots & \vdots \\ \tilde{q}_{n1} & \cdots & \tilde{q}_{nn} \end{bmatrix}$$

Perhatikan bahwa: $\tilde{q}_{ij} = \langle v_i, Q v_j \rangle$.

Selanjutnya, tentukan basis baru $\{v_1, v_2, \dots, v_n\}$ sedemikian rupa sehingga $\tilde{q}_{ij} = 0$ untuk $i \neq j$.

Carilah basis baru dalam bentuk

$v_1 = \alpha_{11}e_1 v_2 = \alpha_{21}e_1 + \alpha_{22}e_2 v_3 = \alpha_{31}e_1 + \alpha_{32}e_2 + \alpha_{33}e_3 v_n = \alpha_{n1}e_1 + \alpha_{n2}e_2 + \dots + \alpha_{nn}e_n$
 Perhatikan bahwa untuk $j = 1, 2, \dots, i - 1$, jika

$$\langle v_i, Qe_j \rangle = 0 = \langle v_i, Qv_j \rangle$$

Selanjutnya tentukan koefisien $\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ii}$ dimana $i = 1, 2, \dots, n$. Sehingga vektor

$$v_i = \alpha_{i1}e_1 + \alpha_{i2}e_2 + \dots + \alpha_{ii}e_i$$

Memenuhi relasi i

$$\langle v_i, Qe_j \rangle = 0, \quad j = 1, 2, \dots, i - 1 \quad \langle v_i, Qv_j \rangle = 1$$

Pada kasus ini, kita punya

$$\tilde{Q} = \begin{bmatrix} \alpha_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \alpha_{nn} \end{bmatrix}$$

Untuk setiap $i = 1, 2, \dots, n$. Relasi i diatas menentukan jumlah $\alpha_{i1}, \dots, \alpha_{ii}e_i$ dengan cara yang khusus. Tentu saja, setelah mengganti pernyataan v_i kedalam persamaan diatas kita memperoleh himpunan persamaan

$$\begin{aligned} \alpha_{i1}q_{11} + \alpha_{i2}q_{12} + \dots + \alpha_{ii}q_{1i} &= 0 \\ \vdots \quad \alpha_{i1}q_{i-1,1} + \alpha_{i2}q_{i-1,2} + \dots + \alpha_{ii}q_{i-1,i} &= 0 \quad \alpha_{i1}q_{i1} + \alpha_{i2}q_{i2} + \dots + \alpha_{ii}q_{ii} \\ &= 1 \end{aligned}$$

Himpunan persamaan diatas dapat dinyatakan dalam bentuk matriks

$$\begin{bmatrix} q_{11} & q_{12} & \dots & q_{1i} \\ \vdots & q_{22} & \dots & q_{2i} \\ \vdots & \dots & \ddots & \vdots \\ q_{i1} & q_{i2} & \dots & q_{ii} \end{bmatrix} \begin{bmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{ii} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Jika minor utama dari matriks Q tidak hilang, maka koefisien α_{ij} dapat diperoleh dengan menggunakan aturan Cramer. Khususnya

$$\alpha_{ii} = \frac{1}{\Delta_i} \det \begin{bmatrix} q_{11} & \dots & q_{1, i-1} & 0 \\ \vdots & \ddots & \vdots & 0 \\ q_{i-1,1} & \dots & q_{i-1, i-1} & 0 \\ q_{i1} & \dots & q_{i, i-1} & 1 \end{bmatrix}$$

Karenanya

$$\tilde{Q} = \begin{bmatrix} 1 & & & & & 0 \\ \Delta_1 & & & & & \\ & \frac{\Delta_1}{\Delta_2} & & & & \\ & & \ddots & & & \\ 0 & & & \frac{\Delta_{n-1}}{\Delta_n} & & \end{bmatrix}$$

Dalam basis baru, bentuk kuadrat dapat dinyatakan sebagai jumlah kuadrat

$$x^T Q x = \tilde{x}^T \tilde{Q} \tilde{x} = \frac{1}{\Delta_1} \tilde{x}_1^2 + \frac{\Delta_1}{\Delta_2} \tilde{x}_2^2 + \dots + \frac{\Delta_{n-1}}{\Delta_n} \tilde{x}_n^2$$

Kondisi perlu dan cukup untuk bentuk kuadrat menjadi definit positif adalah $\Delta_i > 0$, $i = 1, \dots, n$. Syarat cukup jelas, jika $\Delta_i > 0$, $i = 1, \dots, n$ maka dengan argumen sebelumnya, sebuah basis

$$x^T Q x = \tilde{x}^T \tilde{Q} \tilde{x} > 0$$

Untuk setiap $x \neq 0$ (atau ekuivalen $\tilde{x} \neq 0$).

Untuk membuktikannya, pertama-tama tunjukkan bahwa untuk $i = 1, \dots, n$ kita punya $\Delta_i \neq 0$.

Misalkan $\Delta_k = 0$ untuk k. Perhatikan bahwa $\Delta_k = \det Q_k$

$$Q_k = \begin{bmatrix} q_{11} & \dots & q_{1k} \\ \vdots & \ddots & \vdots \\ q_{k1} & \dots & q_{kk} \end{bmatrix}$$

Lalu, terdapat vektor $v \in \mathbb{R}^k, v \neq 0$, sedemikian rupa sehingga $v^T Q_k = 0$. Misalkan $x \in \mathbb{R}^n$ diberikan oleh $x = [v^T, 0^T]^T$. Sehingga

$$x^T Q x = v^T Q_k v = 0$$

Tetapi $x \neq 0$, yang mana kontradiktif dengan bentuk kuadrat f adalah definit positif. Oleh karena itu, jika $x^T Q x > 0$ maka $\Delta_i \neq 0$, $i = 1, \dots, n$. Maka gunakan argument sebelumnya, sehingga kita punya

$$x^T Q x = \tilde{x}^T \tilde{Q} \tilde{x} = \frac{1}{\Delta_1} \tilde{x}_1^2 + \frac{\Delta_1}{\Delta_2} \tilde{x}_2^2 + \dots + \frac{\Delta_{n-1}}{\Delta_n} \tilde{x}_n^2$$

Dimana $\tilde{x} = [v_1, v_2, \dots, v_n] x$

Oleh sebab itu, jika bentuk kuadrat adalah definit positif, maka semua minor utama haruslah positif.

Perhatikan, jika Q tidak simetris, maka kriteria Sylvester tidak bisa digunakan untuk mengecek definit positif dari bentuk kuadrat $x^T Q x$. Perhatikan contoh berikut

$$Q = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix}$$

Minor utama dari Q adalah

$$\Delta_1 = 1 > 0 \Delta_2 = \det Q = 1 > 0$$

Akan tetapi, jika $x = [1, 1]^T$, maka $x^T Q x = -2 < 0$ dan karena bentuk kuadrat tidak pasti definit positif. Perhatikan bahwa

$$x^T Q x = x^T \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} x = \frac{1}{2} x^T \left(\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix} \right) x = x^T \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} x = x^T Q_0 x$$

Minor utama dari Q_0 adalah $\Delta_1 = 1 > 0$ dan $\Delta_2 = \det Q_0 = -3 < 0$.

Teorema 3.7 suatu matriks simetris Q merupakan definit positif (atau semidefinite positif) jika dan hanya jika semua nilai eigen dari Q adalah positif.

Bukti. Untuk setiap x diberikan

$$\begin{aligned} y &= T^{-1}x \\ &= T^T x \end{aligned}$$

Dimana T adalah orthogonal matriks yang mana kolom-kolomnya merupakan vektor eigen dari Q . Maka,

$$x^T Q x = y^T T^T Q T y = \sum_{i=1}^n \lambda_i y_i^2$$

Dengan diagonalisasi, dapat ditunjukkan bahwa semidefinite positif simetris matriks Q mempunyai semidefinite positif (simetris) akar kuadrat $Q^{\frac{1}{2}}$ memenuhi

$$Q^{\frac{1}{2}} Q^{\frac{1}{2}} = Q$$

Selanjutnya, gunakan T diatas dan didefinisikan

$$Q^{\frac{1}{2}} = \begin{bmatrix} \lambda_1^{\frac{1}{2}} & & 0 \\ & \lambda_2^{\frac{1}{2}} & \\ & & \ddots \\ 0 & & & \lambda_n^{\frac{1}{2}} \end{bmatrix} T^T$$

Yang mana dengan mudah diverifikasi sifat-sifat yang diinginkan. Perhatikan bentuk kuadrat $x^T Q x$ dapat dinyatakan sebagai $\|Q^{\frac{1}{2}} x\|^2$.

3.5 Norma Matriks

Norma dari sebuah matriks dapat dipilih dalam berbagai cara. Karena, himpunan matriks $\mathbb{R}^{m \times n}$ dapat dilihat sebagai ruang vektor nyata \mathbb{R}^{mn} . Norma matriks seharusnya tidak berbeda dengan norma vektor biasa. Namun, definisikan norma dari matriks A yang dinotasikan sebagai $\|A\|$ menjadi sebuah fungsi apapun $\|\cdot\|$ yang memenuhi kondisi:

- 1) $\|A\| > 0$ jika $A \neq O$ dan $\|O\| = 0$, dimana O adalah matriks dengan semua nilainya sama dengan nol
- 2) $\|cA\| = |c|\|A\|$, untuk setiap $c \in \mathbb{R}$
- 3) $\|A + B\| \leq \|A\| + \|B\|$

Contoh norma matriks adalah norma Frobenius, didefinisikan sebagai

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n (a_{ij})^2 \right)^{\frac{1}{2}}$$

Dimana $A \in \mathbb{R}^{m \times n}$. Perhatikan bahwa, norma Frobenius ekuivalen dengan norma Euclidean pada $\mathbb{R}^{m \times n}$.

$$4) \|AB\| \leq \|A\|\|B\|$$

Misalkan, $\|\cdot\|_n$ dan $\|\cdot\|_m$ adalah norma vektor masing-masing \mathbb{R}^n dan \mathbb{R}^m . Kita katakan bahwa norma matriks diinduksi oleh, atau sesuai dengan norma vektor yang diberikan jika untuk sembarang matriks $A \in \mathbb{R}^{m \times n}$ dan setiap vektor $x \in \mathbb{R}^n$, pertidaksamaan berikut dipenuhi:

$$\|Ax\|_m \leq \|A\|\|x\|_n$$

Kita dapat mendefinisikan norma matriks induksi sebagai

$$\|A\| = \max_{\|x\|_n=1} \|Ax\|_m$$

Norma yang diinduksi memenuhi kondisi 1 – 4 dan kondisi kompatibilitas, bukti:

Bukti kondisi ke-1

Misalkan $A \neq O$, maka vektor x , $\|x\| = 1$ dapat dicari sedemikian sehingga $Ax \neq 0$, sehingga $\|Ax\| \neq 0$. Oleh karena itu,

$$\|A\| = \max_{\|x\|=1} \|Ax\| \neq 0$$

Jika $A = O$, maka

$$\|A\| = \max_{\|x\|=1} \|O\| = 0$$

Bukti kondisi ke-2

Dengan definisi

$$\|cAx\| = \max_{\|x\|=1} \|cAx\|$$

Maka $\|cAx\| = |c|\|Ax\|$ dan oleh karena itu

$$\|cAx\| = \max_{\|x\|=1} |c|\|Ax\| = |c| \max_{\|x\|=1} \|Ax\| = |c|\|A\|$$

Bukti Kondisi Kompatibilitas

Misalkan $y \neq 0$ divektor apapun. Maka, $x = \frac{y}{\|y\|}$ memenuhi kondisi $\|x\| = 1$. Oleh karena itu

$$\|Ay\| = \|A(\|y\|x)\| = \|y\|\|Ax\| \leq \|y\|\|A\|$$

Bukti kondisi ke-3

Untuk matriks $A + B$, dapat dicari vektor x_0 sehingga $\|A + B\| = \|(A + B)x_0\|$ dan $\|x_0\| = 1$.

Maka kita punya

$$\begin{aligned} \|A + B\| &= \|(A + B)x_0\| = \|Ax_0 + Bx_0\| \leq \|Ax_0\| + \|Bx_0\| \leq \|A\|\|x_0\| + \|B\|\|x_0\| \\ &= \|A\| + \|B\| \end{aligned}$$

Bukti kondisi ke-4

Untuk matriks AB , dapat dicari sebuah vektor x_0 sehingga $\|x_0\| = 1$ dan $\|ABx_0\| = \|AB\|$.

Sehingga kita punya

$$\|AB\| = \|ABx_0\| = \|A(Bx_0)\| \leq \|A\|\|Bx_0\| \leq \|A\|\|B\|\|x_0\| = \|A\|\|B\|$$

Teorema 3.8

Diberikan

$$\|x\| = \left(\sum_{k=1}^n |x_k|^2 \right)^{\frac{1}{2}} = \sqrt{\langle x, x \rangle}$$

Norma matriks induksi oleh norma vektor ini adalah

$$\|A\| = \sqrt{\lambda_1}$$

Dimana λ_1 merupakan nilai eigen terbesar dari matriks $A^T A$

Bukti. Kita punya

$$\|Ax_1\|^2 = \langle Ax, Ax \rangle = \langle x, A^T Ax \rangle$$

Matriks $A^T A$ simetris dan semidefinite positif. Misalkan, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ menjadi nilai eigen dan x_1, x_2, \dots, x_n adalah himpunan orthogonal dari vektor eigen. Ambil vektor arbiter x dengan $\|x\| = 1$ dan sebagai kombinasi linear dari x_i , $i = 1, \dots, n$.

$$x = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

Perhatikan bahwa

$$\langle x, x \rangle = c_1^2 + c_2^2 + \dots + c_n^2 = 1$$

Sehingga

$$\begin{aligned} \|Ax_1\|^2 &= \langle x, A^T Ax \rangle = \langle c_1x_1 + \dots + c_nx_n, c_1\lambda_1x_1 + \dots + c_n\lambda_nx_n \rangle = \lambda_1c_1^2 + \dots + \lambda_nc_n^2 \\ &\leq \lambda_1(c_1^2 + \dots + c_n^2) = \lambda_1 \end{aligned}$$

Untuk vektor eigen x_1 dari $A^T A$ berdasarkan nilai eigen λ_1 , kita punya

$$\|Ax_1\|^2 = \langle x_1, A^T Ax_1 \rangle = \langle x_1, \lambda_1x_1 \rangle = \lambda_1$$

Dan karenanya

$$\max_{\|x\|=1} \|Ax\| = \sqrt{\lambda_1}$$

Pertidaksamaan Rayleigh

Jika suatu $n \times n$ matriks P adalah definit positif simetris maka

$$\lambda_{\min}(P)\|x\|^2 \leq x^T Px \leq \lambda_{\max}(P)\|x\|^2$$

Dimana $\lambda_{\min}(P)$ dinotasikan sebagai nilai eigen terkecil dari P , dan $\lambda_{\max}(P)$ dinotasikan sebagai nilai eigen terbesar P .

Contoh 3.1

Diberikan sebuah matriks

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Dan norma di \mathbb{R}^2

$$\|x\| = \sqrt{x_1^2 + x_2^2}$$

Maka

$$A^T A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}, \text{ dan } \det[\lambda I_2 - A^T A] = \lambda^2 - 10\lambda + 9 = (\lambda - 1)(\lambda - 9)$$

Jadi,

$$\|A\| = \sqrt{9} = 3$$

Vektor eigen dari $A^T A$ berdasarkan $\lambda_1 = 9$ adalah

$$x_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Perhatikan bahwa $\|Ax_1\| = \|A\|$.

$$\|Ax_1\| = \left\| \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = \frac{1}{\sqrt{2}} \left\| \begin{bmatrix} 3 \\ 3 \end{bmatrix} \right\| = \frac{1}{\sqrt{2}} \sqrt{3^2 + 3^2} = 3$$

Karena $A = A^T$ dan

$$\|A\| = \max_{1 \leq i \leq n} |\lambda_i(A)|$$

Dimana $\lambda_1(A), \dots, \lambda_n(A)$ adalah nilai eigen dari A .

Contoh 3.2

Diberikan

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Maka

$$A^T A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Dan

$$\det[\lambda I_2 - A^T A] = \det \begin{bmatrix} \lambda & 0 \\ 0 & \lambda - 1 \end{bmatrix} = \lambda(\lambda - 1)$$

Perhatikan bahwa, 0 adalah hanya nilai eigen dari A . Jadi, untuk $i = 1, 2$

$$\|A\| \Rightarrow |\lambda_i(A)| = 0$$

Latihan Soal 3

3.1. Selesaikan vektor di \mathbb{R}^n dan x adalah kolom koordinat vektor terhadap basis $\{e_1, e_2, \dots, e_n\}$ dan $x^1 \{e_1^1, e_2^1, \dots, e_n^1\}$. Tunjukkan $x^1 = T_x$, dimana T adalah transformasi matriks dari $\{e_1, e_2, \dots, e_n\}$ ke $\{e_1^1, e_2^1, \dots, e_n^1\}$.

Penyelesaian:

Misalkan v adalah vektor sedemikian sehingga x adalah koordinat dari v terhadap $\{e_1, e_2, \dots, e_n\}$ dan x^1 terhadap $\{e_1^1, e_2^1, \dots, e_n^1\}$. Kemudian

$$v = x_1 e_1 + \dots + x_n e_n = [e_1, \dots, e_n] x,$$

dan

$$v = x_1^1 e_1^1 + \dots + x_n^1 e_n^1 = [e_1^1, \dots, e_n^1] x^1$$

Maka $[e_1, \dots, e_n] x = [e_1^1, \dots, e_n^1] x^1$

Sehingga $x^1 = [e_n^1 \dots e_n^1]^{-1} [e_1, \dots, e_n] x = T_x$

- 3.2. Misalkan $\lambda_1, \dots, \lambda_n$ merupakan nilai eigen dari matriks $A \in R^{n \times n}$. Tunjukkan bahwa nilai eigen dari matriks $I_n - A$ adalah $1 - \lambda_1, \dots, 1 - \lambda_n$.

Penyelesaian :

Misalkan v_1, \dots, v_n adalah nilai eigen A yang sesuai $\lambda_1, \dots, \lambda_n$. Kemudian untuk setiap $i = 1, \dots, n$ kita punya $(I_n - A)v_i = v_i - Av_i = v_i - \lambda_i v_i = (1 - \lambda_i)v_i$

Menunjukkan bahwa $1 - \lambda_1, \dots, 1 - \lambda_n$ adalah nilai eigen $I_n - A$.

- 3.3. Misalkan v adalah subruang. Tunjukkan bahwa v^\perp juga subruang.

Penyelesaian:

Misalkan $x, y \in v^\perp$ dan $\alpha, \beta \in R$.

$\alpha x + \beta y \in v^\perp$. Untuk ini, vektor v , kemudian

$$v^T(\alpha x + \beta y) = \alpha v^T x + \beta v^T y = 0$$

Maka $v^T x = v^T y = 0$

- 3.4. Misalkan $A \in R^{m \times n}$ adalah sebuah matriks.

Tunjukkan $R(A)$ adalah subruang dari R^m dan $N(A)$ adalah subruang R^n .

Penyelesaian:

Misalkan $x, y \in R(A)$ dan $\alpha, \beta \in R$. Kemudian terdapat u, v seperti $x = Av$ dan $y = Au$.

$$\alpha x + \beta y = \alpha Av + \beta Au = A(\alpha v + \beta u)$$

Karena $\alpha x + \beta y \in R(A)$, tunjukkan $R(A)$ adalah subruang.

$x, y \in N(A)$ dan $x, y \in R$. kemudian $Ax = 0$ dan $Ay = 0$

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay = 0$$

Dikarenakan $\alpha x + \beta y \in N(A)$. maka $N(A)$ adalah subruang.

- 3.5. Buktikan jika A dan B adalah dua matriks dengan baris m dan $N(A^T)C N(B^T)$, lalu $R(B)CR(A)$. Petunjuk : gunakan bahwa untuk setiap matriks M dengan m baris yang dimiliki $\dim R(M) + \dim N(M^T) = m$

Penyelesaian:

$$v \in R(B), \quad i, e, \dots v = B_x \text{ untuk setiap } x$$

Dengan matriks $[A \ V]$. lalu $N(A^T) = N([A \ U]^T)$. Since $u \in N(A^T)$. then $U \in N(B^T)$, dan karena $U^T v = U^T B_x = x^T B^T U = 0$, maka $\dim R(A) + \dim N(A^T) = m$

Dan

$$\dim R([A \ V]) + \dim N N([A \ V]^T) = m$$

Saat $\dim N(A^T) = \dim N N([A \ V]^T)$, kemudian $\dim R(A) = R([A \ V])$. Dikarenakan v adalah persamaan kombinasi $A, I, e, v \in R(A)$.

3.6. v adalah subruang. Tunjukan $(v^\perp)^\perp = v$. Gunakan soal 3.5.

Penyelesaian:

Tunjukan $v \subset (v^\perp)^\perp$ misalkan $v \in v$ dan U setiap elemen dari v^\perp . Maka $U^T u = V^T u = 0$

Oleh karena itu, $v \in (v^\perp)^\perp$, misalkan $\{a_1, \dots, a_n\}$ baris dari $\{b_1, \dots, b_n\}$ baris dari $(v^\perp)^\perp$ didefinisikan $A = [a_1, \dots, a_k]$ dan $B = [b_1, \dots, b_i]$

Sehingga $v = R(A)$ dan $(v^\perp)^\perp = R(B)$ maka menunjukkan bahwa $R(B) \subset R(A)$. misalkan $x \in N(A^T)$, yang mana berarti $x \in R(A)^\perp = v^\perp$ Karena $R(A)^\perp = N(A^T)$.

Untuk semua y , $(By)^T x = 0 = y^T B^T x$, berarti $B^T x = 0$ sehingga $x \in N(B^T)$.

3.7. Misalkan v dan w adalah subruang. Tunjukan bhwa jika $v \subset w$, maka $w^\perp \subset v^\perp$

Penyelesaian:

Misal $w \in w^\perp$ dan y setiap elemen dari v . karena $v \subset w$ maka $w^\perp y = 0$. sehingga $w \in v^\perp$.

3.8. Misalkan v subruang dari R . Tunjukan bahwa ada matriks v dan v seperti $v = R(v) = N(v)$

Penyelesaian:

Misalkan $r = \dim v$

v_1, \dots, v_r adalah baris untuk v maka v matriks yang ke-I kolom adalah v_i . Maka $v = R(v)$, v_1, \dots, v_n baris v^2 dan v_n matriks baris ke-I adalah u^T . maka $v^2 = R(u^T)$ dan $v = (v^\perp)^\perp = (u^\perp)^T = N(U)$.

3.9. Let P be an orthogonal projector onto a subspace v . Show that.

a. $P_x = x$ for all $x \in v$;

Misalkan $x \in v$, maka $x = P_x + (I - P)x$. Ingat bahwa $P_x x \in v$ dan $(I - P)x \in v^\perp$. $x = P_x + (I - P)x$ adalah dekomposisi ortogonal dari x ke v . Karena dekomposisi ortogonal dari x ke v karena dekomposisi ortogonal unik, maka $x = P_x$.

b. $R(P) = v$

Misalkan P adalah ortogonal proyek dari v . $R(P) \subset v$. Namun, $x = P_x$ untuk semua $x \in v$ dan karenanya $v \subset R(P)$. Maka $R(P) = v$

3.10. Apakah ini bentuk kuadrat

$$x^T \begin{bmatrix} 1 & -8 \\ 1 & 1 \end{bmatrix} x$$

Positif definit, positif semidefinit, negatif definit, negatif semidefinit atau indefinit?

Penyelesaian:

$$\begin{aligned} x^T \left(\frac{1}{2} \begin{bmatrix} 1 & -8 \\ 1 & 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -8 & 1 \end{bmatrix} \right) x \\ = x^T \begin{bmatrix} 1 & -7/2 \\ -7/2 & 1 \end{bmatrix} x \end{aligned}$$

Dengan prinsipal minor adalah $\Delta_1 = 1$ dan $\Delta_2 = -45/4$ sehingga bentuk kuadrat indefinit.

3.11. Misalkan

$$A = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 0 \end{bmatrix}$$

Tunjukkan bahwa meskipun semua minor utama utama dari A adalah nonnegatif A tidak positif setengah pasti.

Penyelesaian:

Dengan principal minor adalah $\Delta_1 = 2$ $\Delta_2 = 0$ $\Delta_3 = 0$, non negatif. Namun, nilai eigen dari A adalah 0, -1.4641, 5.4641. ini berarti bahwa matriks A indefinit. Alternatif solusi lain adalah menunjukkan A bukan positif semidefinit untuk menentukan vektor x sedemikian sehingga $x^T A x < 0$. Misalkan x nilai vektor eigen dari A sesuai nilai eigen negatif.

$$\pi = 1.4641 \text{ maka } x^T A x = x^T(\pi x) = \lambda x^T = \lambda \|x\|^2 < 0 \text{ } x = [0.3251, 0.3251, -0.0881]^T \text{ untuk bisa diverifikasi bahwa } x^T A x = -1.4643.$$

3.12. Perhatikan matriks

$$Q = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

- Apakah matriks definit positif, definit negatif, atau indefinit?
- Apakah matriks ini pasti positif, pasti negatif, atau tidak terbatas pada subruang $M = \{x = x_1 + x_2 + x_3 = 0\}$?

Penyelesaian:

- Matriks Q definit karena $\Delta_2 = 0$ dan $\Delta_3 = 2$
- Misalkan $x \in M$. Maka, $x_2 + x_3 = -x_1$, $x_1 + x_3 = -x_2$ dan $x_1 + x_2 = -x_3$ sehingga.

$$\begin{aligned} x^T Q x &= x_1(x_2 + x_3) + x_2(x_1 + x_3) + x_1(x_1 + x_2) \\ &= -(x_1^2 + x_2^2 + x_3^2) \end{aligned}$$

Berarti matriks Q definit negatif di subruang

3.13. Perhatikan bentuk kuadrat

$$f(x_1, x_2, x_3) = x_1^2 + x_2 + 5x_3^2 + 2\xi x_1 x_2 - 2x_1 x_3 + 4x_2 x_3$$

Nilai parameter yang bentuk kuadratnya pasti positif.

Penyelesaian:

$$Q = \begin{bmatrix} 1 & \xi & -1 \\ \xi & 1 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

Leading principal minor dari $\Delta_1 = 1, \Delta_2 = 1 - \xi^2, \Delta_3 = -\xi^2 4\xi$ untuk bentuk kuadrat positif definit, leading principal minor dari 0 harus positif. Ini harus jika hanya jika $\xi \in (-4/5, 0)$

- 3.14. Perhatikan fungsi $\langle \cdot, \cdot \rangle_Q: R^n \times R^n \rightarrow R$, didefinisikan dengan $\langle x, y \rangle_Q = Q^T Q y$, dimana $x, y \in R^{n \times n}$ adalah matriks definit positif simetris. Tunjukkan bahwa $\langle \cdot, \cdot \rangle_Q$ memenuhi kondisi produk dalam.

Penyelesaian:

Matriks $Q = Q^T > 0$ bisa dilakukan $Q = Q^{1/2} Q^{1/2}$ dimana $Q^{1/2} = (Q^{1/2T})^T > 0$.

$$1. \langle x, y \rangle_Q = (Q^{1/2T} x)^T (Q^{1/2} y) = \| Q^{1/2} x \|^2 \geq 0 \text{ dan}$$

$$\langle x, y \rangle_Q = 0 \Leftrightarrow \| Q^{1/2} x \|^2 = 0$$

$$\Leftrightarrow Q^{1/2} x = 0$$

$$\Leftrightarrow x = 0$$

Karena $Q^{1/2}$ non singular.

$$2. \langle x, y \rangle_Q = Q^T Q Y = Y^T Q^T x = Y^T Q_x = \langle x, y \rangle$$

$$= x^T Q_z + Y^T Q_z$$

$$= \langle x, z \rangle_Q + \langle y, z \rangle_Q$$

$$3. \langle x, y, z \rangle_Q = (x + y)^T Q_z$$

$$4. \langle r_x, y \rangle_Q = (r_x)^T Q Y = r_x^T Q Y = r \langle x, y \rangle_Q$$

- 3.15. Perhatikan norma vektor $\|\cdot\|_{\sim}$ on R^n dimisalkan dengan $\|x\|_{\sim} = \max_i |x_i|$, dimana $x = [x_1, \dots, x_n]^T$.

Demikian pula mendefinisikan norma $\|\cdot\|_{\sim}$ on R^m tunjukkan bahwa norma matriks yang diinduksi oleh norma vektor ini diberikan oleh $\|A\|_{\sim} = \max_i \sum_{k=1}^n |a_{ik}|$,

Dimana a_{ij} adalah elemen ke (i, j) dari $A \in R^{m \times n}$

Penyelesaian:

$$\|A\|_{\sim} = \max\{\|Ax\|_{\sim} : \|x\|_{\sim} = 1\}$$

$$\begin{aligned} \|A\|_{\infty} &= \max_k \left| \sum_{i=1}^n a_{ik} x_k \right| \\ &\leq \max_i \sum_{k=1}^n |a_{ik}| |x_k| \\ &\leq \max_i \sum_{k=1}^n |a_{ik}| \end{aligned}$$

Karena $|x_k| \leq \max_k |x_k| = \|x\|_{\infty} = 1$, maka

$$\|A\|_{\infty} \leq \max_i \sum_{k=1}^n |a_{ik}|$$

Untuk menunjukkan bahwa $\|A\|_{\infty} = \max_i \sum_{k=1}^n |a_{ik}|$, ini berarti untuk menemukan $\bar{x} \in \mathbb{R}^n$, $\|\bar{x}\|_{\infty} = 1$, sedemikian sehingga $\max_k \sum_{i=1}^n |a_{ik}| \bar{x}_k$. Misalkan j:

$$\sum_{k=1}^n |a_{jk}| = \max_i \sum_{k=1}^n |a_{ik}|$$

Definisikan \bar{x} sebagai

$$\bar{x}_k = \begin{cases} |a_{jk}|/a_{jk} & \text{jika } a_{jk} \neq 0 \\ 1 & \text{lainnya} \end{cases}$$

Jelas $\|\bar{x}\|_{\infty} = 1$, lebih lagi untuk $i \neq j$

$$\begin{aligned} \left| \sum_{k=1}^n a_{ik} \bar{x}_k \right| &\leq \sum_{k=1}^n |a_{ik}| \leq \max_i \sum_{k=1}^n |a_{ik}| = \sum_{k=1}^n |a_{jk}| \\ \left| \sum_{k=1}^n a_{jk} \bar{x}_k \right| &= \sum_{k=1}^n |a_{jk}| \end{aligned}$$

Karena itu

$$\|A\bar{x}\|_{\infty} \leq \max_i \left| \sum_{k=1}^n a_{ik} \bar{x}_k \right| = \sum_{k=1}^n |a_{jk}| = \max_i \sum_{k=1}^n |a_{ik}|$$

3.16. Perhatikan norma vektor $\|\cdot\|_1$ on \mathbb{R}^n dimisalkan dengan $\|x\|_1 = \sum_{i=1}^n |x_i|$, dimana $x = [x_1, \dots, x_n]^T$. Sama mendefinisikan norma $\|\cdot\|_1$ on \mathbb{R}^m . Tunjukkan bahwa norma matriks yang diinduksi oleh vektor ini diberikan oleh

$$\|A\|_1 = \max_k \sum_{i=1}^n |a_{ik}|$$

Dimana a_{ij} adalah elemen ke (i, j) dari $A \in R^{m \times n}$

Penyelesaian:

$$\|A\|_1 = \max\{\|A_x\|_1 : \|x\|_1 = 1\}$$

Pertama tunjukkan $\|A\|_1 \leq \max_k \sum_{i=1}^m |a_{ik}|$. Bahwa untuk setiap x sedemikian rupa sehingga

$$\begin{aligned} \|A_x\|_1 &= \sum_{i=1}^m \left| \sum_{k=1}^n a_{ik} x_k \right| \\ &\leq \sum_{i=1}^m \sum_{k=1}^n |a_{ik}| |x_k| \\ &\leq \sum_{k=1}^n |x_k| \sum_{i=1}^m |a_{ik}| \\ &\leq \left(\max_k \sum_{i=1}^m |a_{ik}| \right) \sum_{k=1}^n |x_k| \\ &\leq \max_k \sum_{i=1}^m |a_{ik}| \end{aligned}$$

Ketika $\sum_{k=1}^n |x_k| = \|x\|_1 = 1$. Karena itu

$$\|A\|_1 \leq \max_k \sum_{i=1}^m |a_{ik}|$$

Tunjukkan $\|A\|_1 = \max_k \sum_{i=1}^m |a_{ik}|$, itu tetap untuk menemukan $\bar{x} \in R^m$, $\|\bar{x}\|_1 = 1$, seperti

$\|A_x\|_1 = \max_k \sum_{i=1}^m |a_{ik}|$. Jadi misalkan

$$J = \sum_{i=1}^m |a_{ij}| \max_k \sum_{i=1}^m |a_{ik}|$$

Definisikan \bar{x} sebagai

$$\bar{x} = \begin{cases} 1, & \text{jika } k = j \\ 0, & \text{lainnya} \end{cases}$$

Jelas $\|\bar{x}\| = 1$, lebih lagi

$$\|A\bar{x}\|_1 = \sum_{i=1}^m \left| \sum_{k=1}^n a_{ik} \bar{x}_k \right| = \sum_{i=1}^m |a_{ij}| = \max_k \sum_{i=1}^m |a_{ik}|$$

BAB 4

Konsep Geometri

4.1 Ruang Garis

Dalam analisis berikut, kita hanya berfokus pada \mathbb{R}^n . Unsur-unsur dari ini ruang adalah vektor komponen n , $x = [x_1, x_2, \dots, x_n]^T$.

Ruang garis antara dua titik x dan y di \mathbb{R}^n adalah kumpulan titik pada garis lurus yang terhubung dengan titik x dan y . Perhatikan bahwa jika z terletak di ruang garis antara x dan y , maka

$$z - y = \alpha(x - y)$$

Dimana α merupakan bilangan real dari interval $[0,1]$. Persamaan di atas dapat ditulis ulang sebagai $z = \alpha x + (1 - \alpha)y$. Oleh karena itu, ruang garis antara x dan y dapat ditunjukkan sebagai berikut.

$$\{\alpha x + (1 - \alpha)y : \alpha \in [0,1]\}$$

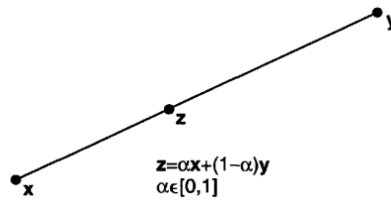
4.2 Hyperplanes dan Varietas linear

Misalkan u_1, u_2, \dots, u_n , $u \in \mathbb{R}$, di mana setidaknya salah satu u_i adalah tak nol. Himpunan dari semua titik $x = [x_1, x_2, \dots, x_n]^T$ yang memenuhi persamaan linear

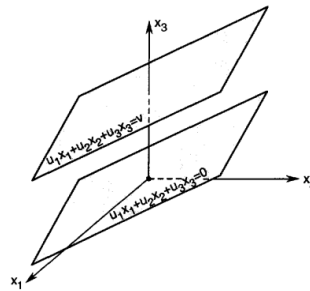
$$u_1 x_1 + u_2 x_2 + \dots + u_n x_n = v$$

Disebut hyperplane dari ruang \mathbb{R}^n . Kita dapat mendefinisikan hyperplane dengan

$$\{x \in \mathbb{R}^n : u^T x = v\},$$



Gambar 4.1 Segmen Garis



Gambar 4.2 Translasi Hyperplane

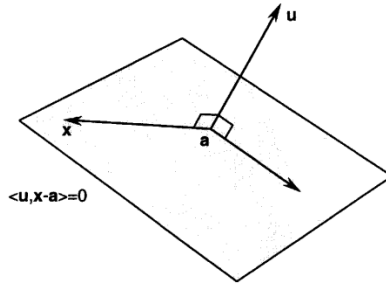
Dimana

$$u = [u_1, u_2, \dots, u_n]^T$$

Hyperplane belum tentu merupakan subruang \mathbb{R}^n , karena, secara umum, itu tidak mengandung keasliannya. Untuk $n = 2$, persamaan hyperplane memiliki bentuk $u_1x_1 + u_2x_2 + \dots + u_kx_k = v$ yang merupakan persamaan dari garis lurus. Dengan demikian, garis lurus adalah hyperplanes di \mathbb{R}^2 . Di \mathbb{R}^3 (ruang tiga dimensi), hyperplanes adalah pesawat biasa. Dengan menerjemahkan hyperplane sehingga mengandung \mathbb{R}^n , menjadi subruang \mathbb{R}^n . Karena dimensi subruang ini adalah $n - 1$, dapat dikatakan bahwa hyperplane memiliki dimensi $n - 1$.

Hyperplane $H = \{x: u_1x_1 + \dots + u_nx_n = v\}$ membagi \mathbb{R}^n menjadi dua ruang setengah. Salah satu setengah ruang ini terdiri dari titik-titik yang memenuhi pertidaksamaan $u_1x_1 + u_2x_2 + \dots + u_nx_n \geq v$, dilambangkan

$$H_+ = \{x \in \mathbb{R}^n: u^T x \geq v\},$$



Gambar 4.3 Hyperplane $H = \{x \in \mathbb{R}^n : u^T(x - a) = 0\}$

Dimana

$$u = [u_1, u_2, \dots, u_n]^T$$

Setengah ruang lainnya terdiri dari titik-titik yang memenuhi pertidaksamaan $u_1x_1 + u_2x_2 + \dots + u_nx_n \leq v$, dilambangkan

$$H_- = \{x \in \mathbb{R}^n : u^T x \leq v\}.$$

Setengah ruang H_+ disebut setengah ruang positif dan setengah ruang H_- disebut setengah ruang negative.

Missal $a = [a_1, a_2, \dots, a_n]^T$ merupakan titik hyperplane yang sering berubah-ubah. Maka $u^T a - v = 0$. dapat dituliskan sebagai berikut.

$$\begin{aligned} u^T x - v &= u^T x - v - (u^T a - v) \\ &= u^T(x - a) \\ &= u_1(x_1 - a_1) + u_2(x_2 - a_2) + \dots + u_n(x_n - a_n) = 0. \end{aligned}$$

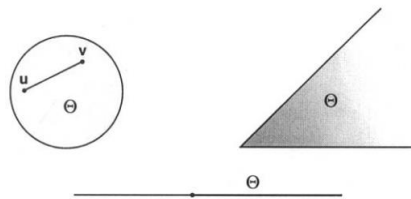
nilai $(x_i - a_i, i = 1, \dots, n)$ adalah komponen vektor $x - a$. Oleh karena itu, hyperplane H terdiri dari titik x dimana $\langle u, x - a \rangle = 0$. Dengan kata lain, hyperplane H terdiri dari titik x dimana vektor u dan $x - a$ adalah ortogonal (lihat Gambar 4.3). Dapat disebut vektor u normal ke hyperplane H. Himpunan H_+ terdiri dari titik-titik x untuk $\langle u, x - a \rangle \geq 0$, dan

H_- terdiri dari titik-titik x untuk $\langle u, x - a \rangle \leq 0$

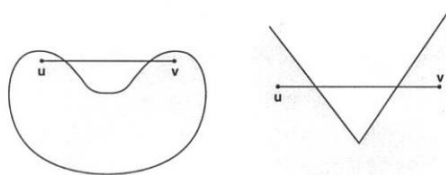
Variasi linier adalah satu himpunan yang berbentuk

$$\{x \in \mathbb{R}^n : Ax = b\}$$

untuk beberapa matriks $x \in \mathbb{R}^{m \times n}$ dan vektor $b \in \mathbb{R}^n$. Jika $\dim \mathcal{N}(A) = r$, dapat dikatakan bahwa variasi linear memiliki dimensi r . Variasi linier adalah subruang jika dan hanya jika $b = 0$. Jika $A = O$, variasi linier adalah \mathbb{R}^n . Jika dimensi varietas linier kurang dari n , maka itu adalah persimpangan dari sejumlah hyperplan yang terbatas.



Gambar 4.4 Himpunan Cembung



Gambar 4.5 Himpunan yang tidak cembung

4.3 Himpunan Cembung

Ingat bahwa segmen garis antara dua titik $u, v \in \mathbb{R}^n$ adalah himpunan $\{w \in \mathbb{R}^n : w = \alpha u + (1 - \alpha)v, \alpha \in [0,1]\}$. Titik $w = \alpha u + (1 - \alpha)v$ (di mana $\alpha \in [0,1]$) disebut kombinasi cembung dari titik-titik u dan v .

Himpunan $\theta \in \mathbb{R}^n$ cembung jika untuk semua $u, v \in \theta$ segmen garis antara u dan v adalah dalam θ . Gambar 4.4 memberikan contoh himpunan cembung, sedangkan Gambar 4.5 memberikan contoh himpunan yang tidak cembung. Perhatikan bahwa θ cembung jika dan hanya jika $\alpha u + (1 - \alpha)v \in \theta$ untuk semua $u, v \in \theta$ dan $\alpha \in (0,1)$.

Contoh himpunan cembung meliputi:

- Himpunan kosong
- Himpunan yang terdiri dari satu titik
- Garis atau segmen garis

- Subruang
- Hyperplane
- Variasi linier
- setengah ruang
- \mathbb{R}^n

Teorema 4.1 Subset cembung dari \mathbb{R}^n memiliki sifat sebagai berikut:

- a) Jika Q adalah himpunan cembung dan β adalah bilangan real, maka himpunan

$$\beta\theta = \{x: x = \beta v, v \in \theta\}$$

juga cembung;

- b) Jika θ_1 dan θ_2 adalah himpunan cembung, maka himpunan

$$\theta_1 + \theta_2 = \{x: x = v_1 + v_2, v_1 \in \theta_1, v_2 \in \theta_2\}$$

juga cembung;

- c) Persimpangan dari setiap koleksi himpunan cembung adalah cembung (lihat Gambar 4.6 untuk ilustrasi hasil dua himpunan).

Bukti :

- a) $\beta_1 v_1, \beta_2 v_2 \in \beta\theta$, dimana $v_1, v_2 \in \theta$. Karena θ cembung, diperoleh $\alpha v_1 + (1 - \alpha)v_2 \in \theta$ untuk setiap $\alpha \in (0,1)$. Oleh karena itu

$$\alpha\beta v_1 + (1 - \alpha)\beta v_2 = \beta(\alpha v_1 + (1 - \alpha)v_2) \in \beta\theta$$

Dimana $\beta\theta$ cembung.

- b) Misal $v_1, v_2 \in \theta_1 + \theta_2$. Maka $v_1 = v'_1 + v''_1$ dan $v_2 = v'_2 + v''_2$ dimana $v'_1, v'_2 \in \theta_1$ dan $v''_1, v''_2 \in \theta_2$. Karena θ_1 dan θ_2 cembung, untuk setiap $\alpha \in (0,1)$,

$$x_1 = \alpha v'_1 + (1 - \alpha)v'_2 \in \theta_1$$

Dan

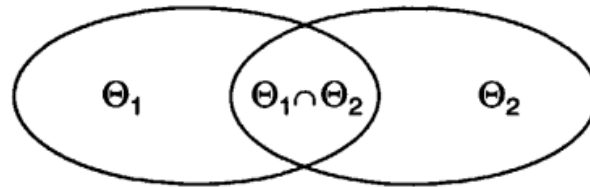
$$x_2 = \alpha v''_1 + (1 - \alpha)v''_2 \in \theta_2$$

Berdasarkan definisi $\theta_1 + \theta_2$, $x_1 + x_2 \in \theta_1 + \theta_2$. Maka

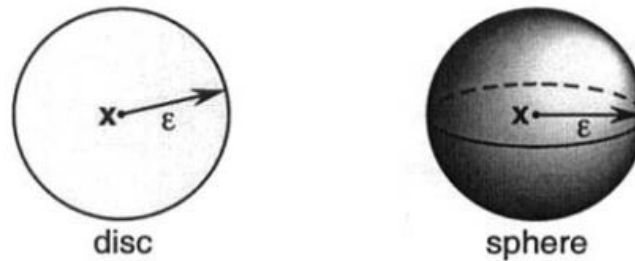
$$= \alpha v_1 + (1 - \alpha)v_2 = \alpha(v'_1 + v''_1) + (1 - \alpha)(v'_2 + v''_2) = x_1 + x_2 \in \theta_1 + \theta_2$$

Dimana $\theta_1 + \theta_2$ cembung.

- c) Misal C koleksi himpunan cembung. Misalkan $x_1, x_2 \in \bigcap_{\theta \in C} \theta$ (di mana $\bigcap_{\theta \in C} \theta$ mewakili persimpangan semua elemen dalam C). Kemudian, $x_1, x_2 \in \theta$ untuk masing-masing $\theta \in C$. Karena setiap $\theta \in C$ cembung, $\alpha x_1 + (1 - \alpha)x_2 \in \theta$ untuk semua $\alpha \in (0,1)$ dan masing-masing $\theta \in C$. Dengan demikian, $\alpha x_1 + (1 - \alpha)x_2 \in \bigcap_{\theta \in C} \theta$.



Gambar 4.6 Persimpangan dua himpunan cembung



Gambar 4.7 contoh lingkungan pada titik di \mathbb{R}^2 dan \mathbb{R}^3

Titik x dalam himpunan cembung θ dikatakan sebagai titik ekstrim θ jika tidak ada dua titik yang berbeda u dan v dalam θ sedemikian sehingga $x = \alpha u + (1 - \alpha)v$ untuk beberapa $\alpha \in (0,1)$. Contoh, pada Gambar 4.4, setiap titik pada batas cakram adalah titik ekstrim, simpul (sudut) dari himpunan di sebelah kanan adalah titik ekstrim, dan titik akhir dari setengah garis juga merupakan titik ekstrim.

4.4 Lingkungan

lingkungan titik $x \in \mathbb{R}^n$ adalah himpunan

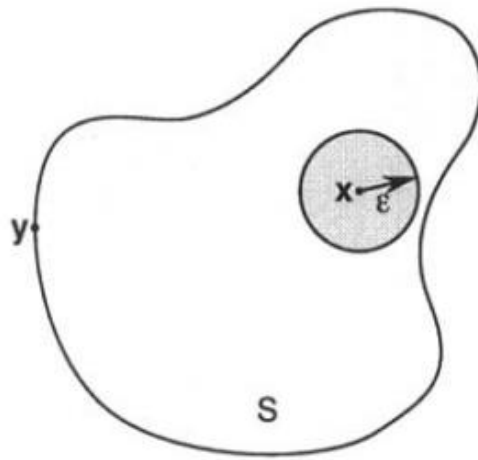
$$\{y \in \mathbb{R}^n : \|y - x\| < \varepsilon\},$$

di mana ε adalah angka positif. Lingkungan ini juga disebut bola dengan radius ε dan pusat x .

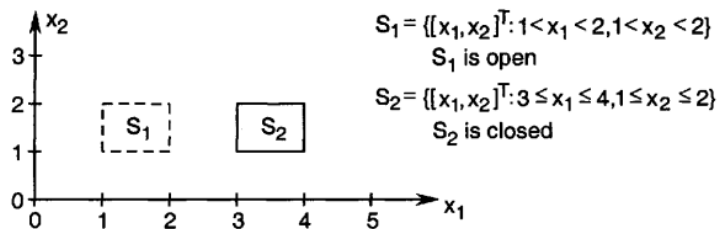
Di ruang \mathbb{R}^2 , lingkungan $x = [x_1, x_2]^T$ terdiri dari semua titik di dalam cakram yang berpusat pada x . Dalam \mathbb{R}^3 , lingkungan $x = [x_1, x_2, x_3]^T$ terdiri dari semua titik-titik di dalam bola yang berpusat pada x (lihat Gambar 4.7).

Titik $x \in S$ dikatakan sebagai titik interior himpunan S jika himpunan S memuat beberapa lingkungan x , yaitu, jika semua titik dalam beberapa lingkungan x juga berada di S . (lihat Gambar 4.8). Himpunan semua titik interior S disebut interior S .

Titik x dikatakan sebagai titik batas dari himpunan S jika setiap lingkungan d x berisi titik di S dan titik tidak dalam S (lihat Gambar 4.8). Perhatikan bahwa batas titik S mungkin atau tidak mungkin elemen dari S . Himpunan semua titik batas S disebut batas dari S .



Gambar 4.8 x merupakan titik interior, sedangkan y merupakan titik batas



Gambar 4.9 Himpunan terbuka dan tertutup

Himpunan S dikatakan terbuka jika memuat lingkungan dari masing-masing titiknya, dimana, jika masing-masing titiknya adalah titik interior, ekuivalen dengan, jika S tidak

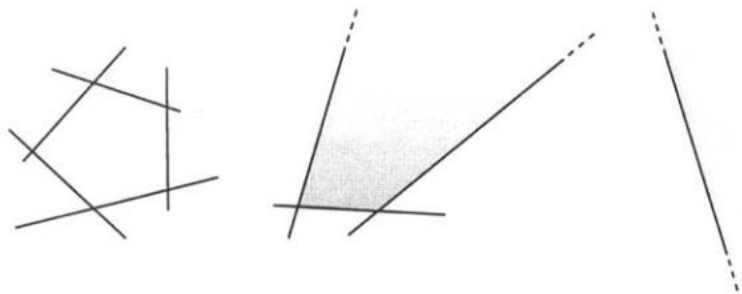
mengandung titik batas. Himpunan S dikatakan tertutup jika memuat batasnya (lihat Gambar 4.9). Kita bisa menunjukkan bahwa sebuah himpunan tertutup jika dan hanya jika komplementnya terbuka.

Himpunan yang memuat bola dalam radius terbatas dikatakan dibatasi. Suatu himpunan dikatakan rapat kompak jika tertutup dan dibatasi. Himpunan rapat penting dalam masalah optimasi untuk alasan berikut.

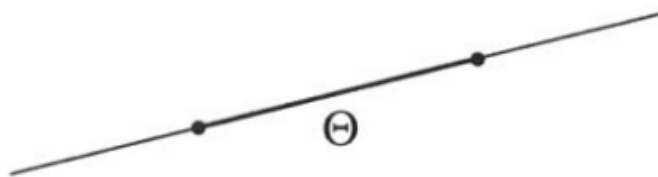
Teorema 4.2 Teorema Weierstrass. Misal $f: \Omega \rightarrow \mathbb{R}$ merupakan fungsi kontinu dimana $\Omega \subset \mathbb{R}^n$ merupakan himpunan rapat. Maka terdapat $x_0 \in \Omega$ dimana $f(x_0) \leq f(x)$ untuk setiap $x \in \Omega$. Dengan kata lain, f mencapai minimum pada Ω .

4.5 Polytopes dan Polyhedra

Misalkan θ merupakan himpunan cembung, dan misalkan y adalah titik batas θ . Sebuah hyperplane melalui y disebut hyperplane dukungan (atau hyperplane pendukung) dari himpunan θ . Jika seluruh himpunan θ terletak sepenuhnya di salah satu dari dua setengah ruang di mana hyperplane membagi ruang \mathbb{R}^n .



Gambar 4.10 Polytopes



Gambar 4.11 Polyhedron dimensi satu

Ingatlah pada Teorema 4.1, persimpangan sejumlah himpunan cembung adalah cembung. Oleh karena itu, kita berfokus pada persimpangan jumlah terbatas setengah ruang. Karena setiap

setengah ruang H_+ atau H_- cembung di \mathbb{R}^n , persimpangan dari setengah ruang adalah himpunan cembung.

Himpunan dapat dinyatakan sebagai persimpangan dari jumlah terbatas setengah ruang disebut cembung (lihat Gambar 4.10). Polytope tanpa batas yang tidak kosong disebut polyhedron (lihat Gambar 4.11). Untuk setiap polyhedron cembung $\theta \in \mathbb{R}^n$, terdapat bilangan bulat nonnegatif $k \leq n$ sehingga θ terkandung dalam berbagai variasi linear dimensi k , tetapi tidak sepenuhnya terkandung dalam setiap variasi linear dimensi $(k - 1)$ dalam \mathbb{R}^n . Selain itu, terdapat hanya satu variasi linear dimensi k yang mengandung θ , yang disebut pembawa polyhedron θ , dan k disebut dimensi θ . Sebagai contoh, polyhedron dimensi nol adalah titik pada \mathbb{R}^n , dan pembawanya adalah dirinya sendiri. Polyhedron satu dimensi adalah segmen, dan pembawanya adalah garis lurus di mana ia berada. Batas dari setiap polyhedron dimensi k , $k > 0$, terdiri dari jumlah terbatas polyhedra dimensi $(k - 1)$. Misalnya, batas polyhedron dimensi satu terdiri dari dua titik yang merupakan titik akhir segmen.

Polyhedra dimensi $(k - 1)$ membentuk batas polyhedron dimensi k yang disebut polyhedron muka. Masing-masing muka ini, pada gilirannya, dimensi muka $(k-2)$. Kami juga mempertimbangkan dimensi muka $(k - 2)$ menjadi muka dari polyhedron dimensi k asli. Dengan demikian, setiap polyhedron dimensi k memiliki dimensi muka $k - 1, k - 2, \dots, 1, 0$. Polyhedron dimensi muka nol disebut simpul, dan polyhedron dimensi muka satu disebut tepi.

Latihan Soal 4

4.1. Tunjukkan bahwa set $S \subset \mathbb{R}^n$ adalah linear variasi jika dan hanya jika untuk semua $x, y \in S$ dan $\alpha \in \mathbb{R}$, dimana memiliki $\alpha x + (1 - \alpha)y \in S$.

Penyelesaian:

Misalkan $S = \{x: Ax = b\}$ menjadi variasi linear. Misalkan $x, y \in S$ dan $\alpha \in \mathbb{R}$. Maka,

$$A(\alpha x + (1 - \alpha)y) = \alpha Ax + (1 - \alpha)Ay = \alpha b + (1 - \alpha)b = b.$$

Jadi, $\alpha x + (1 - \alpha)y \in S$.

Jika S kosong, kita sudah selesai. Jadi, dimisalkan $x_0 \in S$. Pertimbangkanlah set $S_0 = S - x_0 = \{x - x_0: x \in S\}$. Jelas, untuk semua $x, y \in S_0$ dan $\alpha \in \mathbb{R}$, memiliki $\alpha x +$

$(1 - \alpha)y \in S_0$. Perhatikan bahwa $0 \in S_0$ dan S_0 merupakan subruang. Misalkan semua $x, y \in S_0$ dan $\alpha \in \mathbb{R}$. Kemudian, $\alpha x = \alpha x + (1 - \alpha)0 \in S_0$. Selanjutnya, $\frac{1}{2}x + \frac{1}{2}y \in S_0$ dan $x + y \in S_0$ dengan argument sebelumnya. Maka, S_0 merupakan subruang. Jadi, dengan latihan 3.8, terdapat A sedemikian rupa sehingga $S_0 = \mathcal{N}(A) = \{x: Ax = 0\}$. Definisikan $b = Ax_0$. Maka,

$$\begin{aligned} S &= S_0 + x_0 = \{y + x_0: y \in \mathcal{N}(A)\} \\ &= \{y + x_0: Ay = 0\} \\ &= \{y + x_0: A(y + x_0) = b\} \\ &= \{x : Ax = b\}. \end{aligned}$$

4.2. Tunjukkan bahwa set $\{x \in \mathbb{R}^n: \|x\| \leq r\}$ adalah convex, dimana $r > 0$ yang diberikan bilangan real, dan $\|x\| = \sqrt{x^T x}$ adalah Euclidean norm $x \in \mathbb{R}^n$.

Penyelesaian:

Misalkan $u, v \in \theta = \{x \in \mathbb{R}^n: \|x\| \leq r\}$, dan $\alpha \in [0,1]$. Misalkan $z = \alpha u + (1 - \alpha)v$. Untuk menunjukkan bahwa θ adalah convex, sebelumnya perlu ditunjukkan bahwa $z \in \theta$. i. e., $\|z\| \leq r$.

$$\begin{aligned} \|z\|^2 &= (\alpha u^T + (1 - \alpha)v^T)(\alpha u + (1 - \alpha)v) \\ &= \alpha^2 \|u\|^2 + 2\alpha(1 - \alpha)u^T v + (1 - \alpha)^2 \|v\|^2. \end{aligned}$$

Saat $u, v \in \theta$, maka $\|u\|^2 \leq r^2$ dan $\|v\|^2 \leq r^2$. Selanjutnya, dengan pertidaksamaan **Cauchy-Schwarz** memiliki $u^T v \leq \|u\| \|v\| \leq r^2$. Jadi,

$$\|z\|^2 \leq \alpha^2 r^2 + 2\alpha(1 - \alpha)r^2 + (1 - \alpha)^2 r^2 = r^2.$$

4.3. Tunjukkan bahwa terdapat $A \in \mathbb{R}^{m \times n}$ dan vector $b \in \mathbb{R}^m$, dalam set (variasi linear) $\{x \in \mathbb{R}^n: Ax = b\}$ adalah convex.

Penyelesaian:

Misalkan $u, v \in \theta = \{x \in \mathbb{R}^n : Ax = b\}$, dan $\alpha \in [0, 1]$. Misalkan $z = \alpha u + (1 - \alpha)v$. Untuk menunjukkan bahwa θ adalah convex, sebelumnya perlu ditunjukkan bahwa $z \in \theta$. i. e., $Az = b$. Jadi,

$$\begin{aligned} Az &= A(\alpha u + (1 - \alpha)v) \\ &= \alpha Au + (1 - \alpha)Av. \end{aligned}$$

Saat $u, v \in \theta$, maka $Au = b$ dan $Av = b$. Jadi,

$$Az = \alpha b + (1 - \alpha)b = b$$

4.4. Tunjukkan bahwa set $\{x \in \mathbb{R}^n : x \geq 0\}$ adalah convex (dimana $x \geq 0$ berarti setiap komponen x adalah non-negatif).

Penyelesaian:

Misalkan $u, v \in \theta = \{x \in \mathbb{R}^n : x \geq 0\}$, dan $\alpha \in [0, 1]$. Misalkan $z = \alpha u + (1 - \alpha)v$. Untuk menunjukkan bahwa θ adalah convex, sebelumnya perlu ditunjukkan bahwa $z \in \theta$. i. e., $z \geq 0$. Jadi, $x = [x_1, x_2, \dots, x_n]^T$, $y = [y_1, y_2, \dots, y_n]^T$, dan $z = [z_1, z_2, \dots, z_n]^T$. Maka, $z_i = \alpha x_i + (1 - \alpha)y_i$, $i = 1, 2, \dots, n$. Saat $x_i, y_i \geq 0$ dan $\alpha, 1 - \alpha \geq 0$, kita memiliki $z_i \geq 0$.

BAB 5

Elemen Kalkulus

5.1 Baris dan Limit

Barisan bilangan real adalah fungsi yang domainnya adalah himpunan bilangan asli $1, 2, \dots, k, \dots$ dan yang jangkauannya terkandung dalam \mathbb{R} . Jadi, barisan bilangan real dapat dilihat sebagai himpunan bilangan $\{x_1, x_2, \dots, x_3, \dots\}$, yang sering juga dilambangkan sebagai $\{x_k\}$ (atau kadang-kadang sebagai $\{x_k\}_{k=1}^{\infty}$, untuk menunjukkan secara eksplisit rentang nilai yang k dapat mengambil).

Barisan $\{x_k\}$ naik jika $x_1 < x_2 < \dots < x_k \dots$, yaitu $x_k < x_{k+1}$ untuk semua k . Jika $x_k < x_{k+1}$, maka dikatakan barisan tersebut tidak menurun. Demikian pula, kita dapat menentukan barisan turun dan tidak naik. Tidak naik atau tidak turun barisan disebut barisan monoton.

Suatu bilangan $x^* \in \mathbb{R}$ disebut limit barisan $\{x_k\}$ jika untuk sembarang ε ada bilangan K (yang mungkin bergantung pada ε) sehingga untuk semua $k > K$, $|x_k - x^*| < \varepsilon$; yaitu, x_k terletak di antara $x^* - \varepsilon$ dan $x^* + \varepsilon$ untuk semua $k > K$. Dalam hal ini, kita tulis

$$x^* = \lim_{k \rightarrow \infty} x_k$$

Atau

$$x_k \rightarrow x^*$$

Barisan yang memiliki limit disebut barisan konvergen.

Pengertian barisan dapat diperluas menjadi barisan dengan elemen-elemen dalam \mathbb{R}^n . Secara khusus, barisan dalam \mathbb{R}^n adalah fungsi yang domainnya adalah himpunan bilangan asli $1, 2, \dots, k, \dots$ dan yang jangkauannya terkandung dalam \mathbb{R}^n . Kami menggunakan notasi $\{x^{(1)}, x^{(2)}, \dots\}$ atau $\{x^{(k)}\}$ untuk barisan dalam \mathbb{R}^n . Untuk limit barisan dalam \mathbb{R}^n , kami perlu mengganti nilai absolut dengan norma vektor. Dengan kata lain, x^* adalah limit dari $\{x^{(k)}\}$ jika untuk sembarang ε positif ada bilangan K (yang mungkin bergantung pada ε) sedemikian rupa sehingga untuk semua $k > K$, $\|x^{(k)} - x^*\| < \varepsilon$. Seperti sebelumnya, jika barisan $\{x^{(k)}\}$ konvergen, kita tulis $x^* = \lim_{k \rightarrow \infty} x^{(k)}$ atau $x^{(k)} \rightarrow x^*$

Teorema 5.1 Barisan konvergen hanya memiliki satu limit.

Bukti. Kita buktikan hasil ini dengan kontradiksi. Misalkan suatu barisan $\{x^{(k)}\}$ memiliki dua limit yang berbeda, misalkan x_1 dan x_2 . Kemudian, kita memiliki $\|x_1 - x_2\| > 0$. Misal

$$\varepsilon = \frac{1}{2} \|x_1 - x_2\|.$$

Dari definisi suatu limit, terdapat K_1 dan K_2 sehingga untuk $k > K_1$ kita dapat $\|x^{(k)} - x_1\| < \varepsilon$, dan untuk $k > K_2$ kita dapat $\|x^{(k)} - x_2\| < \varepsilon$. Misal $K = \max(K_1, K_2)$. Kemudian, jika $k > K$, kita dapat $\|x^{(k)} - x_1\| < \varepsilon$ dan $\|x^{(k)} - x_2\| < \varepsilon$. Menambahkan $\|x^{(k)} - x_1\| < \varepsilon$ dan $\|x^{(k)} - x_2\| < \varepsilon$ menghasilkan

$$\|x^{(k)} - x_1\| + \|x^{(k)} - x_2\| < 2\varepsilon.$$

Menerapkan pertidaksamaan segitiga memberikan

$$\begin{aligned} \|-x_1 + x_2\| &= \|x^{(k)} - x_1 - x^{(k)} + x_2\| = \|(x^{(k)} - x_1) - (x^{(k)} - x_2)\| \\ &\leq \|x^{(k)} - x_1\| + \|x^{(k)} - x_2\| \end{aligned}$$

Oleh karena itu,

$$\|-x_1 + x_2\| = \|x_1 - x_2\| < 2\varepsilon$$

Namun hal tersebut bertentangan dengan asumsi bahwa $\|x_1 - x_2\| = 2\varepsilon$. Yang menyelesaikan bukti.

Barisan $\{x^{(k)}\}$ dalam \mathbb{R}^n dibatasi jika terdapat bilangan $B > 0$ sehingga $\|x^{(k)}\| \leq B$ untuk semua $k = 1, 2, \dots$

Teorema 5.2 Setiap barisan konvergen memiliki batas.

Bukti. Misalkan $\{x^{(k)}\}$ adalah barisan konvergen dengan limit x^* . Pilih $\varepsilon = 1$. Kemudian, dengan definisi limit, terdapat bilangan asli K sehingga untuk semua $k > K$

$$\|x^{(k)} - x^*\| < 1$$

Dari hasil latihan 2.6, didapatkan

$$\|x^{(k)}\| - \|x^*\| \leq \|x^{(k)} - x^*\| < 1 \text{ untuk semua } k > K.$$

oleh karena itu,

$$\|x^{(k)}\| < \|x^*\| + 1 \text{ untuk semua } k > K.$$

Misalkan

$$B = \max(\|x^{(1)}\|, \|x^{(2)}\|, \dots, \|x^{(K)}\|, \|x^*\| + 1),$$

Didapat

$$B \geq \|x^{(k)}\| \text{ untuk semua } k,$$

Yang berarti barisan $\{x^{(k)}\}$ memiliki batas.

Untuk barisan $\{x_k\}$ pada \mathbb{R} , suatu bilangan B disebut batas atas jika $x_k \leq B$ untuk semua $k = 1, 2, \dots$. Dalam hal ini, kita katakan bahwa $\{x_k\}$ dibatasi ke atas. Demikian pula, B disebut batas bawah jika $x_k \geq B$ untuk semua $k = 1, 2, \dots$. Dalam hal ini, kita katakan bahwa $\{x_k\}$ dibatasi ke bawah. Suatu barisan dibatasi jika keduanya dibatasi ke atas dan dibatasi ke bawah.

Setiap barisan $\{x_k\}$ di \mathbb{R} yang memiliki batas atas merupakan batas atas terkecil (juga disebut supremum), yang merupakan bilangan terkecil B yang merupakan batas atas dari $\{x_k\}$. Demikian pula, setiap barisan $\{x^{(k)}\}$ di \mathbb{R} yang memiliki batas bawah merupakan batas bawah terbesar (juga disebut infimum). Jika B adalah batas atas terkecil dari barisan $\{x_k\}$, maka $x_k \leq B$ untuk

semua k , dan, untuk setiap $\varepsilon > 0$, terdapat bilangan K sedemikian sehingga $x_K > B - \varepsilon$. Pernyataan berlaku untuk batas bawah terbesar: jika B adalah batas bawah terbesar $\{x_k\}$, maka $x_k > B$ untuk semua K , dan untuk setiap $\varepsilon > 0$, ada bilangan K sedemikian rupa sehingga $x_K < B + \varepsilon$.

Teorema 5.3 Setiap barisan monoton yang memiliki batas di \mathbb{R} adalah konvergen

Bukti. Kita buktikan teorema untuk barisan tidak turun. Bukti untuk barisan tidak naik adalah sejalan.

Misalkan $\{x_k\}$ adalah barisan tak turun terbatas di \mathbb{R} , dan x^* memiliki batas atas. Ubah bilangan $\varepsilon > 0$. Kemudian, ada angka K sedemikian rupa sehingga $x_K > x^* - \varepsilon$.

Karena $\{x_k\}$ tidak berkurang, untuk setiap $k \geq K$,

$$x_k \geq x_K > x^* - \varepsilon.$$

Karena x^* adalah batas atas $\{x_k\}$, didapat

$$x_k \leq x^* < x^* + \varepsilon.$$

Oleh karena itu, untuk setiap $k \geq K$,

$$|x_k - x^*| < \varepsilon,$$

Yang berarti bahwa $x_k \rightarrow x^*$.

Misalkan kita diberikan barisan $\{x^{(k)}\}$ dan barisan bilangan asli $\{m_k\}$ yang naik. Barisannya

$$\{x^{(m_k)}\} = \{x^{(m_1)}, x^{(m_2)}, \dots\}$$

disebut subbarisan $\{x^{(k)}\}$. Sebuah subbarisan yang diberikan dengan demikian dapat diperoleh dengan mengabaikan beberapa elemen dari barisan yang diberikan.

Teorema 5.4 Pertimbangkan barisan konvergen $\{x^{(k)}\}$ dengan limit x^* . Kemudian, setiap subbarisan dari $\{x^{(k)}\}$ juga konvergen ke x^* .

Bukti. Misalkan $\{x^{(m_k)}\}$ adalah subbarisan dari $\{x^{(k)}\}$, di mana $\{m_k\}$ adalah kenaikan barisan bilangan asli. Perhatikan bahwa $m_k \geq k$ untuk semua $k = 1, 2, \dots$. Untuk menunjukkan ini, pertama perhatikan bahwa $m_1 \geq 1$ karena m_1 adalah bilangan asli. Selanjutnya, kita lanjutkan dengan induksi dengan mengasumsikan bahwa $m_k \geq k$. Kemudian, kita memiliki $m_{k+1} > m_k \geq k$, yang berarti bahwa $m_{k+1} \geq k + 1$. Oleh karena itu, kami telah menunjukkan bahwa $m_k \geq k$ untuk semua $k = 1, 2, \dots$.

Misalkan $\varepsilon > 0$ diberikan. Kemudian, menurut definisi limit, terdapat K sedemikian sehingga $\|x^{(k)} - x^*\| < \varepsilon$ untuk setiap $k > K$. Karena $m_k \geq k$, kita juga memiliki $\|x^{(m_k)} - x^*\| < \varepsilon$ untuk setiap $k > K$. Ini berarti bahwa

$$\lim_{k \rightarrow \infty} x^{(m_k)} = x^*.$$

Setiap barisan yang memiliki batas mengandung suatu barisan konvergen. Ini disebut teorema Bolzano-Weierstrass.

Sebuah fungsi $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ dan titik $x_0 \in \mathbb{R}^n$. Misalkan ada f^* sedemikian rupa sehingga untuk setiap barisan konvergen $\{x^{(k)}\}$ dengan limit x_0 , didapat

$$\lim_{k \rightarrow \infty} f(x^{(k)}) = f^*$$

Kemudian, kita gunakan notasi

$$\lim_{x \rightarrow x_0} f(x)$$

Untuk merepresentasikan limit f^*

f kontinu di x_0 jika dan hanya jika, untuk setiap barisan konvergen $\{x^{(k)}\}$ dengan limit x_0 , didapat

$$\lim_{k \rightarrow \infty} f(x^{(k)}) = f\left(\lim_{k \rightarrow \infty} x^{(k)}\right) = f(x_0)$$

Oleh karena itu, dengan menggunakan notasi yang diperkenalkan di atas, fungsi f kontinu di x_0 jika dan hanya jika

$$\lim_{x \rightarrow x_0} f(x) = f(x_0).$$

Kami mengakhiri bagian ini dengan beberapa hasil yang melibatkan barisan dan batas matriks. Hasil ini berguna dalam analisis algoritma

Kita katakan bahwa barisan $\{A_k\}$ dari matriks $m \times n$ konvergen ke matriks $m \times n$ A jika

$$\lim_{k \rightarrow \infty} \|A - A_k\| = 0$$

Lemma 5.1 misal $A \in \mathbb{R}^{n \times n}$. Kemudian $\lim_{k \rightarrow \infty} A^k = O$ jika dan hanya jika nilai eigen dari A memenuhi $|\lambda_i(A)| < 1, i = 1, \dots, n$.

Bukti. Untuk membuktikan teorema ini, kami menggunakan bentuk Jordan (see, e.g., [33]). Secara khusus, diketahui bahwa setiap matriks persegi mirip dengan bentuk Jordan, yaitu, ada ada T nonsingular sedemikian rupa sehingga

$$TAT^{-1} = \text{diag}[J_{m_1}(\lambda_1), \dots, J_{m_3}(\lambda_1), J_{n_1}(\lambda_2), \dots, J_{t_v}(\lambda_q)] \triangleq J,$$

Dimana $J_r(\lambda)$ adalah matriks $\times r$:

$$J_r(\lambda) = \begin{bmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$$

$\lambda_1, \dots, \lambda_q$ diatas adalah nilai eigen yang berbeda dari A , banyaknya λ_1 adalah $m_1 + \dots + m_3$ dan seterusnya.

Kita dapat tulis ulang sebagai $A = T^{-1}JT$. Untuk melengkapi bukti perhatikan bahwa

$$J_r(\lambda)^k = \begin{bmatrix} \lambda^k & \binom{k}{1} \lambda^{k-1} & \dots & \binom{k}{k-1} \lambda^{k-r+1} \\ 0 & \lambda^k & \dots & \binom{k}{k-2} \lambda^{k-r+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda^k \end{bmatrix}$$

Dimana

$$\binom{k}{i} = \frac{k!}{i!(k-i)!}$$

Selanjutnya

$$A^k = T^{-1}J^kT$$

Oleh karenanya,

$$\lim_{k \rightarrow \infty} A^k = T^{-1}(\lim_{k \rightarrow \infty} J^k)T = O$$

Jika dan hanya jika $|\lambda_i| < 1, i = 1, \dots, n$.

Lemma 5.2 Deret matriks $n \times n$

$$I_n + A + A^2 + \dots + A^k + \dots$$

Konvergen jika dan hanya jika $\lim_{k \rightarrow \infty} A^k = O$. Dalam hal ini jumlah deretnya sama dengan $(I_n - A)^{-1}$.

Bukti. The necessity of the condition is obvious

Untuk membuktikan syarat cukup, misalkan $\lim_{k \rightarrow \infty} A^k = O$. dari Lemma 5.1 kita dapat menyimpulkan bahwa $|\lambda_i(A)| < 1, i = 1, \dots$ Ini menyiratkan bahwa $\det(I_n - A) \neq 0$, dan maka $(I_n - A)^{-1}$ ada. Pertimbangkan sekarang hubungan berikut:

$$(I_n + A + A^2 + \dots + A^k)(I_n - A) = I_n - A^{k+1}$$

Setelah mengalikan persamaan diatas dengan $(I_n - A)^{-1}$ didapat hasil

$$\lim_{k \rightarrow \infty} \sum_{j=0}^k A^j = (I_n - A)^{-1}$$

Karena $\lim_{k \rightarrow \infty} A^{k+1} = O$. Dengan demikian

$$\sum_{j=0}^{\infty} A^j = (I_n - A)^{-1}.$$

Fungsi bernilai matriks $A : \mathbb{R}^r \rightarrow \mathbb{R}^{n \times n}$ kontinu pada titik $\xi_0 \in \mathbb{R}^r$ jika

$$\lim_{\|\xi - \xi_0\| \rightarrow 0} \|A(\xi) - A(\xi_0)\| = 0$$

Lemma 5.3 Misal $A : \mathbb{R}^r \rightarrow \mathbb{R}^{n \times n}$ menjadi fungsi bernilai matriks $n \times n$ yaitu terus menerus pada ξ_0 . jika $A(\xi_0)^{-1}$, kemudian $A(\xi)^{-1}$ ada untuk ξ cukup dekat dengan ξ_0 dan $A(\cdot)^{-1}$ kontinu pada ξ_0 .

Bukti. Kita buktikan terlebih dahulu keberadaan $A(\xi)^{-1}$ untuk semua ξ cukup mendekati ξ_0 .
Didapatkan

$$A(\xi) = A(\xi_0) - A(\xi_0) + A(\xi) = A(\xi_0)(I_n - K(\xi)),$$

Dimana

$$K(\xi) = A(\xi_0)^{-1}(A(\xi_0) - A(\xi)).$$

Dengan demikian

$$\|K(\xi)\| \leq \|A(\xi_0)^{-1}\| \|A(\xi_0) - A(\xi)\|$$

Dan

$$\lim_{\|\xi - \xi_0\| \rightarrow 0} \|K(\xi)\| = 0$$

Karena A kontinu pada ξ_0 , untuk semua ξ cukup dekat dengan ξ_0 , didapatkan

$$\|A(\xi_0) - A(\xi)\| \leq \frac{\theta}{\|A(\xi_0)^{-1}\|}$$

Dimana $\theta \in (0,1)$, maka

$$\|K(\xi)\| \leq \theta < 1$$

Dan

$$(I_n - K(\xi))^{-1}$$

Ada, tapi kemudian

$$A(\xi)^{-1} = \left(A(\xi_0)(I_n - K(\xi)) \right)^{-1} = (I_n - K(\xi))^{-1} A(\xi_0)^{-1},$$

Yang berarti bahwa $A(\xi)^{-1}$ ada untuk ξ cukup dekat dengan ξ_0

Untuk membuktikan kontinuitas dari $A(\cdot)^{-1}$

$$\begin{aligned} \|A(\xi_0)^{-1} - A(\xi)^{-1}\| &= \|A(\xi)^{-1} - A(\xi_0)^{-1}\| \\ &= \left\| \left((I_n - K(\xi))^{-1} - I_n \right) A(\xi_0)^{-1} \right\| \end{aligned}$$

Namun, karena $\|K(\xi)\| < 1$, maka dari Lemma 5.2

$$(I_n - K(\xi))^{-1} - I_n = K(\xi) + K^2(\xi) + \dots = K(\xi)(I_n + K(\xi) + \dots)$$

Karenanya

$$\left\| (I_n - K(\xi))^{-1} - I_n \right\| \leq \|K(\xi)\|(1 + \|K(\xi)\| + \|K(\xi)\|^2 + \dots) = \frac{\|K(\xi)\|}{1 - \|K(\xi)\|}$$

Ketika $\|K(\xi)\| < 1$.

$$\|A(\xi)^{-1} - A(\xi_0)^{-1}\| \leq \left(\frac{\|K(\xi)\|}{1 - \|K(\xi)\|} \right) \|A(\xi_0)^{-1}\|.$$

Karena

$$\lim_{\|\xi - \xi_0\| \rightarrow 0} \|K(\xi)\| = 0,$$

Diperoleh

$$\lim_{\|\xi - \xi_0\| \rightarrow 0} \|A(\xi)^{-1} - A(\xi_0)^{-1}\| = 0,$$

Yang melengkapi bukti.

5.2 Differensibilitas

Kalkulus diferensial didasarkan pada gagasan untuk mendekati fungsi arbitrer dengan fungsi affine. Fungsi $\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ adalah affine jika ada fungsi linier $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ dan sebuah vektor $y \in \mathbb{R}^m$ seperti yang

$$\mathcal{A}(x) = \mathcal{L}(x) + y$$

Untuk setiap $x \in \mathbb{R}^n$. Pertimbangkan sebuah fungsi $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, dan titik $x_0 \in \mathbb{R}^n$. Kita ingin menemukan fungsi affine \mathcal{A} yang mendekati f di dekat titik x_0 . Pertama, wajar untuk memaksakan kondisi

$$\mathcal{A}(x_0) = f(x_0)$$

Karena $\mathcal{A}(x) = \mathcal{L}(x) + y$, diperoleh $y = f(x_0) - \mathcal{L}(x_0)$. Dengan linearitas \mathcal{L} ,

$$\mathcal{L}(x) + y = \mathcal{L}(x) - \mathcal{L}(x_0) + f(x_0) = \mathcal{L}(x - x_0) + f(x_0)$$

oleh karna itu, kita dapat tulis

$$\mathcal{A}(x) = \mathcal{L}(x - x_0) + f(x_0)$$

Selanjutnya, kita memberikan syarat bahwa $\mathcal{A}(x)$ mendekati $f(x)$ lebih cepat daripada x mendekati x_0 , yaitu

$$\lim_{x \rightarrow x_0, x \in \Omega} \frac{\|f(x) - \mathcal{A}(x)\|}{\|x - x_0\|} = 0$$

Kondisi pada \mathcal{A} memastikan bahwa \mathcal{A} mendekati f didekat x_0 dalam arti bahwa kesalahan dalam pendekatan pada titik tertentu adalah “kecil” dibandingkan dengan jarak titik dari x_0 .

Singkatnya, fungsi $f : \Omega \rightarrow \mathbb{R}^m, \Omega \subset \mathbb{R}^n$, dikatakan terdiferensiasi pada $x_0 \in \Omega$ jika ada fungsi affine yang mendekati f didekat x_0 , yaitu ada fungsi linier $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ seperti

$$\lim_{x \rightarrow x_0, x \in \Omega} \frac{\|f(x) - (\mathcal{L}(x - x_0) + f(x_0))\|}{\|x - x_0\|} = 0$$

Fungsi linier \mathcal{L} di atas secara tunggal ditentukan oleh f dan x_0 , dan disebut turunan dari f pada x_0 . Fungsi f dikatakan terdiferensiasi pada Ω jika f terdiferensiasi pada setiap titik domainnya Ω .

Dalam \mathbb{R} , sebuah fungsi affine berbentuk $ax + b$, dengan $a, b \in \mathbb{R}$. Oleh karena itu sebuah fungsi bernilai real fungsi $f(x)$ dari variabel nyata x yang terdiferensialkan di x_0 dapat didekati dekat x_0 dengan fungsi

$$\mathcal{A}(x) = ax + b$$

Karena $f(x_0) = \mathcal{A}(x_0) = ax_0 + b$, diperoleh

$$\mathcal{A}(x) = ax + b = a(x - x_0) + f(x_0).$$

Barisan linier dari $\mathcal{A}(x)$, yang sebelumnya dilambangkan dengan $\mathcal{L}(x)$, dalam hal ini hanya ax . norma dari bilangan real adalah nilai absolutnya, jadi menurut definisi diferensiasi,

$$\lim_{x \rightarrow x_0} \frac{|f(x) - (a(x - x_0) + f(x_0))|}{|x - x_0|} = 0,$$

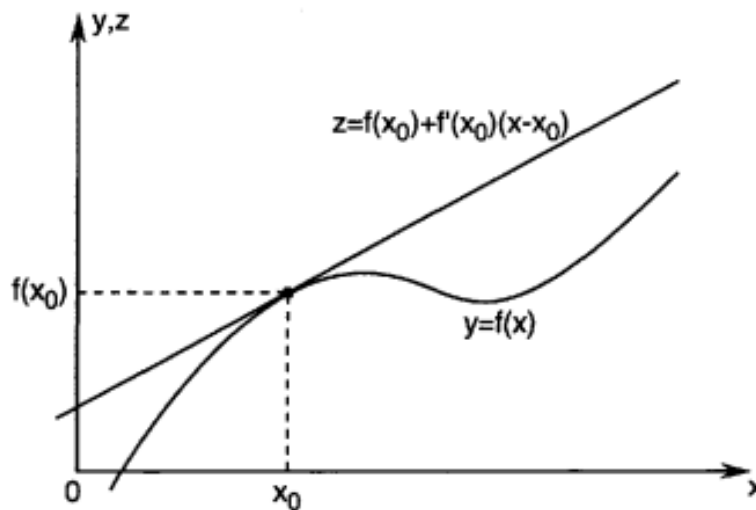
Yang setara dengan

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = a,$$

Bilangan a biasanya dilambangkan dengan $f'(x_0)$, dan disebut turunan dari f pada x_0 . Oleh karena itu, fungsi affine \mathcal{A} diberikan oleh

$$\mathcal{A}(x) = f(x_0) + f'(x_0)(x - x_0)$$

Fungsi affine ini bersinggungan dengan f di x_0



Gambar 5.1 Ilustrasi turunan

5.3 Matriks Derivatif

Transformasi linier dari \mathbb{R}^n ke \mathbb{R}^m , dan khususnya turunan \mathcal{L} dari $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, dapat direpresentasikan oleh matriks $m \times n$. Untuk menemukan representasi matriks L dari turunan \mathcal{L} dari fungsi terdiferensiasi $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, kita menggunakan basis natural $\{e_1, \dots, e_n\}$ untuk \mathbb{R}^n . Perhatikan vektor

$$x_j = x_0 + te, j = 1, \dots, n.$$

Dengan definisi turunan, kita memiliki

$$\lim_{t \rightarrow 0} \frac{f(x_j) - (tLe_j + f(x_0))}{t} = 0$$

untuk $j = 1, \dots, n$. Ini berarti bahwa

$$\lim_{t \rightarrow 0} \frac{f(x_j) - f(x_0)}{t} = Le_j$$

untuk $j = 1, \dots, n$. Tetapi Le_j adalah kolom ke- j dari matriks L . Sebaliknya, vektor x_j berbeda dari x_0 hanya pada koordinat ke- j , dan pada koordinat tersebut perbedaannya hanya pada t . Oleh karena itu, ruas kiri persamaan terakhir adalah turunan parsial

$$\frac{\partial f}{\partial x_j}(x_0).$$

Karena limit vektor dihitung dengan mengambil limit dari masing-masing fungsi koordinat, maka jika

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix},$$

kemudian

$$\frac{\partial f}{\partial x_j}(x_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_j}(x_0) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(x_0) \end{bmatrix},$$

dan matriks L memiliki bentuk

$$\left[\frac{\partial f}{\partial x_1}(x_0) \quad \dots \quad \frac{\partial f}{\partial x_n}(x_0) \right] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \dots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \dots & \frac{\partial f_m}{\partial x_n}(x_0) \end{bmatrix}.$$

Matriks L disebut matriks Jacobian, atau matriks turunan, dari f di x_0 , dan dinotasikan dengan $Df(x_0)$. Untuk memudahkan, kita sering menyebut $Df(x_0)$ hanya sebagai turunan dari f di x_0 . Pembahasan di atas kita rangkum dalam teorema berikut.

Teorema 5.5 Jika suatu fungsi $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ terdiferensial di x_0 , maka turunan dari f di x_0 ditentukan secara unik dan direpresentasikan oleh matriks turunan $m \times n$ $Df(x_0)$.

Pendekatan affine terbaik untuk f mendekati x_0 diberikan oleh

$$\mathcal{A}(x) = f(x_0) + Df(x_0)(x - x_0),$$

dalam arti bahwa

$$f(x) = \mathcal{A}(x) + r(x)$$

dan $\lim_{x \rightarrow x_0} \|r(x)\| / \|x - x_0\| = 0$. Kolom dari matriks turunan $Df(x_0)$ adalah turunan parsial vektor. Vektor

$$\frac{\partial f}{\partial x_j}(x_0)$$

adalah vektor tangen di x_0 terhadap kurva f yang diperoleh dengan memvariasikan hanya koordinat ke- j dari x .

Jika $f : \mathbb{R}^n \rightarrow \mathbb{R}$ terdiferensial, maka fungsi ∇f didefinisikan oleh

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix} = Df(x)^T$$

disebut gradien dari f . Gradien adalah fungsi dari \mathbb{R}^n ke \mathbb{R}^n , dan dapat digambarkan sebagai medan vektor, dengan menggambar panah yang merepresentasikan $\nabla f(x)$ sehingga bagian akhir dimulai di x .

Diberikan $f : \mathbb{R}^n \rightarrow \mathbb{R}$, jika ∇f terdiferensial, kita katakan bahwa f terdiferensial dua kali, dan kita tuliskan turunan dari ∇f sebagai

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

Matriks $D^2f(x)$ disebut matriks Hessian dari f di x , dan dinotasikan dengan $F(x)$.

5.4 Aturan Diferensiasi

Suatu fungsi $f: \Omega \rightarrow \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$, dikatakan terdiferensial kontinu pada Ω jika terdiferensial pada Ω , dan $Df: \Omega \rightarrow \mathbb{R}^{m \times n}$ kontinu, yaitu komponen-komponen f memiliki turunan parsial kontinu. Dalam hal ini, kita menulis $f \in C^1$. Jika komponen f memiliki turunan parsial kontinu orde p , maka kita tulis $f \in C^p$. Perhatikan bahwa matriks Hessian dari suatu fungsi $f: \mathbb{R}^n \rightarrow \mathbb{R}$ di x adalah simetris jika f terdiferensial dua kali secara kontinu di x .

Kita sekarang membuktikan aturan rantai untuk membedakan komposisi $g(f(t))$, dari fungsi $f: \mathbb{R} \rightarrow \mathbb{R}^n$ dan fungsi $g: \mathbb{R}^n \rightarrow \mathbb{R}$.

Teorema 5.6 Misalkan $g: \mathcal{D} \rightarrow \mathbb{R}$ terdiferensial pada himpunan terbuka $\mathcal{D} \subset \mathbb{R}^2$, dan misalkan $f: (a, b) \rightarrow \mathcal{D}$ terdiferensial pada (a, b) . Maka, fungsi komposit $h: (a, b) \rightarrow \mathbb{R}$ yang diberikan $h(t) = g(f(t))$ terdiferensial pada (a, b) , dan

$$h'(t) = Dg(f(t))Df(t) = \nabla g(f(t))^T \begin{bmatrix} f_1'(t) \\ \vdots \\ f_n'(t) \end{bmatrix},$$

Bukti. Menurut definisi,

$$h'(t) = \lim_{s \rightarrow t} \frac{h(s) - h(t)}{s - t} = \lim_{s \rightarrow t} \frac{g(f(s)) - g(f(t))}{s - t}$$

jika limit ada. Dengan Teorema 5.5, kita tulis

$$g(f(s)) - g(f(t)) = Dg(f(t))(f(s) - f(t)) + r(s),$$

Dimana $\lim_{s \rightarrow t} r(s)/(s - t) = 0$. Oleh karena itu,

$$\frac{h(s) - h(t)}{s - t} = Dg(f(t)) \frac{f(s) - f(t)}{s - t} + \frac{r(s)}{s - t}.$$

Misalkan $s \rightarrow t$ menghasilkan

$$h'(t) = \lim_{s \rightarrow t} Dg(f(t)) \frac{f(s) - f(t)}{s - t} + \frac{r(s)}{s - t} = Dg(f(t)) Df(t).$$

Selanjutnya, kita menyajikan aturan perkalian. Misalkan $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ dan $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ adalah dua fungsi yang dapat diturunkan. Tentukan fungsi $h: \mathbb{R}^n \rightarrow \mathbb{R}$ dan $h(x) = f(x)^T g(x)$. Maka, h juga terdiferensial, dan

$$Dh(x) = f(x)^T Dg(x) + g(x)^T Df(x).$$

Kit akhiri bagian ini dengan daftar beberapa rumus yang berguna dari kalkulus multivariabel. Dalam setiap kasus, kita menghitung turunan terhadap x . Misalkan $A \in \mathbb{R}^{m \times n}$ adalah matriks tertentu, dan $y \in \mathbb{R}^m$ vektor tertentu. Kemudian,

$$D(y^T Ax) = y^T A$$

$$D(x^T Ax) = x^T (A + A^T).$$

Ini mengikuti dari rumus pertama di atas bahwa jika $y \in \mathbb{R}^m$, maka

$$D(y^T x) = y^T.$$

Ini mengikuti dari rumus kedua di atas bahwa jika Q adalah matriks simetris, maka

$$D(x^T Qx) = 2x^T Q.$$

Khususnya,

$$D(x^T x) = 2x^T.$$

5.5 Tingkatan Himpunan dan Gradien

Tingkatan himpunan dari fungsi $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tingkatan c adalah himpunan point

$$S = \{x : f(x) = c\}$$

Untuk $f : \mathbb{R}^n \rightarrow \mathbb{R}$, biasanya tertarik pada S adalah kurva, untuk $f : \mathbb{R}^n \rightarrow \mathbb{R}$ himpunan S yang paling sering dianggap adalah permukaan.

Contoh 5.1 Perhatikan fungsi bernilai real berikut pada \mathbb{R}^2 :

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2, \quad x = [x_1, x_2]^T.$$

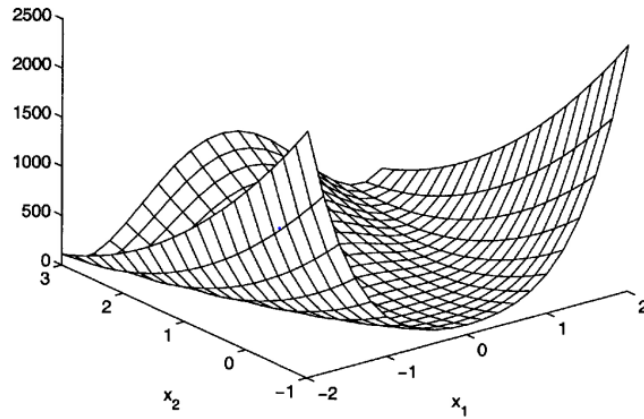
Fungsi di atas disebut fungsi Rosenbrock. Plot fungsi f ditunjukkan pada Gambar 5.2. Kumpulan level f pada level 0,7, 7, 70, 200, dan 700 digambarkan pada Gambar 5.3. Set level ini memiliki bentuk tertentu yang menyerupai pisang. Untuk alasan ini, fungsi Rosenbrock juga disebut "fungsi pisang".

Mengatakan bahwa suatu titik x_0 berada pada level himpunan S pada level c berarti $f(x_0) = c$. Sekarang anggaplah ada kurva γ yang terletak di S dan diparameterisasikan oleh fungsi terdiferensiasi kontinu $g : \mathbb{R} \rightarrow \mathbb{R}^2$. Misalkan juga $g(x_0) = x_0$. dan $Sg(t_0) = v \neq 0$, sehingga v adalah vektor tangen ke γ di x_0 (lihat Gambar 5.4). Menerapkan aturan rantai ke fungsi $h(t) = f(g(x))$ pada t_0 , memberikan

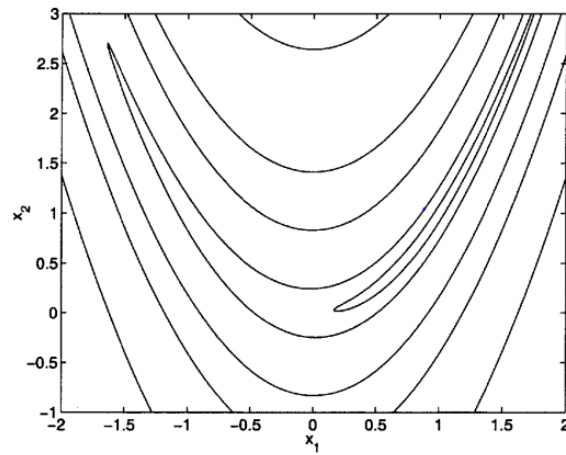
$$h'(t_0) = Df(g(x))Dg(t_0) = Df(x_0)v.$$

Tapi karena γ terletak di S , kita punya

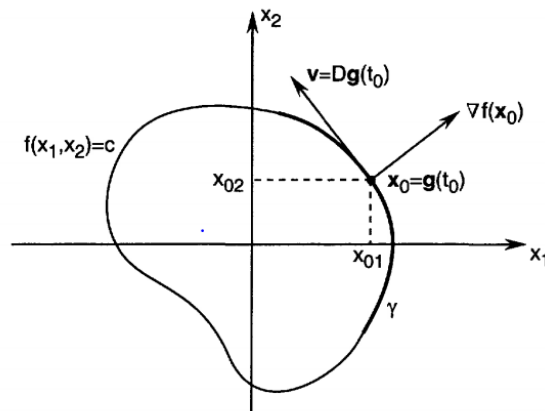
$$h(t) = f(g(t)) = c$$



Gambar 5.3 grafik dari fungsi Rosenbrock.



Gambar 5.2 Himpunan tingkatan fungsi Rosenbrock (banana)



Gambar 5.4 Ortogonalitas gradien himpunan tingkat

yaitu, h konstan. Jadi, $h'(t_0) = 0$ dan

$$Df(x_0)v = \nabla f(x_0)^T v = 0$$

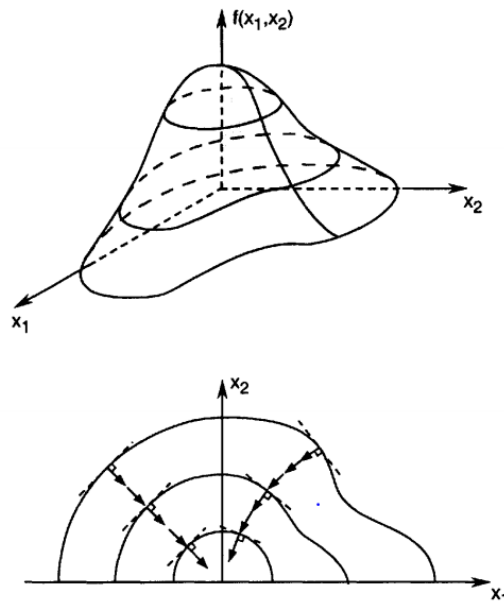
Oleh karena itu, kami telah membuktikan, dengan asumsi f terdiferensiasi terus menerus, teorema berikut (lihat Gambar 5.4).

Teorema 5.7 Vector $\nabla f(x_0)$ orthogonal terhadap vector singgung kurva yang melalui vector x_0 pada tingkat yang di tentukan oleh $f(x) = f(x_0)$. Untuk mengatakan bahwa $\nabla f(x_0)$ orthogonal atau normal ke arah level S yang sesuai dengan x_0 dan untuk mengambil bidang singgung (atau garis) ke S di x_0 himpunan semua titik x memenuhi.

$$\nabla f(x_0)^T (x - x_0) = 0 \text{ jika } \nabla f(x_0) \neq 0$$

$\nabla f(x_0)$ adalah arah laju kenaikan f di x , karena $\nabla f(x_0)$ orthogonal terhadap target yang ditetapkan melalui x_0 yang ditentukan oleh $f(x) = f(x_0)$ dapat disimpulkan bahwa arah laju kenaikan maksimal fungsi terdiferensiasi bernilai nyata pada suatu titik adalah orthogonal ke level himpunan fungsi melalui titik.

Gambar 5.5 mengilustrasikan pembahasan di atas untuk kasus $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Kurva pada bidang yang diarsir pada Gambar 5.5 berjalan dari bawah ke atas memiliki sifat proyeksinya ke bidang (x_1, x_2) selalu ortogonal ke kurva tingkat, dan disebut jalur pendakian paling curam, karena selalu menuju ke arah tingkat kenaikan maksimum untuk f .



Gambar 5.5 Ilustrasi jalur steepest ascent

Grafik dari $f : \mathbb{R}^n \rightarrow \mathbb{R}$ adalah himpunan $\{(x^T, f(x))^T : x \in \mathbb{R}^n\} \subset \mathbb{R}^{n+1}$. Gagasan gradien suatu fungsi memiliki interpretasi alternatif yang berguna dalam hal hiperplane tangen ke grafiknya. Misalkan $x_0 \in \mathbb{R}^n$ dan $z_0 = f(x_0)$. Titik $[x_0^T, z_0]^T \in \mathbb{R}^{n+1}$ adalah titik pada grafik f . Jika f terdiferensiasi pada x_0 , maka grafik memiliki hiperplane tangen nonvertikal $\xi = [x_0^T, z_0]^T$. Hiperplane melalui ξ adalah himpunan semua titik $[x_1, x_2, \dots, x_n, z]^T \in \mathbb{R}^{n+1}$ yang memenuhi persamaan

$$u_1(x_1 - x_{01}) + \dots + u_n(x_n - x_{0n}) + v(z - z_0) = 0$$

dimana vektor $[u_1, u_2, \dots, u_n, v]^T \in \mathbb{R}^{n+1}$ normal pada hiperplane. Dengan asumsi bahwa hiperplane ini nonvertikal, yaitu, $v \neq 0$, misalkan

$$d_i = -\frac{u_i}{v}$$

Dengan demikian, kita dapat menulis ulang persamaan hiperplane di atas sebagai

$$z = d_1(x_1 - x_{10}) + \dots + d_n(x_n - x_{n0}) + z_0$$

Kita dapat menganggap ruas kanan persamaan di atas sebagai fungsi $z : \mathbb{R}^n \rightarrow \mathbb{R}$. Perhatikan bahwa agar hiperplane bersinggungan dengan grafik f , fungsi f dan z harus memiliki turunan

parsial yang sama di titik x_0 . Oleh karena itu, jika f dapat diturunkan di x_0 , hyperplane tangennya dapat ditulis dalam gradiennya, seperti yang diberikan oleh persamaan

$$z - z_0 = Df(x_0)(x - x_0) = (x - x_0)^T \nabla f(x_0)$$

5.6 Deret Taylor

Dasar dari banyak metode numerik dan model untuk optimasi adalah Taylor rumus, yang diberikan oleh teorema Taylor di bawah ini.

Teorema 5.8. Teorema Taylor. Asumsikan bahwa fungsi $f : \mathbb{R} \rightarrow \mathbb{R}$ adalah m waktu terdiferensiasi terus menerus (yaitu, $f \in C^m$) pada interval $[a, b]$. Dilambangkan dengan $h = b - a$. Maka $f(b) = f(a) + \frac{h}{1!} f^{(1)}(a) + \frac{h^2}{2!} f^{(2)}(a) + \dots + \frac{h^{(m-1)}}{(m-1)!} f^{(m-1)}(a) + R_m$

(disebut rumus Taylor) dimana $f^{(i)}$ adalah derivatif ke- i dari f , dan

$$R_m = \frac{h^m (1 - \theta)^{(m-1)}}{(m-1)!} f^{(m)}(a + \theta h) = \frac{h^m}{m!} f^{(m)}(a + \theta' h)$$

Dengan $\theta, \theta' \in (0, 1)$

Bukti. Kita memiliki

$$R_m = f(b) - f(a) + \frac{h}{1!} f^{(1)}(a) + \frac{h^2}{2!} f^{(2)}(a) + \dots + \frac{h^{(m-1)}}{(m-1)!} f^{(m-1)}(a)$$

Disimbolkan dengan $g_m(x)$ fungsi bantu yang diperoleh dari R_m dengan mengganti a dengan x .

Oleh karena itu, $g_m(x)$

$$\begin{aligned} g_m^{(1)} &= -f(x) + \left[f^{(1)}(x) - \frac{b-x}{1!} f^{(2)}(x) \right] + \left[2 \frac{b-x}{2!} f^{(2)}(x) - \frac{(b-x)^2}{2!} f^{(3)}(x) \right] + \dots \\ &\quad + \left[(m-1) \frac{(b-x)^{m-2}}{(m-1)!} f^{(m-1)}(x) - \frac{(b-x)^{m-1}}{(m-1)!} f^{(m)}(x) \right] \\ &= -\frac{(b-x)^{m-1}}{(m-1)!} f^{(m)}(x) \end{aligned}$$

Perhatikan bahwa $g_m(b) = 0$ dan $g_m(a) = R_m$. Dengan menerapkan teorema nilai rata-rata menghasilkan

$$\frac{g_m(b) - g_m(a)}{b - a} = g_m^{(1)}(a + \theta h)$$

Dimana $\theta \in (0,1)$. Persamaan di atas sama dengan

$$-\frac{R_m}{h} = -\frac{(b-a-\theta h)^{m-1}}{(m-1)!} f^{(m)}(a+\theta h) = \frac{h^{m-1}(1-\theta)^{m-1}}{(m-1)!} f^{(m)}(a+\theta h)$$

Sehingga,

$$R_m = \frac{h^{m-1}(1-\theta)^{m-1}}{(m-1)!} f^{(m)}(a+\theta h)$$

Sifat penting dari teorema Taylor muncul dari bentuk sisa R_m . Untuk membahas sifat ini lebih lanjut, maka diperkenalkan apa yang disebut *simbol urutan*, O , dan o .

Misalkan g menjadi fungsi bernilai nyata yang didefinisikan dari $0 \in \mathbb{R}^n$, dengan $g(x) \neq 0$ jika $x \neq 0$. Misalkan $f: \Omega \rightarrow \mathbb{R}^m$ didefinisikan dalam domain $\Omega \rightarrow \mathbb{R}^n$ termasuk 0. Maka, dapat ditulis

- 1) $f(x) = O(g(x))$ berarti hasil bagi $|f(x)|/|g(x)|$ adalah dibatasi dekat 0; sehingga ada bilangan $K > 0$ dan $\delta > 0$ sedemikian sehingga jika $\|x\| < \delta, x \in \Omega$, maka $|f(x)|/|g(x)| \leq K$.
- 2) $f(x) = o(g(x))$ berarti

$$\lim_{x \rightarrow 0, x \in \Omega} \frac{|f(x)|}{|g(x)|} = 0$$

Simbol $O(g(x))$ digunakan untuk menyatakan fungsi yang dibatasi oleh versi skala g di sekitar 0. Contoh fungsinya adalah:

- $x = O(x)$
- $\left[\frac{x^3}{2x^2 + 3x^4} \right] = O(x^2)$
- $\cos x = O(1)$
- $\sin x = O(x)$

Di sisi lain, $o(g(x))$ (baca "sedikit-oh dari $g(x)$ ") merepresentasikan fungsi yang menuju ke nol "lebih cepat" dari $g(x)$ dalam arti bahwa $\lim_{x \rightarrow 0} \frac{o(g(x))}{|g(x)|} = 0$. Contoh fungsi tersebut adalah :

- $x^2 = o(x)$

- $\left[\frac{x^3}{2x^2 + 3x^4} \right] = o(x)$
- $x^3 = o(x^2)$
- $x = o(1)$.

Perhatikan bahwa jika $f(x) = o(g(x))$, maka $f(x) = O(g(x))$ (tetapi kebalikannya belum tentu benar). Juga, jika $f(x) = O(\|x\|^p)$, maka $f(x) = o(\|x\|^{p-\varepsilon})$ untuk setiap $\varepsilon > 0$.

Misalkan $f \in \mathcal{C}^m$. Ingatlah bahwa suku sisa dalam teorema Taylor memiliki bentuk

$$R_m = \frac{h^m}{m!} f^{(m)}(a + \theta h),$$

dimana $\theta \in (0,1)$. Substitusikan persamaan di atas ke dalam rumus Taylor, kita peroleh

$$f(b) = f(a) + \frac{h}{1!} f^{(1)}(a) + \frac{h^2}{2!} f^{(2)}(a) + \dots + \frac{h^{m-1}}{(m-1)!} f^{(m-1)}(a) + \frac{h^m}{m!} f^{(m)}(a + \theta h).$$

Dengan kontinuitas $f^{(m)}$, kita mendapatkan $f^{(m)}(a + \theta h) \rightarrow f^{(m)}(a)$ sebagai $h \rightarrow 0$, yaitu, $f^{(m)}(a + \theta h) = f^{(m)}(a) + o(1)$. Oleh karena itu,

$$\frac{h^m}{m!} f^{(m)}(a + \theta h) = \frac{h^m}{m!} f^{(m)}(a) + o(h^m),$$

karena $h^m o(1) = o(h^m)$. Kita dapat menulis rumus Taylor sebagai

$$f(b) = f(a) + \frac{h}{1!} f^{(1)}(a) + \frac{h^2}{2!} f^{(2)}(a) + \dots + \frac{h^m}{m!} f^{(m)}(a) + o(h^m).$$

Jika, sebagai tambahan, kita berasumsi bahwa $f \in \mathcal{C}^{m+1}$, kita dapat mengganti suku $o(h^m)$ di atas dengan $O(h^{m+1})$. Untuk melihat ini, pertama kita tulis rumus Taylor dengan R_{m+1} :

$$f(b) = f(a) + \frac{h}{1!} f^{(1)}(a) + \frac{h^2}{2!} f^{(2)}(a) + \dots + \frac{h^m}{m!} f^{(m)}(a) + R_{m+1},$$

Dimana

$$R_{m+1} = \frac{h^{m+1}}{(m+1)!} f^{(m+1)}(a + \theta' h),$$

dengan $\theta' \in (0,1)$. Karena $f^{(m+1)}$ terbatas pada $[a, b]$ (dengan Teorema 4.2),

$$R_{m+1} = O(h^{m+1}).$$

Oleh karena itu, jika $f \in \mathcal{C}^{m+1}$, kita dapat menulis rumus Taylor sebagai

$$f(b) = f(a) + \frac{h}{1!} f^{(1)}(a) + \frac{h^2}{2!} f^{(2)}(a) + \dots + \frac{h^m}{m!} f^{(m)}(a) + O(h^{m+1}).$$

Kita sekarang beralih ke ekspansi deret Taylor dari fungsi bernilai nyata $f: \mathbb{R}^n \rightarrow \mathbb{R}$ tentang titik $x_0 \in \mathbb{R}^n$. Misalkan $f \in \mathcal{C}^2$. Misalkan x dan x_0 adalah titik-titik di \mathbb{R}^n , dan misalkan $z(\alpha) = x_0 + \alpha(x - x_0)/\|x - x_0\|$. Tentukan $\phi: \mathbb{R} \rightarrow \mathbb{R}$ dengan:

$$\phi'(\alpha)f(z(\alpha)) = f(x_0 + \alpha(x - x_0)/\|x - x_0\|).$$

Dengan menggunakan aturan rantai, kita peroleh

$$\begin{aligned} \phi'(\alpha) &= \frac{d\phi}{d\alpha}(\alpha) & &= Df(z(\alpha))Dz(\alpha) \\ &= Df(z(\alpha))\frac{(x - x_0)}{\|x - x_0\|} & &= \frac{1}{\|x - x_0\|}(x - x_0)^T Df(z(\alpha))^T, \end{aligned}$$

Dan

$$\begin{aligned} \phi''(\alpha) &= \frac{d^2\phi}{d\alpha^2}(\alpha) \\ &= \frac{d}{d\alpha}\left(\frac{d\phi}{d\alpha}\right)(\alpha) \\ &= \frac{(x - x_0)^T}{\|x - x_0\|} \frac{d}{d\alpha} Df(z(\alpha))^T \\ &= \frac{(x - x_0)^T}{\|x - x_0\|} D(Df)(z(\alpha))^T \frac{dz}{d\alpha}(\alpha) \\ &= \frac{1}{\|x - x_0\|^2} (x - x_0)^T D^2 f(z(\alpha))^T (x - x_0) \\ &= \frac{1}{\|x - x_0\|^2} (x - x_0)^T D^2 f(z(\alpha))(x - x_0), \end{aligned}$$

dimana kita mengingatnya

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

dan $D^2 f = (D^2 f)^T$ karena $f \in \mathcal{C}^2$. Perhatikan bahwa

$$f(x) = \phi(\|x - x_0\|) = \phi(0) + \frac{\|x - x_0\|}{1!} \phi'(0) + \frac{\|x - x_0\|^2}{2!} \phi''(0) + o(\|x - x_0\|^2).$$

Karena,

$$f(x) = f(x_0) + \frac{1}{1!} Df(x_0)(x - x_0) + \frac{1}{2!} (x - x_0)^T D^2 f(x_0)(x - x_0) + o(\|x - x_0\|^2).$$

Jika kita berasumsi bahwa $f \in \mathcal{C}^3$, kita dapat menggunakan rumus untuk suku sisa R_3 untuk menyimpulkan bahwa

$$f(x) = f(x_0) + \frac{1}{1!} Df(x_0)(x - x_0) + \frac{1}{2!} (x - x_0)^T D^2 f(x_0)(x - x_0) + O(\|x - x_0\|^3).$$

Untuk bacaan lebih lanjut dalam kalkulus, lihat [9], [60], [61], [85], [87], [97]. Perlakuan dasar analisis nyata dapat ditemukan di [2], [82], sedangkan perlakuan yang lebih maju disediakan di [66], [81]. Untuk merangsang membaca pada notasi "besar-oh", lihat [56, hlm. 104-108].

Latihan Soal

5.1 Tunjukkan bahwa kondisi yang cukup untuk $\lim_{k \rightarrow \infty} A^k = O$ adalah $\|A\| < 1$

Penyelesaian :

Perhatikan bahwa

$$\|A^k\| \leq \|A^{k-1}\| \|A\| \leq \|A^{k-2}\| \|A\|^2 \leq \dots \leq \|A\|^k$$

Karenanya, jika $\|A\| < 1$ maka $\lim_{k \rightarrow \infty} \|A^k\| = 0$ sehingga berarti bahwa $\lim_{k \rightarrow \infty} A^k = O$.

5.2 Tunjukkan bahwa untuk setiap matriks $A \in \mathbb{R}^{n \times n}$

$$\|A\| \geq \max_{1 \leq i \leq n} |\lambda_i(A)|$$

Penyelesaian :

Untuk kasus ketika A memiliki semua nilai eigen real, pembuktiannya sederhana. Misalkan λ adalah nilai eigen dari A dengan nilai absolut terbesar dan x vector eigen yang sesuai (dinormalisasi) yaitu, $Ax = \lambda x$ dan $\|x\| = 1$. Maka

$$\|A\| \geq \|Ax\| = \|\lambda x\| = \|\lambda\| \|x\| = |\lambda|$$

Yang mana memenuhi pembuktian pada kasus ini.

Secara umum, nilai eigen dari A dan sama dengan vector eigen mungkin bisa kompleks.

Perhatikan matriks

$$B = \frac{A}{\|A\| + \varepsilon}$$

Dimana ε adalah bilangan real positif

$$\|B\| = \frac{\|A\|}{\|A\| + \varepsilon} < 1$$

$B^k \rightarrow 0$ sebagai $k \rightarrow \infty$ dan dengan demikian oleh $|\lambda_i(B)| < 1, i = 1, \dots, n$

$$\lambda_i(B) = \frac{\lambda_i(A)}{\|A\| + \varepsilon}$$

Dan dengan demikian

$$|\lambda_i(B)| = \frac{|\lambda_i(A)|}{\|A\| + \varepsilon} < 1$$

Maka

$$|\lambda_i(B)| < \|A\| + \varepsilon$$

Karena pernyataan di atas berlaku untuk sembarang $\varepsilon > 0$, maka $|\lambda_i(A)| \leq \|A\|$

5.3. Definisikan fungsi $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ dan $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ dengan $f(x) = \frac{x_1^2}{6} + \frac{x_2^2}{4}$, $g(t) = [3t + 5, 2t - 6]^T$. Misalkan $F: \mathbb{R} \rightarrow \mathbb{R}$ diberikan $F(t) = f(g(t))$. Hitung $\frac{dF}{dt}(t)$ menggunakan aturan rantai

Penyelesaian :

Diketahui

$$Df(x) = \left[\frac{x_1}{3}, \frac{x_2}{2} \right]$$

Dan

$$\frac{dg}{dt}(t) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Dengan aturan rantai,

$$\begin{aligned} \frac{d}{dt} F(t) &= Df(g(t)) \frac{dg}{dt}(t) \\ &= \left[\frac{(3t+5)}{3}, \frac{(2t-6)}{2} \right] \begin{bmatrix} 3 \\ 2 \end{bmatrix} \\ &= 5t - 1 \end{aligned}$$

5.4 Perhatikan $f(x) = \frac{x_1 x_2}{2}$, $g(s, t) = [4s + 3t, 2s + t]^T, t \in \mathbb{R}$. Hitung $\frac{\partial g}{\partial s} f(g(s, t))$ dan $\frac{\partial}{\partial t} f(g(s, t))$ dengan aturan rantai

Penyelesaian :

Diketahui

$$Df(x) = \left[\frac{x_2}{2}, \frac{x_1}{2} \right]$$

Dan

$$\frac{\partial g}{\partial s}(s, t) = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \frac{\partial g}{\partial t}(s, t) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Dengan aturan rantai,

$$\frac{\partial}{\partial s} f(g(s, t)) = Df(g(s, t)) \frac{\partial g}{\partial s}(s, t) = \frac{1}{2} [2s + t, 4s + 3t] \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 5s + 3t$$

5.5. Misalkan $x(t) = [e^t + t^3, t^2]^T, t \in \mathbb{R}$ dan $f(x) = x^2 x_2 x_3^2 + x_1 x_2 + x_3, x = [x_1, x_2, x_3]^T, t \in \mathbb{R}^3$. Temukan $\frac{d}{dt} f(x(t))$ di waktu t

Penyelesaian :

Diketahui

$$Df(x) = [3x_1^2 x_2 x_3^2 + x_2, x_1^3 x_3^2 + x_1, 2x_3^2 x_2 x_3 + 1]$$

Dan

$$\frac{dx}{dt}(t) = \begin{bmatrix} 2^t + 3t^2 \\ 2t \\ 1 \end{bmatrix}$$

Dengan aturan rantai,

$$\begin{aligned} \frac{d}{dx} f(x(t)) &= Df(x(t)) \frac{dx}{dt}(t) \\ &= [3x_1(t)^2 x_2(t) x_3(t)^2 + x_2(t), x_1(t)^3 x_3(t)^2 + x_1(t), 2x_1(t)^3 x_2(t) x_3(t) \\ &\quad + 1] \begin{bmatrix} e^t + 3t^2 \\ 2t \\ 1 \end{bmatrix} = 12t(e^t + 3t^2)^3 + 2te^t + 6t^2 + 2t + 1 \end{aligned}$$

5.6 Misalkan $f(x) = o(g(x))$. Tunjukkan bahwa untuk setiap $\varepsilon > 0$, ada $\delta > 0$ sudah jika $\|x\| < \delta$ maka $\|f(x)\| < \varepsilon |o(x)|$.

Penyelesaian :

Misalkan $\varepsilon > 0$. karena $f(x) = o(g(x))$, maka

$$\lim_{x \rightarrow 0} \frac{\|f(x)\|}{g(x)} = 0$$

Maka ada $\delta > 0$, sehingga jika $\|x\| < \delta$, kemudian

$$\frac{\|f(x)\|}{g(x)} < \varepsilon$$

Dapat ditulis ulang sebagai :

$$\| f(x) \| \leq \varepsilon g(x)$$

5.7. Gunakan latihan 5.6 untuk menunjukkan bahwa jika fungsi $f: \mathbb{R}^n \rightarrow \mathbb{R}$ memenuhi $f(x) = -g(x) + o(g(x))$ dan $g(x) > 0$ untuk semua $x \neq 0$. Cukup kecil, kita memiliki $f(x) < 0$

Penyelesaian :

Dengan latihan 5.6 ada $\delta > 0$ sedemikian rupa sehingga jika $\| x \| < \delta$, maka $|o(g(x))| < \frac{g(x)}{2}$. karenanya jika $\| x \| < \delta, x \neq 0$ maka

$$f(x) \leq -g(x) + |o(g(x))| < -g(x) + \frac{g(x)}{2} = -\frac{1}{2} g(x) < 0$$

5.8 Misalkan

$$f_1(x_1, x_2) = x_1^2 - x_2^2, f_2(x_1, x_2) = 2x_1 x_2$$

Gambarkan sketsa himpunan tingkat yang terkait dengan $f_1(x_1, x_2) = 12$ dan $f_1(x_1, x_2) = 6$ pada diagram yang sama. Tunjukkan pada diagram nilai $x = [x_1, x_2]^T$ untuk

$$f(x) = [f_1(x_1, x_2), f_2(x_1, x_2)]^T = [12, 6]^T$$

Penyelesaian :

Diketahui

$$\{x : f_1(x) = 12\} = \{x : x_1^2 - x_2^2 = 12\}$$

$$\{x : f_2(x) = 16\} = \{x : x_2 = 8/x_1\}$$

Untuk menemukan titik potong, substitusi $x_2 = 8/x_1$ menjadi $x_1^2 - x_2^2 = 12$ untuk mendapatkan $x_1^4 - 12x_1^2 - 64 = 0$. Selesaikan $x_1^2 = 16, -4$. Jelas, hanya 2 kemungkinan untuk x_1 adalah $x_1 = +4, -4$, yang mana $x_2 = +2, -2$. Sehingga titik potong terletak di $[4, 2]^T$ dan $[-4, -2]^T$. Himpunan tingkat terkait dengan $f_1(x_1, x_2) = 12$ dan $f_1(x_1, x_2) = 16$ ditunjukkan

5.9. Tuliskan perluasan deret Taylor dari fungsi-fungsi berikut tentang pemberian titik x_0 .
Abaikan ketentuan orde 3 atau lebih tinggi

- $f(x) = x_1 e^{-x_2} + x_2 + 1, x_0 = [1, 0]^T$
- $f(x) = x_1^4 + 2x_1^2 x_2^2 + x_2^4, x_0 = [1, 1]^T$
- $f(x) = e^{x_1 - x_2} + e^{x_1 + x_2} + x_1 + x_2 + 1, x_0 = [1, 0]^T$

Penyelesaian :

$$a. f(x) = f(x_0) + Df(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T D^2 f(x_0)(x - x_0) + \dots$$

Perhitungan,

$$Df(x) = [e^{-x_2}, -x_1 e^{-x_2} + 1]$$

$$D^2 f(x) = \begin{bmatrix} 0 & -e^{-x_2} \\ -e^{-x_2} & x_1 e^{-x_2} \end{bmatrix}$$

Karenanya

$$\begin{aligned} f(x) &= 2 + [1, 0] \begin{bmatrix} x_1 - 1 \\ x_2 \end{bmatrix} + \frac{1}{2} [x_1 - 1, x_2] \begin{bmatrix} 0 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 \end{bmatrix} + \dots \\ &= 1 + x_1 + x_2 - x_1 x_2 + \frac{1}{2} x_2^2 + \dots \end{aligned}$$

$$b. f(x) = x_1^4 + 2x_1^2 x_2^2 + x_2^4, x_0 = [1, 1]^T$$

$$Df(x) = [4x_1^3 + 4x_1 x_2^2, 4x_1^2 x_2 + 4x_2^3] \quad D^2 f(x) = \begin{bmatrix} 12x_1^2 + 4x_2^2 & 8x_1 x_2 \\ 8x_1 x_2 & 4x_1^2 + 12x_2^2 \end{bmatrix}$$

Memperluas f terhadap titik x_0 , menghasilkan

$$\begin{aligned} f(x) &= 4 + [8, 8] \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} + \frac{1}{2} [x_1 - 1, x_2 - 1] \begin{bmatrix} 6 & 8 \\ 8 & 16 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} + \dots \\ &= 8x_1^2 + 8x_2^2 - 16x_1 - 16x_2 + 8x_1 x_2 + 12 + \dots \end{aligned}$$

c. $f(x) = e^{x_1-x_2} + e^{x_1+x_2} + x_1 + x_2 + 1, x_0 = [1,0]^T$

$$Df(x) = [e^{x_1-x_2} + e^{x_1+x_2} + 1, -e^{x_1-x_2} + e^{x_1+x_2} + 1]D^2f(x)$$

$$= \begin{bmatrix} e^{x_1-x_2} + e^{x_1+x_2} & -e^{x_1-x_2} \\ -e^{x_1-x_2} & e^{x_1-x_2} + e^{x_1+x_2} \end{bmatrix}$$

Memperluas f terhadap titik x_0 , menghasilkan

$$f(x) = 2 + 2e + [2e + 1, 1] \begin{bmatrix} x_1 - 1 \\ x_2 \end{bmatrix} + \frac{1}{2} [x_1 - 1, x_2] \begin{bmatrix} 2e & 0 \\ 0 & 2e \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 \end{bmatrix} + \dots$$

$$= 1 + x_1 - e(1 + x_1^2 + x_2^2) \dots$$

BAB 6

Konsep Matematika dalam

Optimasi

6.1 Konsep Dasar Optimasi

Secara umum, optimasi matematik adalah proses dari

- Formulasi dan
- Solusi dari masalah optimasi dengan kendala dari bentuk umum matematis :

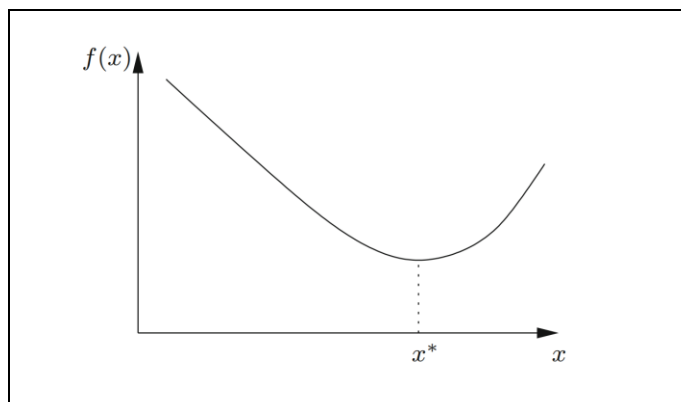
$$\max \text{ or } \min f(x), \quad x = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n$$

Subject to

$$g_j(x) \leq 0, \quad j = 1, 2, \dots, m$$

$$h_j(x) = 0, \quad j = 1, 2, \dots, r$$

dimana $f(x)$, $g_j(x)$ and $h_j(x)$ adalah fungsi scalar dari vektor kolom nyata x



Fungsi dari variabel tunggal dengan optimum di x^*

Diberikan minimalisasi fungsi halus, yaitu kontinu dan terdiferensiasi dua kali kontinu (C^2) dari variabel nyata tunggal, yaitu masalah

$$\min f(x), \quad x \in \mathbb{R}, \quad f \in C^2$$

6.2 Vektor Gradien $f(x)$

Untuk suatu fungsi $f(x) \in C^2$ ada, di sembarang titik x vektor turunan parsial orde pertama, atau vektor gradien. Kita mengarah menuju kondisi untuk titik x^* menjadi minimum lokal. Kita gunakan turunan dari suatu fungsi $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Ingat bahwa turunan orde pertama dari f , dinotasikan sebagai Df , yakni

$$Df = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right]$$

$$\nabla f = (Df)^T$$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

6.3 Matriks Hessian dari $f(x)$

Jika $f(x)$ terdiferensialkan dua kali berturut-turut di titik x terdapat matriks turunan parsial orde dua atau Matriks Hessian

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

$$= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$$

Contoh

Misalkan $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$, Maka,

$$\begin{aligned} Df(x) &= (\nabla f(x))^T = \left[\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x) \right] \\ &= [5 + x_2 - 2x_1, 8 + x_1 - 4x_2] \end{aligned}$$

$$\begin{aligned} F(x) = D^2f(x) &= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) \end{bmatrix} \\ &= \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix} \end{aligned}$$

Fungsi Kuadratik pada \mathbb{R}^n

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c$$

dimana $c \in \mathbb{R}$, b adalah suatu n -vektor nyata dan A adalah matriks real $n \times n$ yang dapat dipilih dengan cara yang tidak unik. Biasanya dipilih simetris dalam hal berikut bahwa

$$\nabla f(x) = Ax + b; \quad H(x) = A$$

Minimum Global

x^* adalah minimum global pada himpunan X jika $f(x) \geq f(x^*)$, $\forall x \in X \subset \mathbb{R}^n$

Minimum Lokal Kuat

x^* adalah minimum lokal kuat $\varepsilon > 0$ sedemikian sehingga

$$f(x) > f(x^*) \forall \{x \mid \|x - x^*\| < \varepsilon\}$$

dimana $\|\cdot\|$ menotasikan norma Euclidean

Titik Balik

$f(x)$ adalah titik balik di $\bar{x} = \begin{bmatrix} x^0 \\ y^0 \end{bmatrix}$ jika terdapat suatu $\varepsilon > 0$ sedemikian sehingga $\forall x, \|x - x^0\| < \varepsilon$ dan setiap $y, \|y - y^0\| < \varepsilon$: $f(x, y^0) \leq f(x^0, y^0) \leq f(x^0, y)$

Teorema Orde Pertama Kondisi yang Diperlukan

Misalkan Ω menjadi suatu himpunan bagian \mathbb{R}^n dan $f \in C^1$ suatu fungsi bernilai real di Ω . Jika x^* adalah minimum lokal dari f terhadap Ω , maka untuk setiap arah yang layak d di x^* , kita memiliki

$$d^T \nabla f(x^*) \geq 0$$

Corollary Kasus Interior

Misalkan Ω menjadi himpunan bagian dari \mathbb{R}^n dan $f \in C^1$ suatu fungsi bernilai real di Ω . Jika x^* adalah minimum lokal dari f terhadap Ω dan jika x^* adalah suatu titik interior Ω , maka

$$\nabla f(x^*) = 0$$

Teorema Orde Kedua Kondisi yang Diperlukan

Misalkan $\Omega \subset \mathbb{R}^n$, $f \in C^2$ adalah suatu fungsi di Ω , x^* adalah suatu minimum lokal f terhadap Ω , dan d adalah suatu arah yang layak di x^* . Jika $d^T \nabla f(x^*) = 0$, maka

$$d^T F(x^*) d \geq 0$$

dimana F adalah Hessian dari f

Theorem Orede Kedua Kondisi Cukup, Kasus Interior

Misalkan $f \in C^2$ adalah didefinisikan pada suatu area yang mana x^* adalah suatu titik interior. Misalkan bahwa

1. $\nabla f(x^*) = 0$; dan
2. $F(x^*) > 0$

Maka, x^* adalah tepat minimum lokal dari f

Definisi 6.1 peminimum lokal. Misalkan bahwa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ adalah nilai fungsi nyata yang didefinisikan pada sebuah himpunan $\Omega \subset \mathbb{R}^n$. Sebuah titik $x^* \in \Omega$ adalah peminimum lokal dari f pada Ω jika terdapat $\epsilon > 0$ sehingga $f(x^*) \leq f(x)$ untuk semua $x \in \Omega \setminus \{x^*\}$ dan $\|x - x^*\| < \epsilon$.

Peminimum global. Sebuah titik $x^* \in \Omega$ adalah peminimum global dari f pada Ω jika $f(x^*) \leq f(x)$ untuk semua $x \in \Omega \setminus \{x^*\}$.

Jika, definisi diatas, kami mengganti " \geq " dengan " $>$ " maka kami punya masing-masing peminimum lokal yang tepat dan peminimum global yang tepat.

Pada gambar 6.1, kita mengilustrasikan secara grafis definisi diatas untuk $n = 1$.

Diberikan sebuah nilai fungsi nyata f , notasi $\arg \min f(x)$ dinotasikan sebagai argumen bahwa fungsi minimum f (titik domain dari f), kami sumpsikan bahwa setiap titik unik. Contohnya, jika $f : \mathbb{R} \rightarrow \mathbb{R}$ diberikan oleh $f(x) = (x + 1)^2 + 3$, maka $\arg \min f(x) = -1$. Jika kita tulis $\arg \min_{x \in \Omega}$, maka kita menganggap Ω sebagai domain dari f . Sebagai contoh, untuk fungsi f diatas, $\arg \min_{x \geq 0} f(x) = 0$. Pada umumnya, kami menduga bahwa $\arg \min_{x \in \Omega} f(x)$ sebagai peminimum global dari f pada Ω (asumsikan ada dan unik).

Sebenarnya, masalah optimasi diselesaikan hanya ketika peminimum global ditemukan. Namun, peminimum global pada umumnya sulit ditemukan. Oleh karena itu, dalam praktiknya, seringkali kita harus puas dengan mencari peminimum lokal.

6.4 Kondisi Peminimum Lokal

Pada bab ini, kita memperoleh kondisi untuk titik x^* menjadi peminimum lokal. Kita gunakan turunan dari fungsi $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Mengingat bahwa turunan orde pertama dari f dinotasikan Df , yaitu

$$Df \triangleq \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right]$$

Perhatikan bahwa gradien ∇f hanya transpose dari Df , yakni, $\nabla f = (Df)^T$. Turunan kedua dari $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (disebut juga Hessian dari f) adalah

$$F(x) \triangleq D^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$$

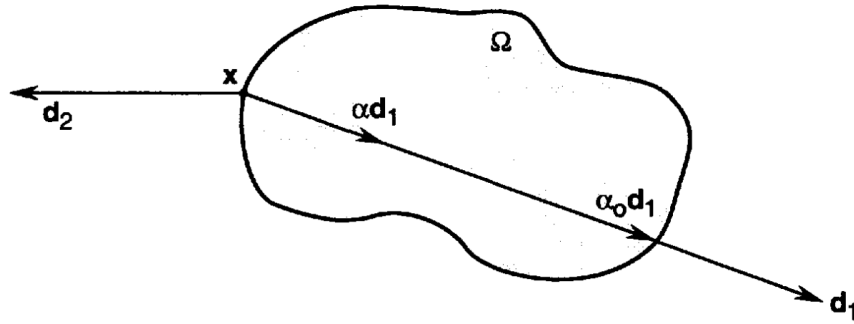
Diberikan masalah optimasi dengan himpunan kendala Ω , sebuah peminimum mungkin terletak di bagian dalam atau pada batas Ω . Untuk mempelajari kasus di mana ia terletak pada batas, kita memerlukan dugaan arah fisibel.

Definisi 6.2 Arah yang layak. Sebuah vektor $d \in \mathbb{R}^n, d \neq 0$, adalah arah yang layak di $x \in \Omega$ jika terdapat $\alpha_0 > 0$ sehingga $x + \alpha d \in \Omega$ untuk semua $\alpha \in [0, \alpha_0]$.

Gambar 6.2 ilustrasi dari dugaan dari arah fisibel.

Misalkan $f : \mathbb{R}^n \rightarrow \mathbb{R}$ adalah fungsi nyata dan misalkan d adalah arah fisibel di $x \in \Omega$. Arah turunan dari f dalam arah d dinotasikan $\partial f / \partial d$, sehingga fungsi nilai nyata didefinisikan sebagai

$$\frac{\partial f}{\partial d}(x) = \lim_{\alpha \rightarrow 0} \frac{f(x + \alpha d) - f(x)}{\alpha}$$



Gambar 6.2 Ilustrasi dimensi dua dari arah fisibel; d_1 adalah arah fisibel, d_2 bukanlah arah fisibel.

Jika $\|d\| = 1$, maka $\partial f / \partial d$ adalah laju kenaikan f di x dalam arah d . Untuk menghitung turunan arah diatas, anggap bahwa x dan d diberikan. Maka, $f(x + \alpha d)$ adalah fungsi dari α dan

$$\frac{\partial f}{\partial d}(x) = \left. \frac{d}{d\alpha} f(x + \alpha d) \right|_{\alpha=0}$$

Menerapkan aturan rantai menghasilkan

$$\begin{aligned} \frac{\partial f}{\partial d}(x) &= \left. \frac{d}{d\alpha} f(x + \alpha d) \right|_{\alpha=0} \\ &= \nabla f(x)^T d \\ &= \langle \nabla f(x), d \rangle \\ &= d^T \nabla f(x) \end{aligned}$$

Kesimpulannya, jika d adalah unit vektor, dengan kata lain, $\|d\| = 1$, maka $\langle \nabla f(x), d \rangle$ adalah laju kenaikan dari f di titik x pada arah d .

Contoh 6.2 Definisikan $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ dengan $f(x) = x_1 x_2 x_3$ dan misalkan

$$d = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}} \right]^T$$

Turunan arah dari f pada arah d adalah

$$\begin{aligned}
\frac{\partial f}{\partial d}(x) &= \nabla f(x)^T d \\
&= [x_2 x_3, x_1 x_3, x_1 x_2] \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \\
&= \frac{x_2 x_3 + x_1 x_3 + \sqrt{2} x_1 x_2}{2}
\end{aligned}$$

Perhatikan bahwa karena $\|d\| = 1$, diatas juga merupakan laju kenaikan f pada x dalam arah d .

Kita siap untuk menyatakan dan membuktikan teorema berikut.

Teorema 6.1 Kondisi FONC yang diperlukan Orde Pertama. Misalkan Ω adalah subset dari \mathbb{R}^n dan $f \in C^1$ adalah fungsi nilai nyata pada Ω . Jika x^* adalah peminimum lokal dari f pada Ω , maka untuk setiap arah fisibel d pada x^* , kita punya

$$d^T \nabla f(x^*) \geq 0$$

Bukti. Definisikan

$$x(\alpha) = x^* + \alpha d \in \Omega$$

Perhatikan bahwa $x(0) = x^*$. Definisikan fungsi campuran

$$\phi(\alpha) = f(x(\alpha))$$

Lalu, dengan teorema Taylor

$$\begin{aligned}
f(x^* + \alpha d) - f(x^*) &= \phi(\alpha) - \phi(0) \\
&= \phi'(0)\alpha + o(\alpha) \\
&= \alpha d^T \nabla f(x(0)) + o(\alpha)
\end{aligned}$$

Dimana $\alpha \geq 0$ (meningat definisi dari $o(\alpha)$ (“sedikit-oh daro α ”) di bagian 1). Jadi, jika $\phi(\alpha) \geq \phi(0)$, yakni, $f(x^* + \alpha d) \geq f(x^*)$ untuk nilai yang cukup kecil dari $\alpha > 0$ (x^* adalah peminimum lokal) maka kita harus memiliki $d^T \nabla f(x^*) \geq 0$ (lihat exercise 5.7)

Teorema diatas adalah diilustrasikan secara grafis dalam gambar 6.3.

Suatu cara alternative untuk menyatakan FONC adalah

$$\frac{\partial f}{\partial d}(x^*) \geq 0$$

Untuk semua arah fisibel d . Dengan kata lain, jika x^* adalah peminimum lokal maka laju kenaikan dari f pada x^* dalam setiap arah fisibel d didalam Ω adalah non negative. Menggunakan turunan terarah, bukti alternative dari teorema 6.1 adalah sebagai berikut. Misalkan bahwa x^* adalah peminimum lokal, maka untuk setiap arah fisibel d , terdapat $\bar{\alpha} > 0$ sehingga untuk semua $\alpha \in (0, \bar{\alpha})$,

$$f(x^*) \leq f(x^* + \alpha d)$$

Oleh karena itu, untuk semua $\alpha \in (0, \bar{\alpha})$, kita punya

$$\frac{f(x^* + \alpha d) - f(x^*)}{\alpha} \geq 0$$

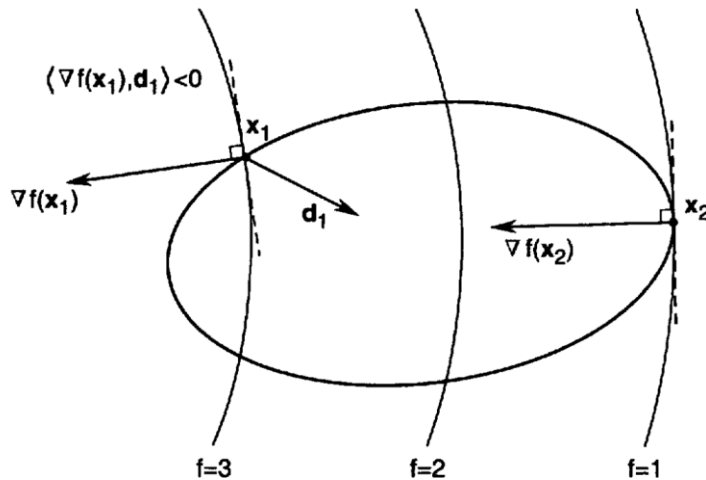
Mengambil limit sebagai $\alpha \rightarrow 0$, kami menyimpulkan bahwa

$$\frac{\partial f}{\partial d}(x^*) \geq 0$$

Suatu kasus special yang menarik adalah ketika x^* adalah titik interior dari Ω (lihat bab 4.4). dalam kasus ini, setiap arah dapat dilakukan dan kita mendapatkan hasil sebagai berikut.

Corollary 6.1 Kasus interior. Misalkan Ω adalah subset dari \mathbb{R}^n dan $f \in C^1$ adalah fungsi nyata pada Ω . Jika x^* adalah peminimum lokal dari f pada Ω dan jika x^* adalah titik interior dari Ω , maka

$$\nabla f(x^*) = 0$$



Gambar 6.3 Ilustrasi dari FONC untuk kasus terkendala, x_1 tidak memenuhi FONC, x_2 memenuhi FONC.

Bukti. Misalkan bahwa f mempunyai peminimum lokal x^* yakni sebuah titik interior dari Ω . Karena x^* adalah sebuah titik interior dari Ω , himpunan dari arah fisibel pada x^* adalah diseluruh \mathbb{R}^n . Jadi, untuk setiap $d \in \mathbb{R}^n$, $d^T \nabla f(x^*) \geq 0$ dan $-d^T \nabla f(x^*) \geq 0$. Oleh karena itu, $d^T \nabla f(x^*) = 0$ untuk semua $d \in \mathbb{R}^n$ yang mengimplikasikan bahwa $\nabla f(x^*) = 0$.

Contoh 6.3 Mempertimbangkan sebuah masalah

$$\min x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$$

Subject to $x_1, x_2 \geq 0$.

Pertanyaan:

- Apakah kondisi orde pertama FONC diperlukan untuk peminimum lokal memenuhi $= [1,3]^T$?
- Apakah FONC untuk peminimum lokal memenuhi $= [0,3]^T$?
- Apakah FONC untuk peminimum lokal memenuhi $= [1,0]^T$?
- Apakah FONC untuk peminimum lokal memenuhi $= [0,0]^T$?

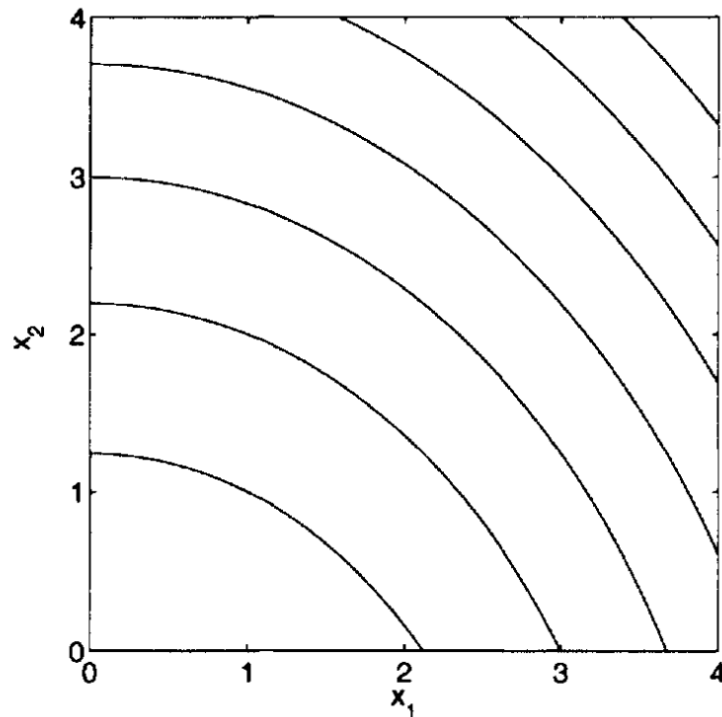
Jawab:

Pertama, misalkan $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ didefinisikan sebagai $f(x) = x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$, dimana $x = [x_1, x_2]^T$. Sebuah plot dari tingkat himpunan dari f ditunjukkan pada gambar 6.4.

a. Pada $x = [1,3]^T$ kita punya

$$\nabla f(x) = [2x_1, x_2 + 3]^T = [2,6]^T$$

Titik $x = [1,3]^T$ adalah titik interior dari $\Omega = \{x: x_1 \geq 0, x_2 \geq 0\}$. Oleh karena itu, FONC membutuhkan $\nabla f(x) = 0$. Titik $x = [1,3]^T$ tidak memenuhi FONC sebagai peminimum lokal.



Gambar 6.4 Tingkat himpunan fungsi pada contoh 6.3

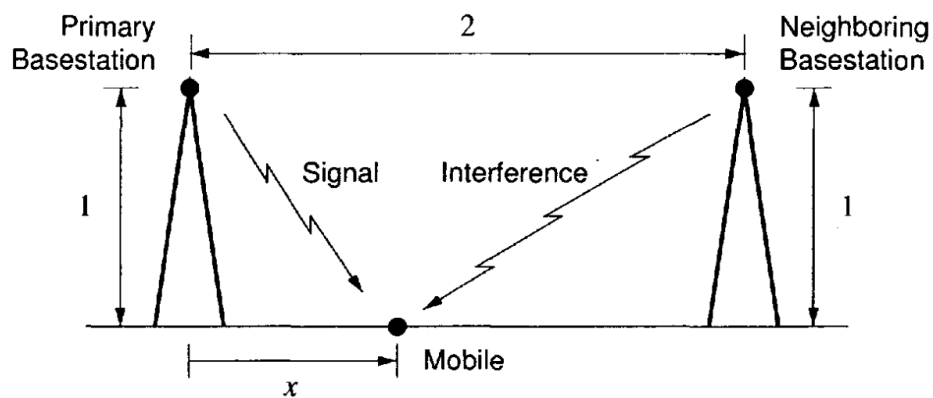
- b. Pada $x = [0,3]^T$ kita punya $\nabla f(x) = x = [0,6]^T$, oleh karena itu $d^T \nabla f(x) = 6d_2$, dimana $d = [d_1, d_2]^T$. Untuk d menjadi mungkin pada x , kita membutuhkan $d_1 \geq 0$ dan d_2 bisa diambil nilai \mathbb{R} yang berubah-ubah. Titik $x = [0,3]^T$ tidak memenuhi FONC untuk peminimum karena d_2 diperbolehkan kurang dari nol. Sebagai contoh, $d = [1, -1]^T$ arah fisibel, tetapi $d^T \nabla f(x) = -6 < 0$.
- c. Pada $x = [1,0]^T$, kita punya $\nabla f(x) = [2,3]^T$, dan oleh karena $d^T \nabla f(x) = 2d_1 + 3d_2$. Untuk d menjadi mungkin, kita membutuhkan $d_2 \geq 0$ dan d_1 bisa ambil nilai \mathbb{R} yang

berubah-ubah. Sebagai contoh, $d = [-5, 1]^T$ adalah arah fisibel. Tetapi $d^T \nabla f(x) = -7 < 0$. Jadi, $x = [1, 0]^T$ tidak memenuhi FONC sebagai peminimum lokal.

- d. Pada $x = [0, 0]^T$, kita punya $\nabla f(x) = [0, 3]^T$ dan oleh karena itu $d^T \nabla f(x) = 3d_2$. Untuk d menjadi mungkin, kita membutuhkan $d_2 \geq 0$ dan $d_1 \geq 0$. Oleh karena itu, $x = [0, 0]^T$ memenuhi FONC sebagai peminimum lokal.

Contoh 6.4 Gambar 6.5 menunjukkan simplifikasi model dari sistem nirkabel seluler (jarak yang ditunjukkan telah diperkecil untuk membuat perhitungan lebih sederhana). Seorang pengguna ponsel (disebut juga "ponsel") terletak di posisi x (lihat Gambar 6.5).

Ada dua antenna stasiun pangkalan, satu untuk stasiun pangkalan utama dan satu lagi untuk stasiun pangkalan tetangga. Kedua antenna mengirimkan sinyal ke pengguna ponsel, dengan daya yang sama. Namun, kekuatan sinyal yang diterima yang diukur oleh ponsel adalah kebalikan dari jarak kuadrat dari antenna terkait (stasiun pangkalan utama atau tetangga). Kami tertarik untuk menemukan posisi ponsel yang memaksimalkan rasio sinyal-ke-interferensi, yang merupakan rasio daya sinyal yang diterima dari basestation utama ke kekuatan sinyal yang diterima dari stasiun pangkalan tetangga.



Gambar 6.5 Simplifikasi sistem nirkabel seluler pada contoh 6.4

Kita menggunakan FONC untuk menyelesaikan masalah ini. Jarak kuadrat dari ponsel ke antenna utama adalah $1 + x^2$, sedangkan jarak kuadrat dari ponsel ke antenna tetangga adalah $1 + (2 - x)^2$. Oleh sebab itu, rasio sinyal-ke-interferensi adalah

$$f(x) = \frac{1 + x^2}{1 + (2 - x)^2}$$

Kita punya

$$f'(x) = \frac{-2x(1 + (2 - x)^2) - 2(2 - x)(1 + x^2)}{1 + (2 - x)^2}$$

Dengan FONC, pada posisi optimal x^* , kita punya $f'(x^*) = 0$. Oleh karena itu, salah satu $x^* = 1 - \sqrt{2}$ atau $x^* = 1 + \sqrt{2}$. Mengevaluasi fungsi objektif pada dua kandidat titik ini, mudah untuk melihat bahwa $x^* = 1 - \sqrt{2}$ adalah posisi optimal.

Kita memperoleh sebuah kondisi orde kedua yang diperlukan untuk memenuhi peminimum lokal.

Teorema 6.2 Kondisi FONC yang diperlukan Orde Kedua (SONC). Misalkan $\Omega \subset \mathbb{R}^n$, $f \in C^2$ sebuah fungsi pada Ω , x^* adalah peminimum lokal dari f pada Ω , dan d sebuah arah fisibel di x^* . Jika $d^T \nabla f(x^*) = 0$, maka

$$d^T F(x^*) d \geq 0,$$

Dimana F adalah Hessian dari f .

Bukti. Kami membuktikan hasilnya dengan kontradiksi. Misalkan ada arah fisibel d di x^* sehingga $d^T \nabla f(x^*) = 0$ dan $d^T F(x^*) d < 0$. Misalkan $x(\alpha) = x^* + \alpha d$ dan tentukan fungsi komposit $\phi(\alpha) = f(x^* + \alpha d) = f(x(\alpha))$. Kemudian, oleh teorema Taylor's

$$\phi(\alpha) = \phi(0) + \phi''(0) \frac{\alpha^2}{2} + o(\alpha^2)$$

Dimana dengan asumsi $\phi'(0) = d^T \nabla f(x^*) = 0$, dan $\phi''(0) = d^T F(x^*) d < 0$. Untuk nilai α cukup kecil,

$$\phi(\alpha) - \phi(0) = \phi''(0) \frac{\alpha^2}{2} + o(\alpha^2) < 0$$

Yaitu

$$f(x^* + \alpha d) < f(x^*)$$

Kontradiksi dengan asumsi bahwa x^* adalah peminimum lokal. Jadi,

$$\phi''(0) = d^T F(x^*) d \geq 0$$

Corollary 6.2 Kasus Interior. Misalkan x^* adalah titik interior dari $\Omega \subset \mathbb{R}^n$. Jika x^* adalah peminimum lokal dari $f : \Omega \rightarrow \mathbb{R}, f \in C^2$, maka

$$\nabla f(x^*) = 0$$

Dan $F(x^*)$ adalah semidefinit positif ($F(x^*) \geq 0$); yaitu untuk semua $d \in \mathbb{R}^n$,

$$d^T(x^*)d \geq 0$$

Bukti. Jika x^* adalah titik interior maka semua arah fisibel. Hasilnya kemudian mengikuti dari corollary 6.1 dan teorema 6.2.

Pada contoh dibawah, kami menunjukkan bahwa kondisi yang diperlukan tidak cukup.

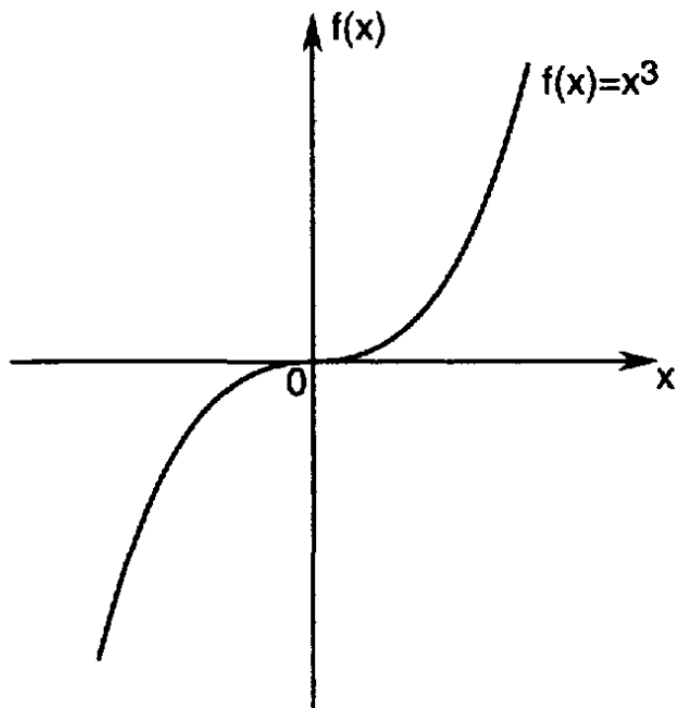
Contoh 6.5 Mempertimbangkan bahwa sebuah fungsi dari salah satu variabel $f(x)x^3, f : \mathbb{R} \rightarrow \mathbb{R}$. Karena, $f'(0) = 0$, dan $f''(0) = 0$, titik $x = 0$ memenuhi keduanya FONC dan SONC. Bagaimanapun, $x = 0$ bukanlah peminimum (lihat gambar 6.6).

Contoh 6.6 mempertimbangkan bahwa fungsi $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, dimana $f(x) = x_1^2 - x_2^2$. FONC mensyaratkan bahwa $\nabla f(x) = [2x_1, -2x_2]^T = 0$. Jadi, $x = [0,0]^T$ memenuhi FONC. Matriks Hessian dari f adalah

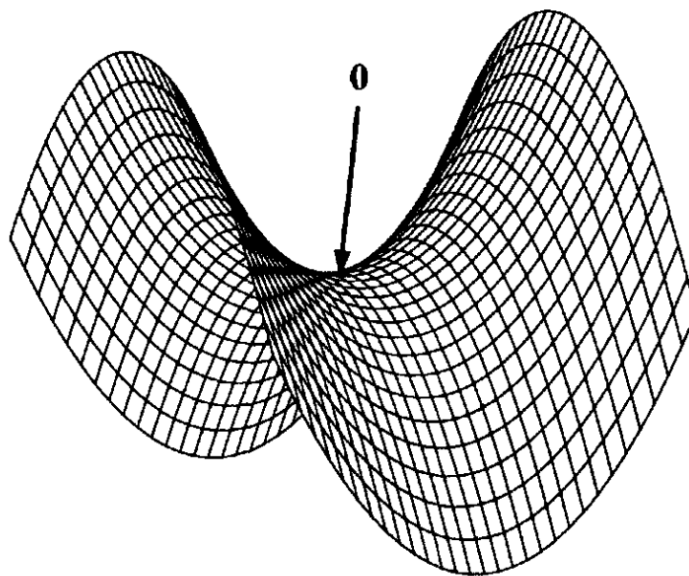
$$F(x) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

Matriks Hessian tak definit yaitu untuk beberapa $d_1 \in \mathbb{R}^2$ kita punya $d_1^T F d_1 > 0$, misalnya $d_1 = [1,0]^T$; dan untuk beberapa d_2 , kita punya $d_2^T F d_2 < 0$, misalnya $d_2 = [0,1]^T$. Jadi, $x = [0,0]^T$ tidak memenuhi SONC dan oleh karena itu bukanlah peminimum. Grafik dari $f(x) = x_1^2 - x_2^2$ ditunjukkan pada gambar 6.7.

Kami sekarang menurunkan kondisi cukup yang menyiratkan bahwa x^* adalah peminimum lokal.



Gambar 6.6 Titik 0 memenuhi FONC dan SONC, tetapi bukanlah peminimum.



Gambar 6.7 Grafik dari $f(x) = x_1^2 - x_2^2$. Titik 0 memenuhi FONC dan SONC dan titik ini bukanlah peminimum.

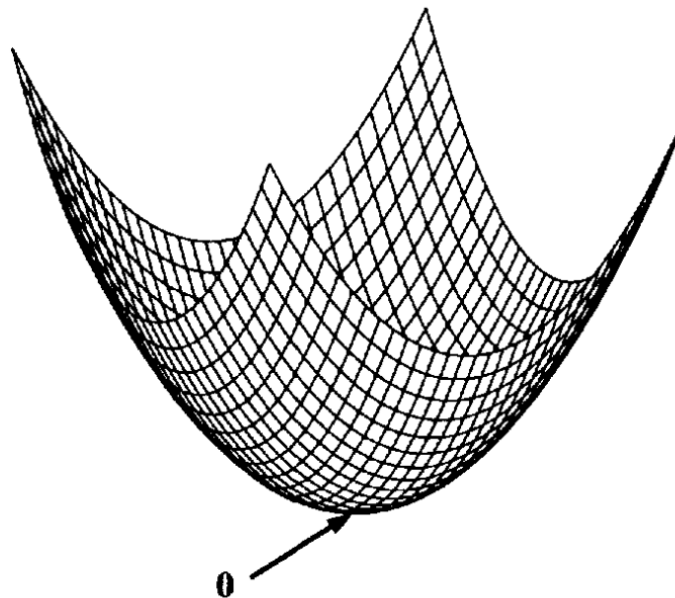
Teorema 6.3 Kondisi Cukup Orde-Dua (SOSC). Kasus interior. Misalkan $f \in C^2$ didefinisikan pada sebuah daerah dengan x^* adalah titik interior. Anggap bahwa

1. $\nabla f(x^*) = 0$
2. $F(x^*) > 0$

Maka, x^* adalah peminimum lokal yang tepat dari f .

Bukti. Karena $f \in C^2$, kita punya $F(x^*) = F^T(x^*)$. Dengan menggunakan asumsi 2 dan ketaksamaan Rayleigh, bahwa jika $d \neq 0$, maka $0 < \lambda_{\min}(F(x^*))\|d\|^2 \leq d^T F(x^*)d$. Dengan teorema Taylor dan asumsi 1,

$$f(x^* + d) - f(x^*) = \frac{1}{2} d^T F(x^*)d + o(\|d\|^2) \geq \frac{\lambda_{\min}(F(x^*))}{2} \|d\|^2 + o(\|d\|^2)$$



Gambar 6.8 Grafik dari $f(x) = x_1^2 + x_2^2$

Karenanya, untuk semua d sedemikian sehingga $\|d\|$ cukup kecil.

$$f(x^* + d) > f(x^*)$$

dan bukti selesai.

Contoh 6.7 Misalkan $f(x) = x_1^2 + x_2^2$. Kita punya $\nabla f(x) = [2x_1, 2x_2]^T = 0$ jika dan hanya jika $x = [0,0]^T$. Untuk semua $x \in \mathbb{R}^2$, kita punya

$$F(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0$$

Titik $x = [0,0]^T$ memenuhi FONC, SONC, dan SOSC. Ini adalah peminimum lokal yang tepat. Sebenarnya $x = [0,0]^T$ adalah peminimum global yang tepat. Gambar 6.8 menunjukkan bahwa grafik dari $f(x) = x_1^2 + x_2^2$.

Dalam bab ini, kami menyajikan dasar teoritis untuk solusi masalah tak terbatas nonlinier. Dalam bab-bab berikut, kita akan membahas metode iteratif untuk memecahkan masalah seperti itu. Metode seperti itu sangat penting dalam praktik. Memang, misalkan seseorang dihadapkan dengan fungsi nonlinier dari 20 variabel. Kemudian, FONC membutuhkan solusi dari 20 persamaan simultan nonlinier untuk 20 variabel. Persamaan ini, karena nonlinier, biasanya memiliki banyak solusi. Selain itu, kita harus menghitung 210 detik turunan (disediakan $f \in C^2$) untuk menggunakan SONC atau SOSC. Kami memulai diskusi kami tentang metode iteratif di bab berikutnya dengan metode pencarian untuk fungsi dari satu variabel.

BAB 7

Optimasi Tanpa Kendala

7.1 Fungsi Univariat

Masalah optimasi sederhana tanpa kendala mungkin adalah pencarian maksimum atau minimum dari fungsi univariate $f(x)$ untuk $-\infty < x < +\infty$ (atau di seluruh domain bilangan real), secara sederhana dituliskan sebagai berikut

$$\max \text{ or } \min f(x), \quad x \in \mathbb{R}$$

untuk masalah optimasi tanpa kendala, optimalitas terjadi pada titik kritis yang diberikan oleh kondisi stasioner $f'(x) = 0$. Namun, kondisi stasioner ini hanya syarat perlu saja, tetapi bukan syarat cukup. Jika $f'(x_*) = 0$ dan $f''(x_*) > 0$, itu adalah minimum lokal. Sebaliknya, jika $f'(x_*) = 0$ dan $f''(x_*) < 0$, maka itu adalah maksimum lokal.

Namun, jika $f'(x_*) = 0$ tetapi $f''(x)$ adalah indefinite (keduanya positif dan negatif) saat $x \rightarrow x_*$, maka x_* sesuai dengan titik balik. Sebagai contoh, $f(x) = x^3$ memiliki titik balik $x_* = 0$ karena $f'(0) = 0$ tetapi f'' perubahan tanda dari $f''(0+) > 0$ ke $f''(0-) < 0$.

Namun, secara umum, suatu fungsi dapat memiliki beberapa titik stasioner. Dalam urutan untuk menemukan maksimum atau minimum global, kita mungkin harus melalui setiap titik stasioner, kecuali fungsi objektif adalah konveks.

Perlu ditunjukkan bahwa notasi $\arg \min$ atau $\arg \max$ digunakan di beberapa buku teks, sehingga optimasi di atas dapat ditulis sebagai

$$\arg \max_{x \in \mathbb{R}} f(x), \quad \arg \min_{x \in \mathbb{R}} f(x)$$

Notasi ini menekankan pada argumen x sehingga tugas optimasi adalah menemukan titik dalam domain $f(x)$ yang memaksimalkan (atau meminimalkan) nilai-nilai fungsi. Di sisi lain, notasi yang kami gunakan dalam Persamaan umum optimasi (\max or $\min f(x)$) menekankan nilai maksimum atau minimum dari fungsi tujuan. Kedua jenis notasi digunakan dalam literatur.

Contoh. Untuk menemukan nilai minimum dari $f(x) = x^2 e^{-x^2}$, kita memiliki kondisi stasioner $f'(x) = 0$ atau

$$f'(x) = 2x \times e^{-x^2} + x^2 \times (-2x)e^{-x^2} = 2(x - x^3)e^{-x^2} = 0$$

Ketika $e^{-x^2} > 0$, kita memiliki

$$x(1 - x^2) = 0, \quad x = 0, \quad x = \pm 1$$

Turunan keduanya adalah

$$f''(x) = 2e^{-x^2}(1 - 5x^2 + 2x^4)$$

Yang mana fungsi genap dengan memperhatikan x

Jadi pada $x = \pm 1$, $f''(\pm 1) = 2[1 - 5(\pm 1)^2 + 2(\pm 1)^4]e^{-(\pm 1)^2} = 4e^{-1} < 0$. Kemudian, terdapat dua maksimum yang terjadi pada $x_* = \pm 1$ dengan $f_{\max} = e^{-1}$. Pada $x = 0$, kita memiliki $f''(0) = 2 > 0$, sehingga minimum dari $f(x)$ terjadi pada $x_* = 0$ dengan $f_{\min}(0) = 0$

Contoh. Untuk $f(x) = 3x^4 - 20x^3 - 24x^2 + 240x + 400$ dimana $-\infty < x < +\infty$, adalah stasioner atau memiliki kondisi optimal

$$f'(x) = 12x^3 - 60x^2 - 48x + 240 = 0$$

Yang mana, tampak tidak mudah untuk diselesaikan dengan cara analitik, namun, dapat dituliskan ulang sebagai

$$f'(x) = 12(x + 2)(x - 2)(x - 5) = 0$$

Sehingga terdapat tiga solusi $x_* = -2, +2, \text{ dan } +5$. Turunan keduanya adalah

$$f''(x) = 36x^2 - 120x - 48$$

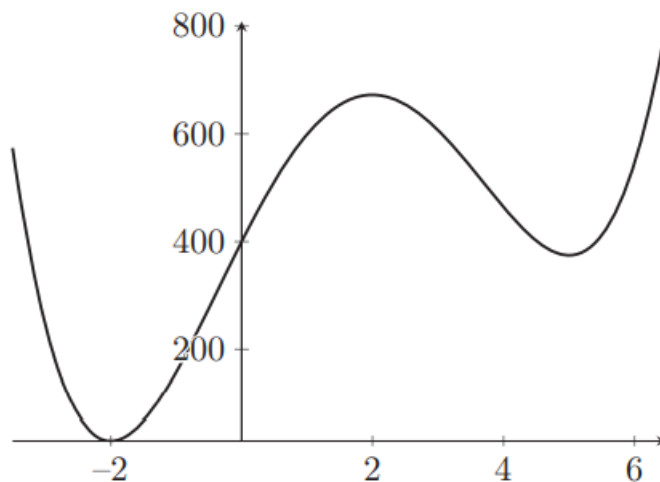
Pada $x = 2$, kita memiliki $f(2) = 672$ dan $f''(2) = -144$, sehingga titik ini adalah maksimum lokal

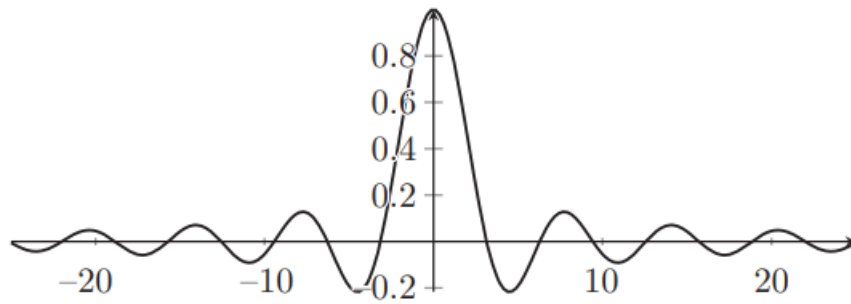
Pada $x = 5$, kita memiliki $f(5) = 375$ dan $f''(5) = 252$, sehingga titik ini adalah minimum lokal

Pada sisi lain, saat $x = -2$, kita memiliki $f''(-2) = 336$ dan $f(-2) = 32$, sehingga titik ini adalah lokal minimum. Bandingkan dua minimum pada $x^* = -2$ dan $x^* = 5$, kita dapat simpulkan bahwa minimum global terjadi saat $x = -2$ dengan $f_{\min} = 32$.

Karena ada satu maksimum lokal saat $x^* = 2$, dapatkah kita simpulkan bahwa maksimum global adalah $f_{\max} = 672$?, jawabannya adalah tidak. Jika kita lihat titik lain seperti $x = 7$ atau $x = -5$, kita memiliki $f(7) = 1247$ dan $f(-5) = 2975$, yang mana lebih besar dari 672, sehingga 672 tidak dapat menjadi maksimum global. Faktanya, fungsi ini adalah tidak terbatas dan dengan demikian maksimumnya adalah $+\infty$.

Seluruh sifat ini menjadi jelas jika melihat gambar dari $f(x)$ seperti yang ditunjukkan dalam gambar berikut, (fungsi semacam ini sering disebut sebagai multimodal)





Maksimum suatu fungsi $f(x)$ dapat diubah menjadi minimum dari $A - f(x)$, dimana A biasanya bilangan positif (walaupun juga bisa $A = 0$). Sebagai contoh, kita tahu bahwa maksimum dari $f(x) = e^{-x^2}$, untuk $x \in (-\infty, \infty)$, adalah 1 saat $x_* = 0$. Masalah ini dapat diubah menjadi masalah minimum $-f(x)$. Untuk alasan ini, masalah optimasi dapat dinyatakan sebagai minimalisasi atau maksimalisasi, tergantung pada konteks formulasi.

Untuk masalah optimasi fungsi sederhana dengan satu variabel bebas, prinsip matematika mungkin mudah untuk dipahami, dan optimalitas terjadi pada $f'(x) = 0$ (titik stasioner) atau pada batas (limit sederhana interval). Namun, mungkin tidak mudah untuk menemukan solusi optimal yang sebenarnya, bahkan untuk fungsi yang tampaknya sederhana sekalipun seperti $\text{sinc}(x) = \sin(x)/x$.

Contoh. Untuk $f(x) = \sin(x)/x$, memiliki jumlah minimum dan maksimum lokal yang tak terhingga dengan maksimum global $f_{\max} = 1$ yang terjadi pada $x = 0$. Namun, dengan menggunakan metode analitik mungkin tidak begitu mudah

Diketahui

$$f'(x) = \frac{\cos(x)}{x} - \frac{\sin(x)}{x^2} = 0, \quad (x \neq 0)$$

Yang mengarah ke

$$\tan(x) = x, \quad (x \neq 0)$$

Tetapi tidak ada formula eksplisit untuk solusi ini, kecuali untuk beberapa aproksimasi. Bahkan kita dapat menyelesaikannya secara numerik untuk menemukan beberapa akar, tetapi kita tidak dapat menemukan semua akarnya (banyak tak terhingga).

Selain itu, akar-akar ini tidak memberikan indikasi yang jelas bahwa $x = 0$ sesuai dengan maksimum global. Nyatanya, maksimum pada $x = 0$ hanya dapat diperoleh dengan metode lain seperti mengambil limit dari bentuk alternatif atau dengan beberapa integral kompleks.

Seperti yang telah kita lihat dari contoh ini, metode numerik harus digunakan untuk menemukan titik optimal yang sebenarnya. Ini menyoroti masalah utama: bahkan kita tahu teori dasar optimasi, itu mungkin tidak secara langsung banyak membantu dalam memecahkan kelas masalah tertentu seperti masalah optimasi multimodal yang sangat nonlinier. Faktanya, metode analitik hanya dapat menyelesaikan sejumlah kecil masalah. Untuk sebagian besar masalah, algoritma numerik menjadi sangat penting.

7.2 Fungsi Multivariat

Untuk fungsi multivariat $f(x)$ dimana $x = (x_1, \dots, x_n)^T$, optimasinya dapat dituliskan mirip seperti masalah optimasi univariat

$$\max \text{ atau } \min f(x), \quad x \in \mathbb{R}^n$$

Disini, kita telah menggunakan notasi \mathbb{R}^n untuk menyatakan bahwa vektor x berada dalam ruang berdimensi n yang mana setiap komponen x_i adalah bilangan real. Yakni, $-\infty < x_i < +\infty$ untuk $i = 1, 2, \dots, n$.

Untuk fungsi $f(x)$, kita dapat memperluasnya secara lokal menggunakan deret Taylor mengenai suatu titik $x = x_*$ sehingga $x = x_* + \epsilon u$,

$$f(x + \epsilon u) = f(x_*) + \epsilon u G(x_*) + \frac{1}{2} \epsilon^2 u^T H(x_*) u + \dots,$$

Dimana G dan H masing-masing adalah vektor gradien dan matriks Hessian. ϵ adalah suatu parameter kecil dan u adalah vektor. Misalnya, fungsi kuadrat umum

$$f(x) = \frac{1}{2} x^T A x + k^T x + b$$

Dimana A adalah konstanta matriks persegi, k adalah vektor gradien dan b adalah vektor konstanta, kita memiliki

$$f(x_* + \epsilon u) = f(x_*) + \epsilon u^T k + \frac{1}{2} \epsilon^2 u^T A u + \dots,$$

Dimana

$$f(x_*) = \frac{1}{2} x_*^T A x_* + k^T x_* + b$$

Jadi, untuk mempelajari sifat lokal dari fungsi kuadrat, perlu mempelajari G dan H . Selain itu, untuk mempermudah, kita dapat mengambil $b = 0$ karena itu adalah vektor konstan.

Pada titik stasioner x_* , turunan pertama adalah nol atau $G(x_*) = 0$, oleh karena itu, persamaan

$$f(x_* + \epsilon u) = f(x_*) + \epsilon u^T k + \frac{1}{2} \epsilon^2 u^T A u + \dots$$

Menjadi

$$f(x_* + \epsilon u) \approx f(x_*) + \frac{1}{2} \epsilon^2 u^T H u$$

Jika $H = A$, maka

$$A v = \lambda v$$

Bentuk masalah nilai eigen. Untuk matriks A $n \times n$, akan ada n nilai eigen $\lambda_j (j = 1, \dots, n)$ dengan n vektor eigen yang sesuai v . Seperti yang telah kita lihat sebelumnya, bahwa A adalah simetris, vektor-vektor eigen ini ortonormal atau dapat diubah menjadi ortonormal, yakni

$$v_i^T v_j = \delta_{ij}$$

Dekat setiap titik stasioner x_* , jika kita mengambil $u_j = v_j$ sebagai koordinat lokal dari sistem, kita kemudian memiliki

$$f(x_* + \epsilon v_j) = f(x_*) + \frac{1}{2} \epsilon^2 \lambda_j$$

Yang berarti bahwa variasi dari $f(x)$, ketika x bergerak menjauh dari titik stasioner x_* sepanjang arah v_j , dicirikan oleh nilai eigen. Jika $\lambda_j > 0, |\epsilon| > 0$ akan mengarah ke $|\Delta f| = |f(x) - f(x_*)| > 0$. Dengan kata lain, $f(x)$ akan meningkat saat $|\epsilon|$ menurun. Sebaliknya jika $\lambda_j < 0$, $f(x)$ akan menurun saat $|\epsilon| > 0$ meningkat. Jelas, dalam kasus khusus $\lambda_j = 0$, fungsi $f(x)$ akan tetap konstan sepanjang arah yang sesuai dari v_j

BAB 8

Metode Pencarian Satu Dimensi

8.1 Pengantar Plot Grafik *Non Linear*

```
import numpy as np
import matplotlib.pyplot as plt

x=np.linspace(-2,2,1000)
#four examples of nonlinear functions
f1=x**2-x-1
f2=x**3-3*np.sin(x)
f3=np.exp(x)-2
f4=1-x**2+np.sin(50 / (1 + x**2))

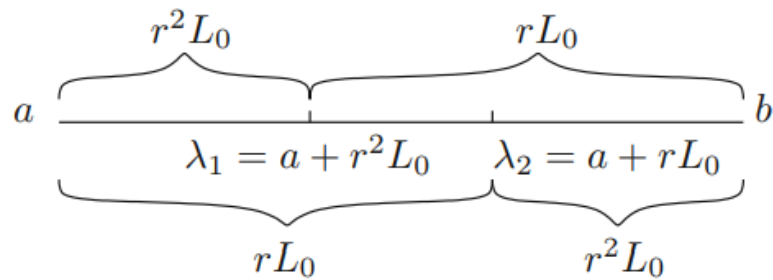
# plot each function
fig, axes = plt.subplots(1, 4, figsize=(12, 3), sharey=True)

for n, f in enumerate([f1, f2, f3, f4]):
    axes[n].plot(x, f, lw=1.5)
    axes[n].axhline(0, ls=':', color='r')
    axes[n].set_ylim(-5, 5)
    axes[n].set_xticks([-2, -1, 0, 1, 2])
    axes[n].set_xlabel(r'$x$', fontsize=18)

lab=[r'$f(1)$',r'$f(2)$',r'$f(3)$',r'$f(4)$']
for n, ylabel in enumerate(lab):
    axes[n].set_ylabel(ylabel, fontsize=18)

titles = [r'$f(x)=x^2-x-1$', r'$f(x)=x^3-3\sin(x)$',
          r'$f(x)=\exp(x)-2$', r'$f(x)=\sin\left(50/(1+x^2)\right)+1-x^2$']
for n, title in enumerate(titles):
    axes[n].set_title(title)
```

8.2 Metode Golden Section Search



Algoritma :

Diberikan interval $[a, b]$ dan tentukan akurasi/toleransi ε ; kemudian atur $i = 0$; $L_0 = b - a$, dan lakukan langkah berikut :

- 1) Atur $\lambda_1 = a + r^2 L_0$; $\lambda_2 = a + r L_0$
- 2) Menghitung $F(\lambda_1)$ dan $F(\lambda_2)$; atur $i = i + 1$
- 3) Jika $F(\lambda_1) > F(\lambda_2)$ maka
Atur $a = \lambda_1$; $\lambda_1 = \lambda_2$; $L_i = (b - a)$; dan $\lambda_2 = a + r L_i$
Else
Atur $b = \lambda_2$; $\lambda_2 = \lambda_1$; $L_i = (b - a)$; dan $\lambda_1 = a + r^2 L_i$
- 4) Jika $L_i < \varepsilon$ maka
Atur $\lambda^* = \frac{b+a}{2}$; hitung $F(\lambda^*)$ dan stop
Else kembali ke langkah 2

Dengan menggunakan Bahasa python :

```
import math
import numpy as np
def gsm(a,b,tol):
    r=0.618
    lam1=a+(r**2)*(b-a)
    lam2=a+r*(b-a)
    i=0
    L=b-a
    while abs(L) > tol:
```

```

f1=- (lam1) *math.cos (lam1)
f2=- (lam2) *math.cos (lam2)
if f1>f2:
    a=lam1
    lam1=lam2
    L=b-a
    lam2=a+r*L
else:
    b=lam2
    lam2=lam1
    L=b-a
    lam1=a+(r**2)*L
i=i+1
hasil = (a+b)/2
fhasil=- (hasil) *math.cos (hasil)
return (hasil,fhasil)

```

```
gsm(0,math.pi/2,0.01)
```

Dengan menggunakan R Studio :

```

min<-function(f,a,b,tol=0.0000001){
  rasio<-(sqrt(5)-1)/2
  x<-a+(rasio^2)*(b-a)
  y<-a+rasio*(b-a)
  f1<-f(x)
  f2<-f(y)
  while(abs(b-a)>tol) {
    if(f1>f2){
      a<-x
      x<-y
      f1<-f2
      y<-a+rasio*(b-a)
      f2<-f(y)
    } else{
      b<-y
      y<-x
      f2<-f1
      x<-a+(rasio^2)*(b-a)
      f1<-f(x)
    }
  }
}

```

```

    }
  }
  return ( (b+a) /2)
}

f <- function(x) {
  -x*cos(x)
}

nilai_x=min(f,0,pi/2)
nilai_x
f(nilai_x)

```

8.3 Barisan Fibonacci

Fibonacci adalah barisan yang diawali dengan angka dari 0 dan 1, kemudian angka berikutnya didapat dengan cara menambahkan kedua bilangan yang berurutan sebelumnya. Dengan aturan ini, maka barisan bilangan Fibonacci yang pertama adalah:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, 2584, 4181, 6765, 10946,...

- 1) Mulai
- 2) Masukkan n, yaitu banyaknya bilangan fibonacci yang diinginkan
- 3) Tentukan nilai i untuk suku pertama dan kedua yaitu $i=0$ atau $i=1$
- 4) Jika $i=0$ atau $i=1$ cetak $\text{fibonacci}(i)=i$
- 5) Isi nilai $i>1$ dimana $i\leq n$
- 6) Tambahkan nilai pada variabel fibonacci di bilangan ke $(i-1)$ dengan nilai pada variabel fibonacci di bilangan ke $(i-2)$ kemudian nyatakan ke variabel fibonacci pada indeks ke i .
- 7) Cetak nilai variabel $\text{fibonacci}(i)$
- 8) Tambahkan i dengan 1
- 9) Tanya apakah pengguna masih ingin menampilkan deret fibonacci
- 10) Bila iya lanjut ke langkah 5

11) Bila tidak, maka program dapat diakhiri

12) Selesai

Algoritma Menentukan Deret Bilangan Fibonacci

Deklarasi:

```
i, n, fibonacci(i): integer
```

Deskripsi:

```
input n
if i=0 atau i=1 then
cetak "fibonacci(i)=i"
else
while i>1 dan i≤n do
fibonacci(i)=fibonacci(i-1)+fibonacci(i-2)
cetak fibonacci(i)
i=i+1
end
```

Dengan menggunakan Python:

```
# Program untuk menampilkan barisan Fibonacci hingga suku ke-n

nterms = int(input("Banyaknya Suku ? "))

# dua suku pertama
n1, n2 = 0, 1
count = 0

# periksa jika banyaknya suku valid
if nterms <= 0:
    print("Mohon Inputkan Bilangan Bulat")
# if there is only one term, return n1
elif nterms == 1:
    print("Barisan Fibonacci hingga ke", nterms, ":")
    print(n1)
# menghasilkan deret fibonacci
else:
    print("Barisan Fibonacci:")
    while count < nterms:
        print(n1)
        nth = n1 + n2
```



```

# update nilai
n1 = n2
n2 = nth
count += 1

```

Dengan menggunakan R Studio:

```

barisan=function(n) {
  fib=c()
  fib[1]=1
  fib[2]=1
  if(n<3){
    return(fib)
  }
  else{
    for(i in 3:n){
      fib[i]=fib[i-1]+fib[i-2]
    }
    return(fib)
  }
}

```

```

baris=barisan(50);baris
# Rasio bilangan Fibonacci konvergen menuju 1.618 (Rasio Emas)
n=c(2:50)
rasio=(baris[n]+baris[n-1])/baris[n];rasio

```

8.4 Metode Biseksi/Metode Bagi Dua

```

import numpy as np
import matplotlib.pyplot as plt

# mendefinisikan fungsi, toleransi, dan interval awal [a, b]
f = lambda x: x**2-4*np.sin(x)
tol = 0.02
a, b = 1,2
x = np.linspace(-2.1, 2.1, 1000)

# menggambar fungsi f
fig, ax = plt.subplots(1, 1,figsize=(12, 4))

ax.plot(x, f(x), lw=1.5)

```

```

ax.axhline(0, ls=':', color='k')
ax.set_xticks([-2, -1, 0, 1, 2])
ax.set_xlabel(r'$x$', fontsize=18)
ax.set_ylabel(r'$f(x)$', fontsize=18)

# menemukan akar menggunakan metode biseksi dan visualisasi
# langkah-langkah dalam metode dengan grafik
fa, fb = f(a), f(b)

ax.plot(a, fa, 'ko')
ax.plot(b, fb, 'ko')
ax.text(a, fa + 0.5, r"$a$",
        ha='center', fontsize=18)
ax.text(b, fb + 0.5, r"$b$",
        ha='center', fontsize=18)

n = 1
while b - a > tol:
    m = a + (b - a)/2
    fm = f(m)

    ax.plot(m, fm, 'ko')
    ax.text(m, fm - 0.5, r"$m_{}d$" % n,
            ha='center')
    n += 1
    if np.sign(fa) == np.sign(fm):
        a, fa = m, fm
    else:
        b, fb = m, fm

ax.plot(m, fm, 'r*', markersize=10)
ax.annotate("Akar aproksimasi ada pada %.3f" % m,
           fontsize=14, family="serif",
           xy=(a, fm), xycoords='data',
           xytext=(-150, +50), textcoords='offset points',
           arrowprops=dict(arrowstyle="->",
                           connectionstyle="arc3, rad=-.5"))
ax.set_title("Metode Biseksi / Metode Bagi Dua")

```

8.5 Metode Newton

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Contoh

$$f(x) = x^3 - 2x^2 + 3x - 6$$

$$f'(x) = 3x^2 - 4x + 3$$

Dengan nilai awal $x_0 = 3$

Iterasi pertama, menentukan nilai x_1

$$x_0 = 3 \rightarrow f(x_0) = f(3) = 3^3 - 2(3)^2 + 3(3) - 6 = 12$$

$$f'(x_0) = f'(3) = 3(3)^2 - 4(3) + 3 = 18$$

$$k = 0 \rightarrow x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

$$x_{0+1} = x_0 - \frac{f(x_0)}{f'(x_0)}$$

$$x_1 = 3 - \frac{12}{18} = 2.33333333$$

Iterasi kedua, menentukan nilai x_2

$$x_1 = 2.33333333 \rightarrow f(x_1) = f(2.33333333) = 2.814814815$$

$$f'(x_1) = f'(2.33333333) = 10$$

$$k = 1 \rightarrow x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

$$x_{1+1} = x_1 - \frac{f(x_1)}{f'(x_1)}$$

$$x_2 = 2.33333333 - \frac{2.814814815}{10} = 2.05185$$

Iterasi ketiga, menentukan nilai x_3

$$x_2 = 2.05185 \rightarrow f(x_2) = f(2.05185) = 0.3738431$$

$$f'(x_2) = f'(2.05185) = 7.422865$$

$$k = 2 \rightarrow x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

$$x_{2+1} = x_2 - \frac{f(x_2)}{f'(x_2)}$$

$$x_3 = 2.05185 - \frac{0.3738431}{7.422865} = 2.001486$$

Iterasi keempat, menentukan nilai x_4

$$x_3 = 2.001486 \rightarrow f(x_3) = f(2.001486) = 0.01041084$$

$$f'(x_3) = f'(2.001486) = 7.011895$$

$$k = 3 \rightarrow x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

$$x_{3+1} = x_3 - \frac{f(x_3)}{f'(x_3)}$$

$$x_4 = 2.001486 - \frac{0.01041084}{7.011895} = 2.000001$$

Iterasi kelima, menentukan nilai x_5

$$x_4 = 2.000001 \rightarrow f(x_4) = f(2.000001) \approx 0.000007$$

$$f'(x_4) = f'(2.000001) \approx 7.000008$$

$$k = 4 \rightarrow x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

$$x_{4+1} = x_3 - \frac{f(x_3)}{f'(x_3)}$$

$$x_5 = 2.000001 - \frac{0.000007}{7.000008} = 2$$

Dengan menggunakan Python:

```
from sympy import *
import numpy as np
import matplotlib.pyplot as plt

# definisikan fungsi, pilih toleransi, dan nilai awal xk
tol = 0.001
xk = 30
```

```

sx = symbols("x")
s_f = sx**3-2*sx**2+3*sx-6

f = lambda x: lambdify(sx, s_f, 'numpy')(x)
fp = lambda x: lambdify(sx,diff(s_f, sx), 'numpy')(x)

x = np.linspace(-35, 35, 1000)

# grafik untuk visualisasi langkah-langkah pencarian akar
fig, ax = plt.subplots(1, 1, figsize=(12,9))

ax.plot(x, f(x))
ax.axhline(0, ls=':', color='k')
# iterasi metode Newton sampai konvergen ke toleransi tercapai
n = 0
while f(xk)>tol:
    xk_new = xk-f(xk)/fp(xk)

    ax.plot([xk, xk], [0, f(xk)], color='k', ls=':')
    ax.plot(xk, f(xk), 'ko')
    ax.text(xk, -.5, r'$x_{%d}$' % n, ha='center')
    ax.plot([xk, xk_new], [f(xk), 0], 'k-')

    xk = xk_new
    n += 1

ax.plot(xk, f(xk), 'r*', markersize=15)
ax.annotate("Akar aproksimasi ada pada %.3f" % xk,
           fontsize=14, family="serif",
           xy=(xk, f(xk)), xycoords='data',
           xytext=(-150, +50), textcoords='offset points',
           arrowprops=dict(arrowstyle="->",
                           connectionstyle="arc3, rad=-.5"))

ax.set_title("Metode Newton")
ax.set_xlim([-35,35])

```

BAB 9

Algoritma Mesin Learning

9.1 Data Mining

Data mining telah berkembang melampaui teknik pemodelan data tradisional seperti model statistik dan metode regresi. Data mining sekarang juga mencakup pengelompokan dan klasifikasi, pemilihan fitur dan ekstraksi fitur, dan teknik pembelajaran mesin seperti metode pohon keputusan, model Markov hidden, ANN dan SVM. Metode pengelompokan dan klasifikasi kaya dengan spektrum metode yang luas. Metode K-means dasar untuk pengelompokan dan mesin SVM untuk klasifikasi. ANN adalah metode dengan variasi dan varian yang berbeda. ANN dapat memiliki banyak aplikasi dalam rentang yang beragam bidang, termasuk pengelompokan, klasifikasi, pembelajaran mesin, kecerdasan komputasi, ekstraksi fitur dan seleksi, dan lain-lain.

9.1.1 Hierarchy Clustering

Untuk sekumpulan n pengamatan tertentu, tujuannya adalah untuk membaginya menjadi beberapa kluster (misalnya, k kluster yang berbeda) untuk meminimalkan ukuran atau tujuan pengelompokan tertentu. Ada banyak masalah utama penggunaan Hierarchy Clustering ini. Pertama, tidak tahu berapa banyak cluster yang mungkin dimiliki data secara intrinsik. Kedua, kumpulan data bisa sangat besar ($n \gg 1$, sebagai contoh, $n = O(10^9)$ atau bahkan $n = O(10^{18})$). Ketiga, data mungkin tidak cukup bersih (seringkali dengan informasi yang tidak berguna dan/atau data yang salah). Akhirnya, data bisa tidak lengkap, dan dengan demikian mungkin kekurangan informasi yang diperlukan untuk pengelompokan yang benar. Selain itu, juga ada masalah lain seperti faktor waktu, data tidak terstruktur, dan metrik jarak.

Pengelompokan Hierarchy biasanya bekerja dengan baik untuk kumpulan data kecil. Ini dimulai dengan setiap titik dalam clusternya sendiri (yaitu, $k=n$ untuk n titik data), diikuti dengan prosedur iteratif sederhana:

- 1) Setiap titik milik cluster sendiri $k=n$.
- 2) Sementara (kriteria berhenti),
- 3) Pilih 2 cluster terdekat untuk digabungkan menjadi 1 cluster,
- 4) Perbarui $k \leftarrow k - 1$;
- 5) Ulangi sampai ukuran metrik naik.
- 6) Selesai.

Prosedur iteratif ini dapat menghasilkan satu cluster besar $k=1$ pada akhirnya. Tapi itu menghasilkan pohon keputusan yang kompleks, yang memberikan ringkasan informatif dan beberapa wawasan tentang struktur dan hubungan di dalam data. Namun, jika metrik jarak seperti metrik Euclidean didefinisikan dengan benar, metrik akan mulai berkurang pada tahap awal ketika dua cluster digabungkan. Pada tahap akhir, metrik ini biasanya mulai meningkat, yang meruokan indikasi untuk berhenti dan jumlah cluster dapat menjadi jumlah cluster yang sebenarnya. Namun, ini tidak langsung dalam praktiknya, dan mungkin tidak ada nilai k unik sama sekali. Faktanya k adalah hyper-parameter, yang membutuhkan beberapa studi penyetelan dan parameter dalam praktiknya.

9.1.2 K-Means Clustering

Tujuan utama dari metode k-means clustering adalah untuk membagi satu set n pengamatan menjadi k cluster yang berbeda sedemikian rupa sehingga setiap titik dalam cluster terdekat dengan jarak terpendek ke mean cluster atau centroid yang sesuai.

Misalkan kita memiliki n titik pengamatan x_1, x_2, \dots, x_n dalam ruang vektor berdimensi d , tujuan kita adalah membagi pengamatan ini menjadi k cluster (S_1, S_2, \dots, S_k) dengan rata-rata centroid $(\xi_1, \xi_2, \dots, \xi_k)$ sehingga jumlah kuadrat berdasarkan cluster, juga disebut kuadrat dalam cluster, dapat diminimalkan. Itu adalah,

$$\text{Minimize } \sum_{j=1}^k \sum_{x_i \in S_j} \|x_i - \xi_j\|^2,$$

Dimana $1 < k \leq n$ dan biasanya $k \ll n$.

Metode k-means untuk membagi titik menjadi k cluster ini dapat diringkas secara sistematis sebagai berikut:

- 1) Pilih secara acak k titik sebagai centroid awal dari k cluster.
- 2) Untuk setiap sisa titik i .
- 3) Tetapkan i ke cluster dengan centroid terdekat.
- 4) Perbarui centroid dari cluster itu (berisi i);
- 5) Selesai .

Ada beberapa masalah utama mengenai metode k-means ini. Pemilihan titik k sebagai centroid awal tidak efisien. Dalam kasus terburuk, k poin yang dipilih secara acak dapat dimiliki oleh cluster yang sama. Salah satu solusi yang mungkin adalah memilih k titik dengan jarak terbesar satu sama lain. Hal ini sering dilakukan dengan memulai dari titik acak, kemudian mencoba mencari titik kedua yang sejauh mungkin ke titik pertama, dan kemudian mencoba mencari titik ketiga yang sejauh mungkin dari dua titik sebelumnya. Ini berlanjut sampai k poin pertama diinisialisasi. Metode ini merupakan perbaikan dari metode pemilihan acak sebelumnya, namun masih belum ada jaminan bahwa pemilihan titik awal ini akan mengarah pada solusi clustering terbaik. Oleh karena itu, diperlukan semacam restart acak dan beberapa proses.

Untuk kumpulan data tertentu, sulit untuk mengetahui k apa yang harus digunakan karena k adalah parameter hiper. Idealnya, pilihan awal k harus cukup dekat dengan jumlah aktual kelas intrinsik dalam data, dan kemudian bervariasi di sekitar tebakan awal. Namun dalam praktiknya, ini mungkin memerlukan beberapa pengalaman atau metode lain untuk memahami data. Selain itu, jarak pengelompokan juga dapat digunakan untuk memeriksa apakah k adalah pilihan yang tepat dalam banyak kasus.

9.1.3 Distance Metric (Jarak metrik)

Sebagian besar metode pengelompokan menggunakan jarak Euclidean dan kesamaan Jaccard, meskipun jarak lain seperti jarak edit dan jarak hamming juga banyak digunakan.

Secara singkat, jarak Euclidean $d(x, y)$ antara dua titik data x dan y adalah L_p norma diberikan oleh

$$d(x, y) = \left(\sum_i |x_i - y_i|^p \right)^{1/p} \quad 9.2$$

Dalam kebanyakan kasus, $p=2$ digunakan. Indeks kesamaan Jaccard dari dua himpunan U dan V didefinisikan sebagai:

$$J(U, V) = \frac{|U \cap V|}{|U \cup V|}, \quad 9.3$$

Yang mengarah ke $0 \leq J(U, V) \leq 1$, dan jarak Jaccard didefinisikan sebagai

$$d_J(U, V) = 1 - J(U, V). \quad 9.4$$

Jarak edit antara dua string U dan V adalah jumlah penyisipan dan penghapusan karakter tunggal terkecil yang akan mengubah U menjadi V .

9.2 Data Mining for Big Data

Ilmu big data menjadi semakin penting saat ini, didorong oleh internet, media sosial, dan internet of things. Banyak aplikasi sekarang didorong oleh data secara dinamis. Dibandingkan dengan database tradisional dan analitik data, data besar memiliki beberapa karakteristik utama dan dengan demikian teknik yang diperlukan untuk mengatasi big data tersebut juga lebih canggih.

9.2.1 Karakteristik Big Data

Karakteristik utama dari big data dapat diringkas sebagai 5V yaitu volume, velocity, variety, value dan veracity.

- Volume: volume data telah meningkat secara dramatis dalam beberapa tahun terakhir, didorong oleh internet, multimedia, dan media sosial.
- Velocity (Kecepatan): tingkat akumulasi data juga meningkat secara dramatis. Misalnya, diperkirakan ada sekitar 20 triliun GB yang ditambahkan setiap tahun.
- Variety (Ragam): ragam data juga beragam dan data dapat terstruktur, tidak terstruktur dari berbagai jenis, sumber, dan media. Misalnya, astronomi digital dapat memiliki kumpulan data besar berupa gambar dan gambar survei langit, tetapi sebagian besar

merupakan data terstruktur. Sebagai perbandingan, data besar dari media sosial, ekonomi digital dan internet of things dapat mengumpulkan sejumlah besar tipe data yang berbeda. Selain itu, dimensi data besar juga bisa tinggi, tergantung banyak faktor, fitur atau variabel yang berbeda, baik yang eksplisit maupun yang tersembunyi.

- Value (Nilai): tujuan utama dari analisis dan pemrosesan adalah untuk mengekstrak beberapa fitur yang berguna untuk mendapatkan wawasan tentang data dan membuat prediksi idelanya, tujuan utamanya adalah untuk memahami data sehingga berpotensi memprediksi kejadian di masa depan.
- Veracity (Kebenaran): tidak peduli seberapa besar datanya, data tersebut dapat menjadi tidak lengkap dengan kesalahan dan ketidakpastian, tunduk pada perubahan dinamis. Meski begitu, kualitas data dan analisis selanjutnya akan memiliki akurasi dan nilai yang berbeda.

9.2.2 Sifat Statistik Big Data

Metode tradisional masih digunakan dengan beberapa modifikasi untuk menangani data besar, termasuk regresi, pohon keputusan, model hidden Markov, neural networks, SVM, dan lainnya. Namun, dari sudut pandang statistik, kita harus menyadari prinsip Bonferroni's ketika berhadapan dengan data besar. Jika Anda mencoba menghitung beberapa jumlah kejadian yang diharapkan dari pola tertentu dan jika angka ini secara signifikan lebih besar dari jumlah contoh nyata yang ingin Anda temukan. Maka Anda dapat menyimpulkan bahwa hampir semua hal yang Anda temukan salah (yang berarti itu adalah artefak dalam arti statistik, bukan bukti aktual). Dengan kata lain, jika Anda mencari terlalu keras untuk pola menarik yang mungkin didukung oleh data Anda, Anda pasti akan menemukan pola yang salah. Oleh karena itu kehati-hatian harus diambil ketika menafsir data.

9.2.3 Mining Big Data

Untuk big data, selain tantangan lain seperti penyimpangan dan pengambilan, salah satu tantangannya adalah data yang akan diproses jauh lebih besar dari memori fisik utama komputer, sehingga tidak mungkin memuat semua data ke dalam komputer, memori utama untuk diproses. Beberapa jenis pengambilan sampel dan pemrosesan segmen demi segmen mungkin diperlukan, sebelumnya menerapkan algoritme apa pun.

Perkembangan terakhir menunjukkan bahwa metode baru mungkin lebih cocok untuk kumpulan data yang besar. Sebagai contoh, algoritma Bradley-Fayya-Reina (BFR) dan algoritma clustering Using Representative (CURE) telah menunjukkan hasil yang baik. Pada intinya, algoritma BFR merupakan perluasan dari k-means ke data berdimensi tinggi. Diasumsikan bahwa data didistribusikan di sekitar centroid menurut distribusi Gaussian dalam ruang Euclidean. Inti dari algoritma BFR dapat diuraikan sebagai berikut:

- 1) Pilih subset kecil (dengan resampling) dari data besar menggunakan metode k-means atau hierarchy untuk menemukan k cluster awal.
- 2) Ambil setiap potongan data (subset) dari kumpulan data besar, lakukan hal berikut
 - Tetapkan titik data dan rangkum cluster (lihat detail di bawah). Kemudian buang datanya.
 - Kompres dan gabungkan titik-titik yang berdekatan satu sama lain (membentuk corn set yang ditekan).
 - Pertahankan titik data yang tidak ditetapkan tau tidak dekat satu sama lain (membentuk himpunan yang dipertahankan).

Untuk setiap titik N tertentu (dalam subset), hitung SUM_i dan $SUMSQ_i$ dimana SUM_i adalah jumlah vektor data dalam dimensi ke- i , dan $SUMSQ_i$ adalah jumlah kuadrat dari semua titik data dalam dimensi ke- i .

Kemudian, centroid dapat dipernarui pada SUM_i/N dalam dimensi ke- i dan variasinya dalam dimensi ke- i dapat diperkirakan sebagai $SUMSQ_i - (SUM_i/N)^2$. Keuntungan dari perhitungan dimensi demi dimensi seperti itu adalah bahwa baik SUM_i dan $SUMSQ_i$ menjadi jumlah sederhana saat menggabungkan dua cluster.

Saat memutuskan apakah titik baru x_i , cukup dekat dengan salah satu k cluster. Dapat menggunakan aturan berikut:

- i. Menambahkan titik x ke cluster jika memiliki centroid paling dekat dengan x_i . Atau
- ii. Menetapkan x_i ke cluster dengan jarak Mahalanobis terkecil.

Jarak Mahalanobis antara $x = (x_1, x_2, \dots, x_n)$ ke centroid cluster $c = (c_1, c_2, \dots, c_n)$ adalah

$$d_m = \sqrt{\sum_{i=1}^n \left(\frac{x_i - c_i}{\sigma_i} \right)^2} \quad (9.5)$$

Dimana σ_i adalah simpangan baku kluster dalam dimensi ke-i. Karena itu, jarak ini adalah jarak skala berbasis varians.

Untuk subset titik data (sepotong dari data besar), perhitungkan rinci adalah sebagai berikut:

- 1) Untuk titik data yang dekat dengan pusat cluster, tambahkan titik-titik ini ke cluster itu. Kemudian, perbarui centroid dan metrik lainnya seperti SUM_i dan $SUMSQ_i$.
- 2) Untuk titik-titik yang tidak dekat dengan centroid manapun, kelompokkan titik-titik tersebut bersama-sama dengan set yang dipertahankan. Kemudian, gabungkan semua mini-cluster bila perlu. Untuk titik yang ditetapkan ke kluster (termasuk kluster mini apa pun), perbarui centroid dan metrik lainnya (lalu buang titik tersebut).
- 3) Pada tahap terakhir (setelah melalui semua subset data atau memuat potongan terakhir data), pascaproses set yang dipertahankan dan det terkompresi dengan menetapkan setiap titik ke kluster centroid terdekat, atau membuangnya sebagai outlier.

Meskipun algoritma BFR efisien, ini terutama untuk data yang simetris di sekitar cluster, sehingga tidak dapat menangani bentuk-S atau cincin secara efektif. Untuk kumpulan data yang rumit seperti itu, dapat menggunakan algoritma lain yang kuat, yang disebut algoritma CURE, yang merupakan algoritma penetapan titik dalam ruang Euclidean tanpa asumsi tentang bentuk cluster (tidak mungkin asumsi distribusi normal dalam algoritma BFR). Hasilnya, algoritma ini dapat menangani pengelompokan data bentuk ganjil seperti tikungan, bentuk S dan cincin.

Langkah-langkah utama dari algoritma CURE dapat diringkas sebagai berikut:

- 1) Pilih sampel kecil (subset kecil) dari data untuk dikelompokkan dalam memori utama menggunakan metode yang baik seperti k-means dan metode hierarchy, yang memberikan cluster awal.
- 2) Pilih satu set kecil titik dari setiap cluster untuk menjadi titik representatif (titik harus sejauh mungkin satu sama lain).

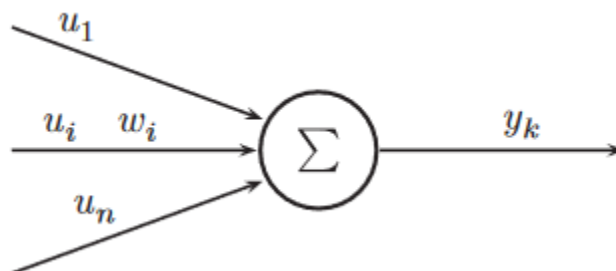
- 3) Pindahkan masing-masing titik perwakilan sebagian kecil (biasanya 20%) dari jarak antara lokasinya ke centroid clusternya.
- 4) Gabungkan dua cluster jika memiliki pasangan titik representasi (satu dari setiap cluster) yang cukup dekat.
- 5) Ulangi langkah diatas dan gabungkan samapai tidak ada lagi cluster yang dekat untuk digabungkan.
- 6) Melaksanakan penugasan titik untuk semua sisa titik dalam data besar.

Agar metode apa pun menjadi efisien dan berguna, matriks besar harus dihindari dan tidak perlu mencoba setiap kombinasi yang mungkin. Metode yang digunakan untuk memecahkan masalah skala besar harus cukup efisien untuk menghasilkan hasil yang baik dalam skala waktu yang dapat diterima secara praktis. Namun, secara umum tidak ada jaminan bahwa optimalitas global dapat ditemukan.

Ada metode lain seperti algoritma GRGPF dan BIRCH untuk mengelompokkan kumpulan data besar. Selain itu, algoritma dan implementasi dapat dipararelkan untuk mempercepat algoritma ini.

9.3 Artificial Neural Networks (Jaringan Syaraf Tiruan)

Ide dasar dari ANN adalah belajar dari data dan membuat prediksi. Untuk satu set data input tertentu, Neural Networks memetakan data input ke dalam beberapa output. Hubungana antara input dan output cukup rumit, dan biasanya tidak mungkin untuk mengungkapkan hubungan tersebut dalam bentuk eksak atau analitis. Dengan membandingkan keluaran dengan keluaran sebenarnya, sistem Neural Networks dapat menyesuaikan bobotnya agar lebih sesuai dengan keluarannya. Jika ada jumlah data yang cukup, jaringan dapat “dilatih” dengan baik dan dengan demikian jaringan yang terlatih bahkan dapat membuat prediksi untuk data baru, yang merupakan inti dari ANN.



Gambar 9.1 Neuron Sederhana

9.3.1 Model Neuron

Neuron buatan dengan n input atau impuls dan output y_k akan aktif jika kekuatan sinyal mencapai ambang θ tertentu. Setiap input memiliki bobot respon yang sesuai w_i (lihat Gambar 9.1). Output dari neuron ini diberikan oleh

$$y_l = \phi(\xi) , \xi = \sum_{i=1}^n w_i u_i \quad (9.6)$$

Dimana jumlah bobot ξ adalah kekuatan sinyal total, dan ϕ adalah apa yang disebut fungsi aktivasi, yang dapat diambil sebagai fungsi langkah. Artinya kita memiliki,

$$\phi(\xi) = \begin{cases} 1 & \text{jika } \xi \geq 0, \\ 0 & \text{jika } \xi < 0 \end{cases} \quad (9.7)$$

Dapat dilihat bahwa output hanya diaktifkan ke nilai bukan nol jika kekuatan sinyal keseluruhan lebih besar dari ambang θ .

Fungsi aktivasi langkah adalah biner dan memiliki diskontinuitas; terkadang, lebih mudah menggunakan fungsi nonlinier, halus untuk aktivasi, yang disebut fungsi Sigmoid

$$S(\xi) = \frac{1}{1 + e^{-\xi}} , \quad (9.8)$$

Yang mendekati 1 sebagai $\xi \rightarrow \infty$, dan menjadi 0 sebagai $\xi \rightarrow -\infty$. Proporsi yang menarik erty dari fungsi ini adalah

$$S'(\xi) = S(\xi)[1 - S(\xi)]. \quad (9.9)$$

Model aktivasi neuron lain yang berguna adalah model linier yang diperbaiki, juga disebut unit linier yang diperbaiki (ReLU), yang dapat ditulis sebagai



$$y(\xi) = \xi^{+1} = \max\{0, \xi\} = \begin{cases} \xi, & \text{jika } \xi > 0, \\ 0, & \text{sebaliknya} \end{cases}$$

(9.10)

Fungsi ini dapat didekati dengan fungsi yang lebih halus

$$y(\xi) = \log(1 + e^\xi) \tag{9.11}$$

Yang turunannya adalah $y'(\xi) = e^\xi / (1 + e^\xi) = S(\xi)$.

Sebenarnya, ada banyak model neuron lain untuk mendekati dan memperlancar fungsi aktivasi, termasuk model linier terkoneksi stokastik, tangen hiperbolik, model linier eksponensial, dan lain-lain. Misalnya tangen hiperbolik dapat ditulis sebagai

$$\tanh(\xi) = \frac{e^\xi - e^{-\xi}}{e^\xi + e^{-\xi}} = \frac{1 - e^{-2\xi}}{1 + e^{-2\xi}} \tag{9.12}$$

Yang mendekati +1 sebagai $\xi \rightarrow \infty$ dan -1 sebagai $\xi \rightarrow -\infty$

9.3.2 Neural Networks

Konstruksi Neural Networks melibatkan estimasi bobot yang sesuai dari sistem jaringan dengan beberapa set data training. Tugas training adalah menemukan bobot yang sesuai, sehingga Neural Networks tidak hanya dapat menyesuaikan data yang diketahui dengan baik, tetapi juga dapat memprediksi output untuk input yang baru. ANN yang baik harus dapat meminimalkan kedua kesalahan secara bersamaan- kesalahan fitting/learning dan kesalahan prediksi.

Kesalahan dapat didefinisikan sebagai perbedaan antara keluaran v_k yang dihitung (atau diprediksi), dan keluaran nyata y_k , untuk semua N neuron keluaran dalam kuadrat terkecil.

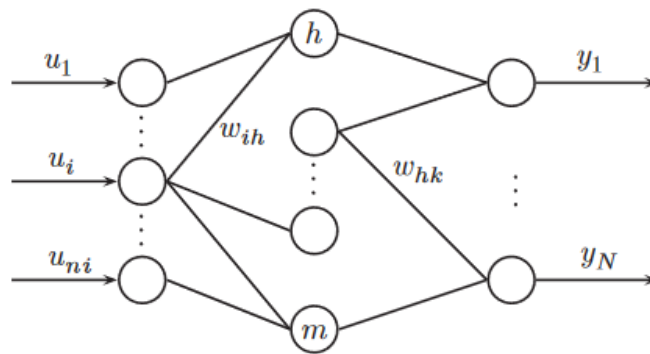
$$E = \frac{1}{2} \sum_{k=1}^N (v_k - y_k)^2 \tag{9.13}$$

Disini output v_k adalah fungsi dari input/aktivasi dan bobot. Ini juga dapat ditulis sebagai

$$E = \frac{1}{2} \|v - y\|_2^2, \quad (9.14)$$

Dimana v adalah vektor keluaran dan y adalah vektor keluaran nyata atau yang diinginkan. Disini, fungsi error E sering disebut sebagai fungsi loss dari ANN. Tujuan utama dari training Neural Networks adalah untuk meminimalkan kerugian atau training error dan juga untuk memaksimalkan akurasi prediksi.

Untuk meminimalkan kesalahan ini, dapat menggunakan teknik minimalisasi standar untuk menemukan solusi dari bobot.



Laapisan masukan (i) lapisan tersembunyi(h) Neuron keluaran(k)

Teknik yang sederhana namun efisien adalah metode penurunan paling curam. Untuk setiap bobot acak awal, kenaikan bobot untuk w_{hk} adalah

$$\Delta w_{hk} = -\eta \frac{\partial E}{\partial w_{hk}} = -\eta \frac{\partial E}{\partial v_k} \frac{\partial v_k}{\partial w_{hk}} \quad (9.15)$$

Yang mengikuti aturan rantai dasar diferensial. Disini, $0 < \eta \leq 1$ adalah kecepatan learning. Dalam kasus khusus, kita dapat menggunakan $\eta = 1$ untuk diskusi

Dari

$$S_k = \sum_{h=1}^m w_{hk} v_h \quad (k = 1, 2, \dots, N) \quad (9.16)$$

Dan

$$v_k = f(S_k) = \frac{1}{1 + e^{-S_k}}, \quad (9.17)$$

Kita memiliki

$$f' = f(1 - f), \quad (9.18)$$

$$\frac{\partial v_k}{\partial w_{hk}} = \frac{\partial v_k}{\partial S_k} \frac{\partial S_k}{\partial w_{hk}} = v_k(1 - v_k)v_h, \quad (9.19)$$

Dan

$$\frac{\partial E}{\partial v_k} = (v_k - y_k). \quad (9.20)$$

Oleh karena itu, kita mempunyai

$$\Delta w_{hk} = -\eta \delta_k v_h, \quad (9.21)$$

Dimana

$$\delta_k = v_k(1 - v_k)(v_k - y_k) \quad (9.22)$$

Secara umum, berdasarkan Persamaan (9.15), kenaikan untuk setiap bobot w , terkait dengan kesalahan diberikan oleh

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \quad (9.23)$$

Yang dapat ditulis secara ringkas sebagai rumus berulang

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta \nabla E(\mathbf{w}^t) \quad (9.24)$$

Dimana ∇E adalah vektor gradien. Namun, untuk masalah skala besar dengan banyak keluaran, perhitungan vektor gradien bisa sangat mahal, sehingga beberapa rumus iteratif atau prog\pagasi lebih disukai.

Untuk struktur Neural Networks tertentu, ada banyak cara menghitung bobot dengan pembelajaran terawasi. Salah satu metode yang paling sederhana dan banyak digunakan adalah

dengan menggunakan algoritma back propagation untuk melatih Neural Networks, yang sering disebut dengan back propagation neural networks (BPNN).

9.3.3 Algoritma Back Propagation

Ide dasar dari BPNN adalah memulai dari lapisan keluaran dan menyebar ke belakang untuk memperkirakan dan memperbarui bobot (lihat Algoritma 9.1). Dari sembarang matriks pembobotan acak awal w_{ih} (untuk menghubungkan simpul masukan ke lapisan tersembunyi) dan w_{hk} (untuk menghubungkan lapisan tersembunyi ke simpul keluaran), kita dapat menghitung keluaran dari lapisan tersembunyi v_h , sebagai

$$v_h = \frac{1}{1 + \exp[-\sum_{i=1}^{n_i} w_{ih} u_i]} \quad (h = 1, 2, \dots, m), \quad (9.25)$$

Dan output untuk mode output

$$v_k = \frac{1}{1 + \exp[-\sum_{h=1}^{n_i} w_{hk} v_h]} \quad (k = 1, 2, \dots, N), \quad (9.26)$$

Kesalahan untuk simpul keluaran diberikan oleh

$$\delta_k = v_k (1 - v_k)(v_k - y_k), \quad (k = 1, 2, \dots, N) \quad (9.27)$$

Algoritma 9.1 Pseudocode dari back propagation neural networks

Inisialisasi matriks bobot w_{ih} dan w_{hk} secara acak

for semua titik data training

while (kesalahan residual yang tidak sedikit)

 Hitung output untuk lapisan tersembunyi v_h menggunakan persamaan (9.25)

 Hitung output untuk lapisan output v_k menggunakan persamaan (9.26)

 Hitung kesalahan δ_k dan δ_h menggunakan persamaan (9.27) dan (9.28)

 Perbarui bobot w_{ih} dan w_{hk} melalui persamaan (9.29)

dan (9.30)

end while
end for

dimana y_k ($k = 1, 2, \dots, N$) adalah data (output nyata) untuk input u_i ($i = 1, 2, \dots, n_i$).
Demikian pula, kesalahan untuk node tersembunyi dapat ditulis sebagai

$$\delta_h = v_h(1 - v_h) \sum_{k=1}^N w_{hk} \delta_k \quad (h = 1, 2, \dots, m) \quad (9.28)$$

Rumus pemuutakhiran bobot pada iterasi t adalah:

$$w_{hk}^{t+1} = w_{hk}^t - \eta \delta_k v_h, \quad (9.29)$$

Dan

$$w_{ih}^{t+1} = w_{ih}^t - \eta \delta_h u_i, \quad (9.30)$$

Dimana $0 < \eta \leq 1$ adalah kecepatan belajar.

Dapat kita lihat bahwa kenaikan berat badan adalah

$$\Delta w_{ih} = -\eta \delta_h u_i \quad (9.31)$$

Dengan rumus pembaruan serupa untuk w_{hk} ,

Rumus di atas dapat ditulis dalam rumus umum berikut:

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta \nabla E(\mathbf{w}^t), \quad E = \frac{1}{2} \sum_{k=1}^N |v_k(\mathbf{w}) - y_k|^2 \quad (9.32)$$

9.3.4 Fungsi Kerugian ANN

Bentuk kerugian lainnya fungsi dalam literatur, yang mungkin merupakan alternatif yang baik untuk tugas-tugas tertentu dan jenis ANN tertentu.

Jika $y = (y_1, y_2, \dots, y_N)^T$ adalah vektor dari nilai yang diprediksi dan $\bar{y} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_N)$ adalah vektor dari nilai sebenarnya, fungsi kerugian L_2 biasanya didefinisikan sebagai:

$$E = \|y - \bar{y}\|_2 = \|\bar{y} - y\|_2 = \sum_{i=1}^N (\bar{y}_i - y_i)^2$$

Ini juga dapat dikonversi ke bentuk kesalahan kuadrat rata-rata sebagai;

$$E_m = \frac{1}{N} \sum_{i=1}^N (\bar{y}_i - y_i)^2$$

Dibandingkan dengan L_2 -norm loss yang mulus, fungsi loss berbasis L_1 diberikan oleh;

$$E_1 = \|y - \bar{y}\|_1 = \sum_{i=1}^N |\bar{y}_i - y_i|$$

Kesalahan atau kerugian rata – rata didefinisikan sebagai berikut ;

$$\frac{1}{N} \sum_{i=1}^N |\bar{y}_i - y_i|$$

Untuk klasifikasi biner, entropi silang E_e dapat digunakan sebagai fungsi kerugian;

$$E_e = -\frac{1}{N} [\bar{y}_i \log y_i + (1 - \bar{y}_i) \log (1 - y_i)]$$

Jadi, entropi silang yang lebih besar berarti perbedaan yang lebih besar juga.

Fungsi kerugian Kullbeck-Leibler didefinisikan sebagai berikut;

$$E_{KL} = \frac{1}{N} \sum_{t=1}^N \bar{y}_i \log(\bar{y}_i) - \frac{1}{N} \sum_{t=1}^N \bar{y}_i \log y_i$$

9.3.5 Turunan Gradien Stokastik

Dalam penurunan ini kita perlu menghitung nilai komponen n dari vector gradient ∇E dengan n bervariasi. SGD ini menggunakan pengamatan tunggal atau contoh data $\nabla E_i (i = 1, 2, \dots, n)$ untuk mengaproksimasi ∇E , yaitu ; $w^{t+1} = w - \eta \nabla E_i$, Estimasi yang lebih baik dapat diperoleh dengan subset dari rata-rata gradient stokastik, atau lainnya seperti;

$$w^{t+1} = w - \eta \nabla E_i + \tau \Delta w,$$

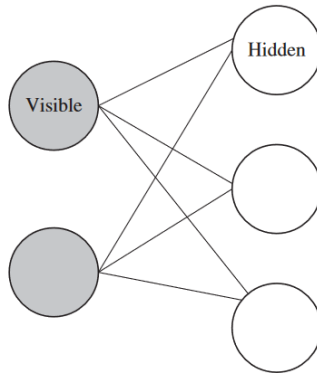
Dimana $0 \leq \tau \leq 1$ adalah control parameter yang mengendalikan momentum, dan ini disebut metode momentum.

9.3.6 Boltzmann Terbatas

Mesin Boltzmann Terbatas adalah alat untuk aplikasi pembelajaran mendalam dengan matematika, yang merupakan dua grup dengan m adalah $v_i = (i = 1, 2, \dots, m)$ dan n adalah $h_j = (j = 1, 2, \dots, n)$ dimana v_i dan h_j adalah biner.

Dengan notasi dan konfigurasi yang diberikan, sebuah RBM dapat diwakili oleh pasangan (v, h) dimana $v = (v_1, v_2, \dots, v_m)^T$ dan $h = (h_1, h_2, \dots, h_n)^T$. Energi system didefinisikan sebagai :

$$E(v, h) = - \sum_{i \in \text{visible}} \alpha_i v_i - \sum_{j \in \text{hidden}} \beta_j h_j - \sum_{i,j} w_{ij} v_i h_j$$



Probabilitas jaringan yang terkait dengan setiap kemungkinan pasangan yang terlihat vektor v dan vektor tersembunyi h diasumsikan mengikuti distribusi Boltzmann:

$$p(v, h) = \frac{1}{Z} e^{-E(v, h)}$$

di mana Z adalah konstanta normalisasi dan disebut fungsi partisi, yaitu dasarnya penjumlahan atas semua konfigurasi yang mungkin.

$$Z = \sum_{v, h} e^{-E(v, h)}$$

Probabilitas marjinal dari jaringan yang terkait dengan v dapat dihitung dengan jumlahkan semua vektor tersembunyi yang mungkin dalam Persamaan. (9.45), maka diperoleh

$$p(v, h) = \frac{1}{Z} \sum_{v, h} e^{-E(v, h)}$$

sehingga gambar pelatihan dapat memaksimalkan probabilitas jaringan yang terkait (sehingga meminimalkan energi yang sesuai). Untuk set pelatihan, maksimalisasi probabilitas gabungan $p(v)$ adalah ekuivalen untuk memaksimalkan log probabilitas log yang diharapkan $p(v)$. Dimana

$$: \frac{\partial \log p(v)}{\partial w_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{RBM Model}$$

kita dapat menghitung penyesuaian bobot dengan menggunakan gradien stokastik

metode $\Delta w_{ij} = \eta (\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{RBM Model})$, dimana $\langle v_i h_j \rangle$ adalah ekspektasi atas distribusi terkait.

9.4 Support Vector Machines

Tujuannya adalah untuk memisahkan beberapa data kompleks menjadi beberapa data yang berbeda kategori. Misalnya, dalam pengenalan pola, kita mungkin hanya perlu memisahkan lingkaran dari kotak. Yaitu dengan melabeli mereka ke dalam dua kelas yang berbeda.

9.4.1 Teori Pembelajaran Statistika

Untuk kasus klasifikasi dua kelas, kami memiliki contoh atau data pembelajaran sebagai (x_i, y_i) dimana $i = 1, 2, \dots, n$ dan $y_i \in \{-1, +1\}$ $i = 1, 2, \dots, n$ dan $y_i \in \{-1, +1\}$. Tujuan dari pembelajaran tersebut adalah untuk menemukan fungsi $f_\beta(x)$ dari fungsi yang diizinkan $\{f_\beta: \beta \in \Omega\}$ dalam parameter ruang Ω sedemikian rupa sehingga

$$f_\beta(x_i) \mapsto y_i \quad (i = 1, 2, \dots, n),$$

Dan risiko bahwa $E(\beta)$ yang diharapkan minimal. Artinya, minimalisasi risiko $E(\beta) = 1$

$$E(\beta) = \frac{1}{2} \int |f_\beta(x) - y| dP(x, y),$$

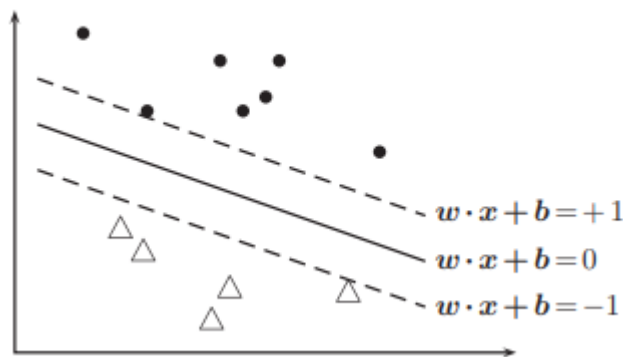
di mana $P(x, y)$ adalah distribusi probabilitas yang tidak diketahui, yang membuatnya tidak mungkin untuk menghitung $E(\beta)$ secara langsung. Pendekatan sederhana adalah menggunakan apa yang disebut risiko empiris

$$E_P(\beta) \approx \frac{1}{n} \sum_{i=1}^n \frac{1}{2} |f_\beta(x_i) - y_i|.$$

Kelemahan utama dari pendekatan ini adalah bahwa risiko atau kesalahan kecil pada pelatihan himpunan tidak selalu menjamin kesalahan kecil pada prediksi jika angka n data pelatihan kecil.

Dalam kerangka minimalisasi risiko struktural dan teori pembelajaran statistik, ada batas atas untuk kesalahan tersebut. Untuk probabilitas tertentu dari setidaknya $1 - p$, Vapnik yang terikat untuk kesalahan dapat ditulis sebagai

$$E(\beta) \leq E_P(\beta) + \mathcal{O} \left[\frac{h}{n}, \frac{\log(p)}{n} \right]$$



Gambar 9.4 Hyperplane, margin maksimum, dan mesin vektor dukungan linier

dimana

$$\mathcal{O} \left[\frac{h}{n}, \frac{\log(p)}{n} \right] = \sqrt{\frac{1}{n} \left[h \left(\log \frac{2n}{h} + 1 \right) - \log \left(\frac{P}{4} \right) \right]}.$$

Di sini h adalah parameter, sering disebut sebagai dimensi Vapnik–Chervonenskis (atau hanya dimensi VC). Dimensi ini menjelaskan kapasitas untuk prediksi himpunan fungsi f_β . Dalam

klasifikasi biner paling sederhana dengan hanya duanilai +1 dan -1, h pada dasarnya adalah jumlah maksimum poin yang dapat diklasifikasikan ke dalam dua kelas yang berbeda dalam semua kemungkinan kombinasi 2^h .

9.4.2 Support Vector Machine (SVM) Linier

Ide dasar klasifikasi adalah mencoba memisahkan sampel yang berbeda ke dalam kelas yang berbeda. Untuk klasifikasi biner seperti segitiga dan bola (atau solid dots) seperti yang ditunjukkan pada

$$\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = 0,$$

sehingga sampel ini dapat dibagi menjadi beberapa kelas dengan segitiga di satu sisi dan bola-bola di sisi lain. Di sini, vektor normal \mathbf{w} dan \mathbf{b} memiliki ukuran yang sama dengan \mathbf{x} , dan mereka dapat ditentukan menggunakan data, meskipun metodenya menentukan mereka tidak mudah. Ini membutuhkan keberadaan hyperplane; jika tidak, pendekatan ini tidak akan berhasil. Dalam hal ini, kita harus menggunakan metode lain.

Perlu ditunjukkan bahwa $\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = 0$ juga dapat ditulis sebagai

$$\mathbf{w}^T \mathbf{x} + \mathbf{b} = 0,$$

tetapi bentuk produk titik dari $\mathbf{w} \cdot \mathbf{x}$ secara eksplisit menyoroti sifat hyperplane

diatur oleh arah normal \mathbf{w} .

Secara matematis, ini setara dengan dua persamaan

$$\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = +1, \text{ dan}$$

$$\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = -1.$$

Dari kedua persamaan ini, sangat mudah untuk memverifikasi bahwa jarak normal (tegak lurus) antara kedua hyperplane ini terkait dengan norma $\|\mathbf{w}\|$ melalui

$$d = \frac{2}{\|\mathbf{w}\|}$$

Dari titik optimasi yang terlihat , maksimalisasi margin dapat ditulis sebagai

$$\text{Minim } \frac{1}{2} \|w\|^2 = \frac{1}{2} \|w\|_2^2 = \frac{1}{2} (\mathbf{w} \cdot \mathbf{w}).$$

Jika kita dapat mengklasifikasikan semua sampel sepenuhnya, untuk sampel apa pun (x_i, y_i) dimana $i = 1, 2, \dots, n$, mempunyai

$$w \cdot x_i + b \geq +1, \quad \text{jika } (x_i, y_i) \in \text{satu kelas}$$

dan

$$w \cdot x_i + b \leq -1, \quad \text{jika } (x_i, y_i) \in \text{kelas lainnya.}$$

Seperti $y_i \in \{+1, -1\}$, dua persamaan di atas dapat digabungkan sebagai

$$y_i(w \cdot x_i + b) \geq 1 \quad (i = 1, 2, \dots, n)$$

Namun, pada kenyataannya, tidak selalu mungkin untuk membangun pemisahan seperti itu

hyperplane. Pendekatan yang sangat berguna adalah dengan menggunakan variabel slack nonnegative

$$\eta_1 \geq 0 \quad (i = 1, 2, \dots, n),$$

Jadi,

$$y_i(w \cdot x_i + b) \geq 1 - \eta_1 \quad (i = 1, 2, \dots, n).$$

Sekarang masalah optimasi untuk SVM menjadi

$$\text{meminimalkan } \Psi = \frac{1}{2} \|w\|^2 + \lambda \sum_{i=1}^n \eta_1,$$

s.t

$$y_i(w \cdot x_i + b) \geq 1 - \eta_1,$$

$$\eta_1 \geq 0 \quad (i = 1, 2, \dots, n),$$

di mana $\lambda > 0$ adalah parameter penalti yang harus dipilih dengan tepat. Dengan menggunakan pengganda Lagrange $\alpha_i \geq 0$, kita dapat menulis ulang batasan di atas optimasi ke dalam versi yang tidak dibatasi, dan kami memiliki

$$L = \frac{1}{2} \|w\|^2 + \lambda \sum_{i=1}^n \eta_i - \sum_{i=1}^n \alpha_i [y_i(w \cdot x_i + b) - (1 - \eta_i)].$$

Maka diperoleh

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0,$$

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^n \alpha_i y_i = 0, \quad y_i(w \cdot x_i + b) - (1 - \eta_i) \geq 0$$

$$\alpha_i [y_i(w \cdot x_i + b) - (1 - \eta_i)] = 0, \quad (i = 1, 2, 3, \dots, n)$$

$$\alpha_i \geq 0, \quad \eta_i \geq 0, \quad (i = 1, 2, \dots, n)$$

Dari pertama kondisi KKT, diperoleh

$$w = \sum_{i=1}^n \alpha_i y_i x_i$$

Ada masalah ganda untuk masalah optimasi SVM ini, dan dapat ditunjukkan bahwa solusi untuk α_i dapat ditemukan dengan menyelesaikan pemrograman kuadrat berikut;

$$\text{maximize } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

s. t

$$\sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha \leq \lambda \quad (i = 1, 2, \dots, n)$$

Maka diperoleh fungsi

$$f(x) = \text{sign} \left[\sum_{i=1}^n \alpha_i y_i (x \cdot x_i) + b \right]$$

Dimana $\text{sign}(x)$ adalah fungsi sign klasik, lalu $\text{sign}(x) = +1$ jika $x > 0$. -1 jika $x < 0$ dan jika $x = 0$.

9.4.3 Fungsi Kernel dan Nonlinear SVM

Pada kenyataannya, sebagian besar masalah adalah nonlinier, dan SVM linier di atas tidak dapat digunakan. Idealnya, kita harus menemukan beberapa transformasi nonlinier sehingga data dapat dipetakan ke ruang dimensi tinggi di mana klasifikasi menjadi linier (lihat Gambar 9.5). Transformasi harus dipilih dengan cara tertentu sehingga produk titiknya mengarah ke fungsi gaya kernel

$$K(x, x_i) = \phi(x) \cdot \phi(x_i)$$

$$f(x) = \text{sign} \left[\sum_{i=1}^n \alpha_i y_i K(x, x_i) + b \right]$$

Dari teori fungsi eigen, kita mengetahui bahwa adalah mungkin untuk memperluas fungsi dalam bentuk fungsi eigen. Sebenarnya, kita tidak perlu mengetahui transformasi seperti itu, kita bisa langsung menggunakan fungsi kernel $K(x, x_i)$ untuk menyelesaikan tugas ini. Inilah yang disebut trik fungsi kernel. Sekarang tugas utamanya adalah memilih fungsi kernel yang sesuai untuk masalah yang diberikan.

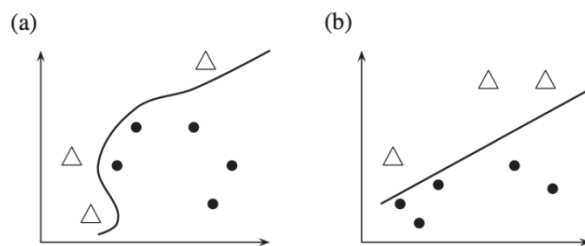


Figure 9.5 Kernel functions by nonlinear transformation. (a) Input space and (b) feature space.

untuk masalah nonlinear SVM

$$K(x, x_i) = (x \cdot x_i)^d$$

untuk klasifikasi polynomial

$$K(x, x_i) = \tanh[k(x \cdot x_i) + \Theta]$$

untuk neural network menggunakan basis Radial Gaussian Fungsi (RBF)

$$\begin{aligned} K(x, x_i) &= \exp\left[-\frac{\|x-x_i\|^2}{(2\sigma)^2}\right] \\ &= \exp[-\gamma\|x-x_i\|^2] = \exp[-\gamma r^2] \end{aligned}$$

untuk pengklasifikasi nonlinier. Di sini, $r = \|x - x_i\|$. Kernel ini dapat dengan mudah diperluas ke dimensi tinggi apa pun. Di sini, 2 adalah varians dan $1/2\sigma^2$ adalah konstanta. Faktanya, adalah parameter hiper, yang perlu disetel untuk setiap SVM.

Mengikuti prosedur serupa seperti yang dibahas sebelumnya untuk SVM linier, kita dapat memperoleh koefisien α_i dengan memecahkan masalah optimasi berikut:

$$\text{maximize} \sum_{i=1}^n \alpha_i - \frac{1}{2} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

Perlu ditunjukkan bahwa, ketika matriks $A = y_i y_j K(x_i, x_j)$ adalah matriks definit positif simetris, masalah maksimisasi di atas menjadi masalah pemrograman kuadratik, dan dengan demikian dapat diselesaikan secara efisien dengan teknik pemrograman kuadratik standar.

Ada banyak paket perangkat lunak (komersial atau open source) yang mudah didapat, jadi kami tidak akan membahas implementasinya. Selain itu, beberapa metode dan variannya masih merupakan bidang penelitian aktif. Pembaca yang tertarik dapat merujuk ke literatur yang lebih maju.

BAB 10

Masalah Optimasi dengan Kendala

Berupa Persamaan

10.1 Metode Pengali Lagrange

Kadang-kadang kita perlu mencari nilai ekstrem suatu fungsi yang daerah asalnya adalah dikendalai berada pada suatu subhimpunan tertentu dari bidang. Pada bab ini kita akan mempelajari suatu metode untuk mencari nilai ekstrem fungsi kendala yaitu metode pengali Lagrange.

Pada bagian ini, kita sajikan kondisi yang diperlukan untuk masalah ekstrim dengan kendala. Hasilnya adalah teorema Lagrange yang terkenal. Untuk lebih memahami ide yang mendasari teorema ini, pertama-tama kita perhatikan fungsi dari dua variabel dan hanya satu kendala berupa persamaan. Misalkan $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ menjadi fungsi kendala. Ingat bahwa pada setiap titik x dari domain, vector gradient $\nabla h(x)$ adalah orthogonal terhadap tingkatan himpunan yang melalui titik tersebut. Mari kita pilih titik $x^* = [x_1^*, x_2^*]^T$, $h(x^*) = 0$, dan asumsikan $\nabla h(x^*) \neq 0$. Tingkatan himpunan melalui titik x^* adalah himpunan $\{x : h(x) = 0\}$. Kami kemudian membuat parameter level ini di persekitaran x^* dengan kurva $\{x(t)\}$ yaitu fungsi vektor yang terdiferensiasi kontinu $x : \mathbb{R} \rightarrow \mathbb{R}^2$ sedemikian sehingga

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad t \in (a, b), \quad x^* = x(t^*), \quad \dot{x}(t^*) \neq 0, \quad t^* \in (a, b)$$

Kita dapat menunjukkan bahwa $\nabla h(x^*)$ adalah ortogonal terhadap $\dot{x}(t^*)$. Memang, karena h adalah konstan pada kurva $\{x(t) : t \in (a, b)\}$, kami memiliki itu untuk semua $t \in (a, b)$,

Karenanya, $\forall t \in (a, b)$,

$$\frac{d}{dt} h(x(t)) = 0$$

Dengan menerapkan aturan rantai, kita mendapatkan

$$\frac{d}{dt} h(x(t)) = \nabla h(x(t))^T \dot{x}(t) = 0$$

Oleh karenanya, $\nabla h(x^*)$ adalah ortogonal terhadap $\dot{x}(t)$.

Sekarang perhatikan bahwa x^* adalah peminimum dari $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ pada himpunan $\{x : h(x) = 0\}$. Kami mengklaim bahwa $\nabla f(x^*)$ adalah ortogonal terhadap $\dot{x}(t^*)$. Untuk melihat ini, cukup dengan mengamati bahwa fungsi komposit t diberikan oleh

$$\phi(t) = f(x(t))$$

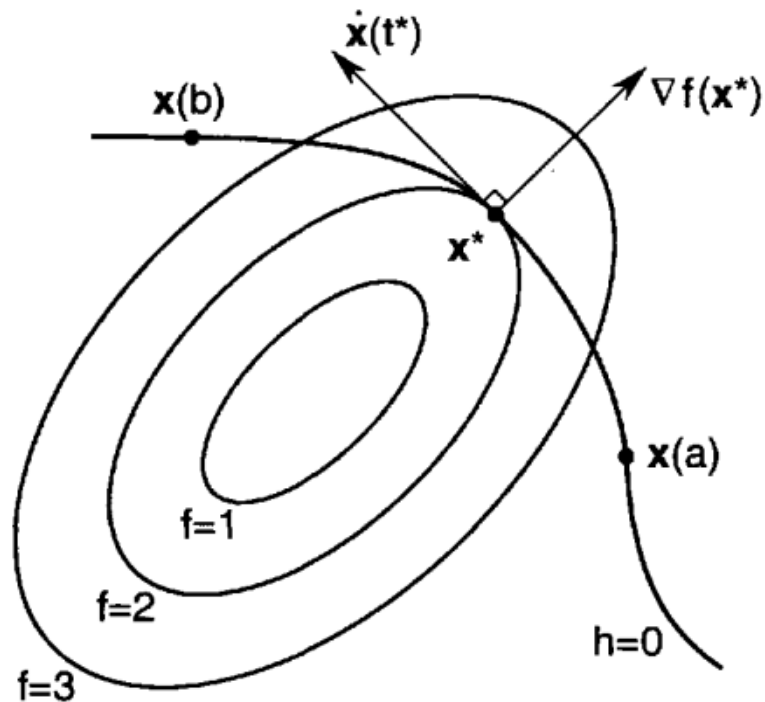
mencapai minimum pada t^* . Akibatnya, kondisi perlu orde pertama untuk masalah ekstrem tanpa kendala menyiratkan

$$\frac{d\phi}{dt}(t^*) = 0$$

Dengan menerapkan aturan rantai menghasilkan

$$0 = \frac{d}{dt} \phi(t^*) = \nabla f(x(t^*))^T \dot{x}(t^*) = \nabla f(x^*)^T \dot{x}(t^*)$$

Dengan demikian, $\nabla f(x^*)$ adalah ortogonal terhadap $\dot{x}(t^*)$. Sesungguhnya bahwa $\dot{x}(t^*)$ adalah bersinggungan dengan kurva $\{x(t)\}$ pada x^* yang berarti bahwa $\nabla f(x^*)$ adalah ortogonal terhadap kurva di x^*

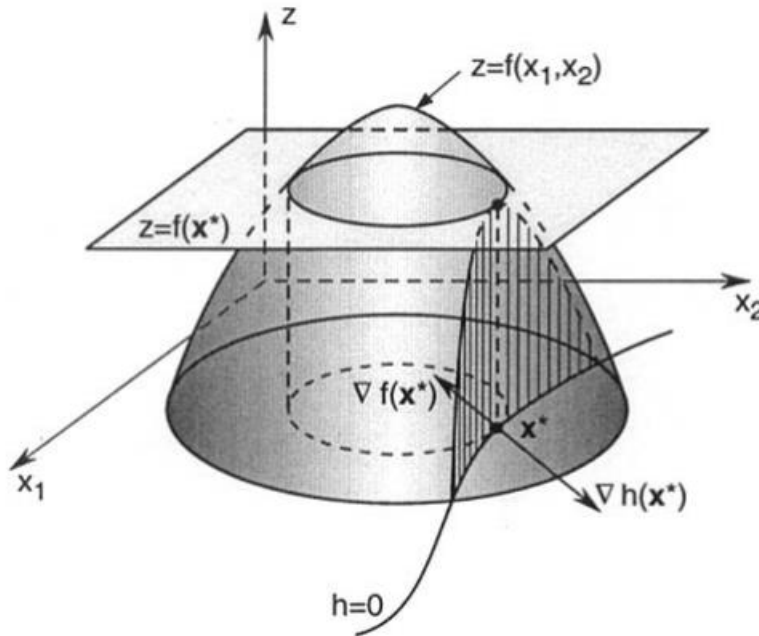


Ingat bahwa $\nabla h(x^*)$ juga ortogonal terhadap $\dot{x}(t^*)$. Oleh karena itu, vektor $\nabla h(x^*)$ dan $\nabla f(x^*)$ adalah “paralel”, yaitu $\nabla f(x^*)$ adalah pengali skalar dari $\nabla h(x^*)$. Pengamatan di atas memungkinkan kita sekarang untuk merumuskan teorema Lagrange untuk fungsi dua variabel dengan satu kendala

Teorema. Teorema Lagrange’s untuk $n = 2$, $m = 1$. Misalkan titik x^* menjadi peminimum dari $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ subject to kendala $h(x) = 0$, $h : \mathbb{R}^2 \rightarrow \mathbb{R}$. Maka, $\nabla f(x^*)$ dan $\nabla h(x^*)$ adalah paralel. Yakni jika $\nabla h(x^*) \neq 0$, maka terdapat suatu skalar λ^* sedemikian sehingga

$$\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0$$

Pada teorema diatas, kita merujuk kepada λ^* sebagai pengali Lagrange. Perhatikan bahwa teorema ini juga berlaku untuk pemaksimal. Gambar berikut memberikan ilustrasi dari teorema Lagrange untuk kasus dimana x^* adalah pemaksimum dari f pada himpunan $\{x : h(x) = 0\}$.

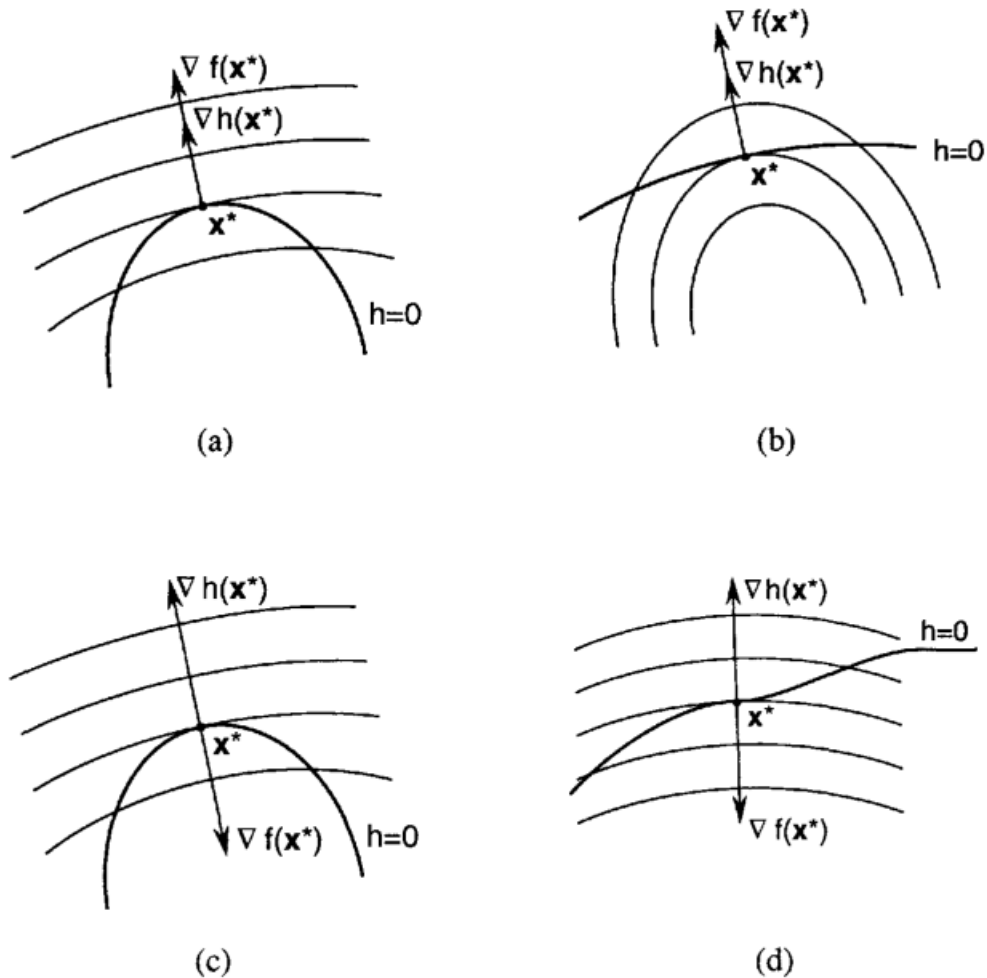


Teorema Lagrange memberikan kondisi perlu orde pertama agar suatu titik menjadi peminimal lokal. Kami menyebut kondisi ini sebagai kondisi Lagrange, yang terdiri dari dua persamaan :

$$\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0$$

$$h(x) = 0$$

Perhatikan bahwa kondisi Lagrange hanya diperlukan tetapi tidak cukup. Pada Gambar berikut, kami mengilustrasikan berbagai titik di mana kondisi Lagrange terpenuhi, termasuk kasus di mana titik tersebut bukan ekstremiser (bukan pemaksimal atau peminimal).



Kami sekarang menggeneralisasi teorema Lagrange untuk kasus ketika $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dan $h : \mathbb{R}^n \rightarrow \mathbb{R}^m, m \leq n$.

Teorema Misalkan x^* menjadi peminimum lokal/pemaksimum lokal dari $f : \mathbb{R}^n \rightarrow \mathbb{R}$, subject to $h(x) = 0, h : \mathbb{R}^n \rightarrow \mathbb{R}^m, m \leq n$. Asumsikan bahwa x^* adalah titik teratur. Maka, terdapat $\lambda^* \in \mathbb{R}^m$ sedemikian sehingga

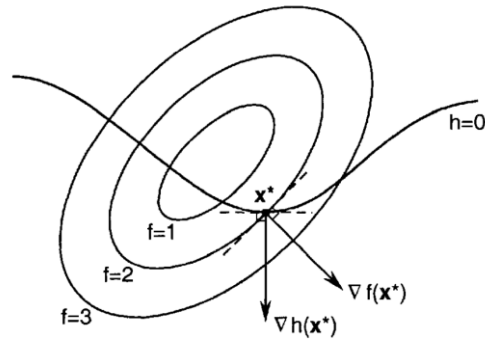
$$Df(x^*) + \lambda^{*T} Dh(x^*) = 0^T$$

Teorema Lagrange menyatakan bahwa jika x^* adalah ekstremiser, maka gradien dari fungsi tujuan dapat dinyatakan sebagai kombinasi linier dari gradien kendala. Kita mengacu pada vektor λ^* dalam teorema di atas sebagai vektor pengali Lagrange, dan komponennya adalah pengali Lagrange.

Perhatikan bahwa x^* tidak dapat menjadi ekstremiser jika

$$\nabla f(x^*) \notin N(x^*)$$

Situasi ini diilustrasikan pada Gambar berikut



Lebih mudah untuk memperkenalkan apa yang disebut fungsi Lagrangian $l : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, diberikan oleh

$$l(x, \lambda) \triangleq f(x) + \lambda^T h(x)$$

Kondisi Lagrange untuk minimizer lokal x^* dapat direpresentasikan menggunakan fungsi Lagrangian sebagai :

$$Dl(x^*, \lambda^*) = 0^T$$

untuk beberapa λ^* , di mana operasi turunan D dengan memperhatikan seluruh argumen $[x^T, \lambda^T]^T$. Dengan kata lain, kondisi perlu dalam teorema Lagrange ekuivalen dengan kondisi perlu orde pertama untuk optimasi tak terbatas yang diterapkan pada fungsi Lagrangian.

Untuk melihatnya, nyatakan turunan dari l terhadap x sebagai $D_x l$, dan turunan dari l terhadap λ sebagai $D_\lambda l$. Kemudian,

$$Dl(x, \lambda) = [D_x l(x, \lambda), D_\lambda l(x, \lambda)]$$

Perhatikan bahwa $D_x l(x, \lambda) = Df(x) + \lambda^T Dh(x)$ dan $D_\lambda l(x, \lambda) = h(x)^T$. Oleh karena itu, teorema Lagrange untuk peminimum lokal x^* dapat dinyatakan sebagai

$$D_x l(x^*, \lambda^*) = 0^T$$

$$D_{\lambda}l(x^*, \lambda^*) = 0^T$$

Untuk suatu λ^* , yang mana ekuivalen dengan

$$Dl(x^*, \lambda^*) = 0^T$$

Dengan kata lain, kondisi Lagrange dapat di tuliskan sebagai $Dl(x^*, \lambda^*) = 0^T$.

Kondisi Lagrange digunakan untuk menemukan kemungkinan ekstrimis. Ini memerlukan penyelesaian persamaan :

$$D_x l(x^*, \lambda^*) = 0^T$$

$$D_{\lambda} l(x^*, \lambda^*) = 0^T$$

Persamaan di atas mewakili $n + m$ persamaan dalam $n + m$ tidak diketahui. Ingatlah bahwa kondisi Lagrange hanya diperlukan, tetapi tidak cukup; yaitu, titik x^* yang memenuhi persamaan di atas tidak perlu menjadi ekstremiser

Contoh 10.1.1 Diberikan luas karton tetap, kita ingin membuat kotak kardus tertutup dengan volume maksimum. Kita dapat merumuskan dan menyelesaikan masalah ini menggunakan kondisi Lagrange. Nyatakan dimensi kotak dengan volume maksimum sebesar x_1 , x_2 , dan x_3 , dan misalkan luas tetap karton yang diberikan adalah A . Masalahnya kemudian dapat dirumuskan sebagai berikut

$$\max \quad x_1 x_2 x_3$$

Subject to

$$x_1 x_2 + x_2 x_3 + x_3 x_1 = \frac{A}{2}$$

Kita notasikan $f(x) = -x_1 x_2 x_3$, dan $h(x) = x_1 x_2 + x_2 x_3 + x_3 x_1 - \frac{A}{2}$. Kita memiliki $\nabla f(x) = -[x_2 x_3, x_1 x_3, x_1 x_2]^T$ dan $\nabla h(x) = [x_2 + x_3, x_1 + x_3, x_1 + x_2]^T$. Perhatikan bahwa semua titik

yang layak adalah reguler dalam kasus ini. Dengan kondisi Lagrange, dimensi kotak dengan volume maksimum memenuhi

$$x_2x_3 - \lambda(x_2 + x_3) = 0$$

$$x_1x_3 - \lambda(x_1 + x_3) = 0$$

$$x_1x_2 - \lambda(x_1 + x_2) = 0$$

$$x_1x_2 + x_2x_3 + x_3x_1 = \frac{A}{2}$$

Dimana $\lambda \in \mathbb{R}$

Sekarang kita selesaikan persamaan diatas. Pertama, kita tunjukan bahwa x_1, x_2, x_3 , dan λ adalah tak nol. Misalkan $x_1 = 0$. Dengan kendala, kita memiliki $x_2x_3 = \frac{A}{2}$. Namun, persamaan kedua dan persamaan ketiga pada kondisi Lagrange menghasilkan $\lambda x_2 = \lambda x_3 = 0$, yang mana bersama-sama dengan persamaan pertama berarti $x_2x_3 = 0$. Ini kontradiksi dengan kendala. Argumen serupa berlaku untuk x_2 dan x_3 .

Selanjutnya, misalkan $\lambda = 0$. Maka, jumlah dari ketiga persamaan Lagrange memberikan $x_2x_3 + x_1x_3 + x_1x_2 = 0$, yang mana kontradiksi dengan kendala.

Sekarang kita selesaikan untuk x_1, x_2 , dan x_3 pada persamaan Lagrange. Pertama, kalikan persamaan pertama dengan x_1 dan kedua dengan x_2 , dan kurangi satu dari yang lain. Kita sampai pada $x_3\lambda(x_1 - x_2) = 0$. Karena baik x_3 maupun λ tidak bisa sama dengan nol, kita simpulkan bahwa $x_1 = x_2$. Kita juga menyimpulkan bahwa $x_2 = x_3$. Dari kendala berupa persamaan, kita memperoleh $x_1 = x_2 = x_3 = \sqrt{\frac{A}{6}}$.

Perhatikan bahwa kita telah mengabaikan kendala yang x_1, x_2 , dan x_3 positif sehingga kita dapat menyelesaikan masalah menggunakan teorema Lagrange. Namun, hanya ada satu solusi untuk persamaan Lagrange, dan solusinya adalah positif. Oleh karena itu, jika ada solusi untuk masalah dengan kendala positif pada variabel x_1, x_2 , dan x_3 , maka solusi ini harus sama dengan solusi di atas yang diperoleh dengan mengabaikan kendala positif.

Selanjutnya kita berikan contoh dengan fungsi tujuan kuadratik dan kendala kuadratik.

Contoh. Diberikan permasalahan dari ekstrimisasi fungsi objektif

$$f(x) = x_1^2 + x_2^2$$

Pada elips

$$\{[x_1, x_2]^T : h(x) = x_1^2 + 2x_2^2 - 1 = 0\}$$

Kita memiliki

$$\nabla f(x) = [2x_1, 2x_2]^T$$

$$\nabla h(x) = [2x_1, 4x_2]^T$$

Dengan demikian

$$D_x l(x, \lambda) = D_x [f(x) + \lambda h(x)] = [2x_1 + 2\lambda x_1, 2x_2 + 4\lambda x_2]$$

Dan

$$D_\lambda l(x, \lambda) = h(x) = x_1^2 + 2x_2^2 - 1$$

Atur $D_x l(x, \lambda) = 0^T$ dan $D_\lambda l(x, \lambda) = 0$ kami memperoleh tiga persamaan dalam tiga yang tidak diketahui

$$2x_1 + 2\lambda x_1 = 0$$

$$2x_2 + 4\lambda x_2 = 0$$

$$x_1^2 + 2x_2^2 = 1$$

Semua titik yang layak dalam masalah ini adalah reguler. Dari persamaan pertama di atas, kita mendapatkan $x_1 = 0$ atau $\lambda = -1$. Untuk kasus dimana $x_1 = 0$, persamaan kedua dan ketiga berarti bahwa $\lambda = -\frac{1}{2}$ dan $x_2 = \pm 1$ dan $x_2 = 0$. Dengan demikian titik-titik memenuhi kondisi Lagrange untuk ekstrem adalah

$$x^{(1)} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad x^{(2)} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \end{bmatrix}, \quad x^{(3)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad x^{(4)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

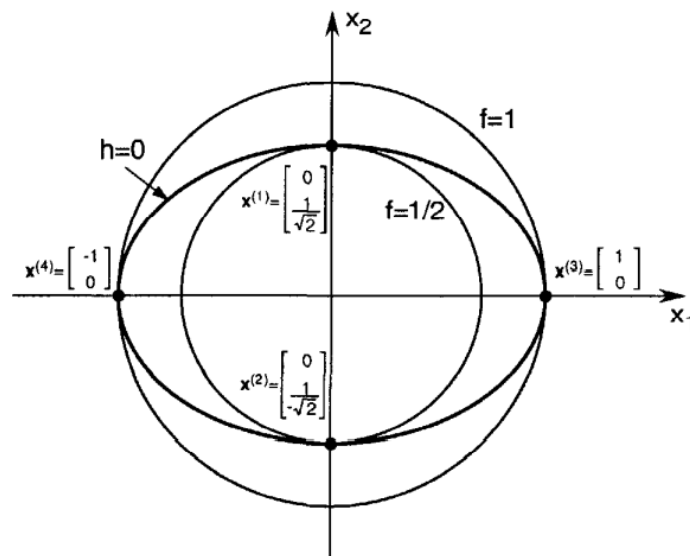
Karena

$$f(x^{(1)}) = f(x^{(2)}) = \frac{1}{2}$$

Dan

$$f(x^{(3)}) = f(x^{(4)}) = 1$$

Kita menyimpulkan bahwa jika ada peminimal, maka mereka berada di $x^{(1)}$ dan $x^{(2)}$, dan jika terdapat pemaksimum, maka lokasinya berada pada $x^{(3)}$ dan $x^{(4)}$. Ternyata, memang, $x^{(1)}$ dan $x^{(2)}$ adalah peminimal dan $x^{(3)}$ dan $x^{(4)}$ adalah pemaksimal. Masalah ini dapat diselesaikan secara grafis, seperti yang diilustrasikan pada Gambar berikut



Pada contoh di atas baik fungsi tujuan f dan fungsi kendala h adalah fungsi kuadratik. Dalam contoh berikutnya, kita melihat lebih dekat pada sebuah kelas dari masalah dimana fungsi objektif f dan fungsi kendala h adalah fungsi n variabel.

Contoh 10.1.2 Diberikan masalah seperti berikut ini :

$$\max \frac{x^T Q x}{x^T P x}$$

Dimana $Q = Q^T \geq 0$, dan $P = P^T > 0$. Perhatikan bahwa jika titik $x = [x_1, \dots, x_n]^T$ adalah solusi untuk masalah, maka begitu juga kelipatan skalar bukan nol darinya,

$$tx = [tx_1, \dots, tx_n]^T, \quad t \neq 0$$

Sesungguhnya

$$\frac{(tx)^T Q(tx)}{(tx)^T P(tx)} = \frac{t^2 x^T Qx}{t^2 x^T Px} = \frac{x^T Qx}{x^T Px}$$

Oleh karena itu, untuk menghindari banyaknya solusi, kami lebih lanjut menentukan kendala

$$x^T Px = 1$$

Masalah optimasi menjadi

$$\max x^T Qx$$

Subject to

$$x^T Px = 1$$

Mari kita tuliskan

$$f(x) = x^T Qx$$

$$h(x) = 1 - x^T Px$$

Setiap titik yang layak untuk masalah ini adalah reguler. Sekarang kita terapkan metode Lagrange. Pertama, bentuk dari fungsi Lagrangian

$$l(x, \lambda) = x^T Qx + \lambda(1 - x^T Px)$$

Menerapkan kondisi Lagrange menghasilkan

$$D_x l(x, \lambda) = 2x^T Q - 2\lambda x^T P = 0^T$$

$$D_\lambda l(x, \lambda) = 1 - x^T Px = 0$$

Persamaan pertama di atas dapat direpresentasikan sebagai

$$Qx - \lambda Px = 0$$

Atau

$$(\lambda P - Q)x = 0$$

Representasi ini memungkinkan karena $P = P^T$ dan $Q = Q^T$. Dengan asumsi $P > 0$, karenanya P^{-1} ada. Sebelum mengalikan $(\lambda P - Q)x = 0$ dengan P^{-1} , kita memperoleh

$$(\lambda I_n - P^{-1}Q)x = 0$$

Atau ekuivalen dengan

$$P^{-1}Qx = \lambda x$$

Oleh karena itu, solusinya, jika ada, adalah vektor eigen dari $P^{-1}Q$, dan pengali Lagrange adalah nilai eigen yang sesuai. Misalkan x^* dan λ^* menjadi solusi optimal. Karena $x^{*T}Px^* = 1$, dan $P^{-1}Qx^* = \lambda^*x^*$, kita memiliki

$$\lambda^* = x^{*T}Qx^*$$

Namun, λ^* adalah maksimum dari fungsi objektif, dan oleh karenanya, fakta bahwa nilai eigen maksimal $P^{-1}Q$

Dalam masalah di atas, kita dapat menemukan titik yang merupakan kandidat untuk ekstremisator dari fungsi tujuan subject to pada kendala dengan persamaan. Titik-titik kritis ini merupakan satu-satunya kandidat karena hanya titik-titik tersebut yang memenuhi kondisi Lagrange. Untuk mengklasifikasikan titik kritis seperti peminimal, pemaksimal, atau bukan keduanya, kita memerlukan kondisi yang lebih kuat—mungkin kondisi perlu dan cukup. Pada bagian berikutnya, kita membahas kondisi perlu orde kedua dan kondisi cukup orde kedua untuk minimizer.

10.2 Masalah Fungsi Penalty

Pada bagian ini, diberikan masalah optimasi dengan kendala dengan bentuk

$$\min f(x)$$

Subject to

$$\begin{aligned}
g_1(x) &\leq 0 \\
g_2(x) &\leq 0 \\
&\vdots \\
g_p(x) &\leq 0
\end{aligned}$$

Dimana $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, p$. Perhatikan hanya kendala pertidaksamaan tidak restriktif, karena kendala persamaan bentuk $h(x) = 0$ adalah ekuivalen untuk kendala berupa pertidaksamaan $||h(x)||^2 \leq 0$. Sekarang kita diskusikan suatu metode untuk menyelesaikan masalah optimasi dengan kendala di atas dengan menggunakan teknik dari optimasi tanpa kendala. Secara khusus, kami memperkirakan masalah optimasi dengan kendala di atas dengan masalah optimasi tanpa kendala

$$\min f(x) + \gamma P(x)$$

Dimana $\gamma \in \mathbb{R}$ adalah konstanta positif, dan $P : \mathbb{R}^n \rightarrow \mathbb{R}$ adalah fungsi yang diberikan. Kami kemudian menyelesaikan masalah berkaitan optimasi tanpa kendala, dan menggunakan solusi sebagai pendekatan untuk meminimalkan masalah asli. Konstanta γ disebut parameter pinalti, dan fungsi P disebut Penyimpangan pinalti. Secara formal, kita mendefinisikan fungsi penalti sebagai berikut.

Definisi 3.1. Suatu fungsi $P : \mathbb{R}^n \rightarrow \mathbb{R}$ disebut fungsi penalty untuk masalah optimasi dengan kendala di atas jika memenuhi tiga kondisi berikut :

- 1) P adalah kontinu
- 2) $P(x) \geq 0, \forall x \in \mathbb{R}^n$
- 3) $P(x) = 0$ jika dan hanya jika x adalah fisibel, yakni $g_1(x) \leq 0, \dots, g_p(x) \leq 0$

Jelasnya, agar masalah tanpa kendala di atas menjadi pendekatan yang baik untuk masalah asli, fungsi penalti P harus dipilih dengan tepat. Peran fungsi pinalti adalah untuk “menghukum” titik yang berada di luar himpunan yang layak. Oleh karena itu, wajar jika fungsi penalti didefinisikan dalam bentuk fungsi kendala g_1, \dots, g_p . Pilihan yang mungkin untuk P adalah

$$P(x) = \sum_{i=1}^p g_i^+(x)$$

Dimana

$$g_i^+(x) = \max(0, g_i(x)) = \begin{cases} 0, & \text{jika } g_i(x) \leq 0 \\ g_i(x), & \text{jika } g_i(x) > 0 \end{cases}$$

Kami menyebut fungsi penalti di atas sebagai fungsi penalti nilai absolut, karena itu sama dengan $\sum |g_i(x)|$, dimana penjumlahan diambil alih semua kendala yang dilanggar di x . Kami menggambarkan fungsi penalti ini dalam contoh berikut.

Contoh 10.2.1 Misalkan $g_1, g_2 : \mathbb{R} \rightarrow \mathbb{R}$ didefinisikan oleh $g_1(x) = x - 2$, $g_2(x) = -(x + 1)^3$. Himpunan fisibel/layak didefinisikan dengan $\{x \in \mathbb{R} : g_1(x) \leq 0, g_2(x) \leq 0\}$ hanyalah interval $[-1, 2]$. Pada contoh ini, kita memiliki

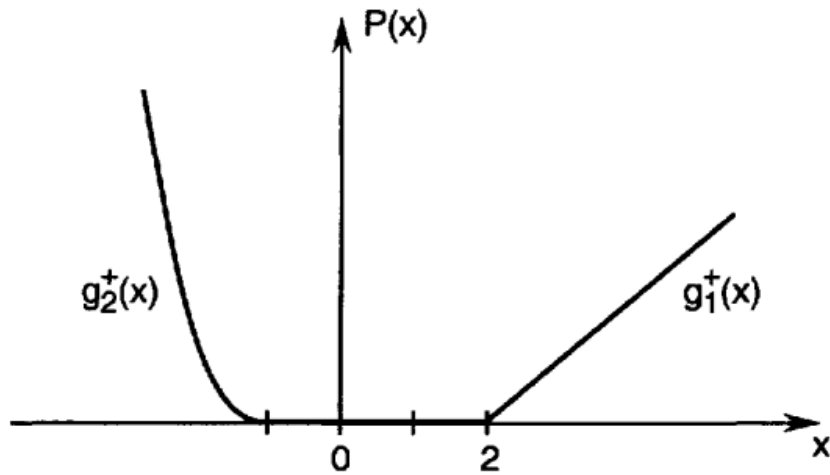
$$g_1^+(x) = \max(0, g_1(x)) = \begin{cases} 0, & \text{jika } x \leq 2 \\ x - 2, & \text{lainnya} \end{cases}$$

$$g_2^+(x) = \max(0, g_2(x)) = \begin{cases} 0, & \text{jika } x \leq 2 \\ -(x + 1)^3, & \text{lainnya} \end{cases}$$

Dan

$$P(x) = g_1^+(x) + g_2^+(x) = \begin{cases} x - 2, & \text{jika } x > 2 \\ 0, & \text{jika } -1 \leq x \leq 2 \\ -(x + 1)^3, & \text{jika } x < -1 \end{cases}$$

Gambar berikut memberikan ilustrasi grafis g^+ untuk contoh ini.



Fungsi penalti nilai absolut mungkin tidak terdiferensialkan pada titik x di mana $g_i(x) = 0$, seperti halnya pada titik $x = 1$ pada contoh di atas (perhatikan, bahwa pada contoh diatas, P terdiferensial pada $x = -1$). Oleh karena itu, dalam kasus seperti itu kita tidak dapat menggunakan teknik optimasi yang melibatkan turunan. Suatu bentuk fungsi penalti yang dijamin dapat dibedakan adalah yang disebut fungsi penalti Courant-Beltrami, yang diberikan oleh

$$P(x) = \sum_{i=1}^p (g_i^+(x))^2$$

Dalam pembahasan berikut, kita tidak mengasumsikan bentuk tertentu dari fungsi penalti P . Kita hanya mengasumsikan bahwa P memenuhi kondisi 1-3 yang diberikan dalam Definisi 3.1.

Metode fungsi penalti untuk menyelesaikan masalah optimasi berkendala melibatkan konstruksi dan penyelesaian masalah optimasi tanpa kendala yang terkait, dan menggunakan solusi untuk masalah tanpa kendala sebagai solusi untuk masalah kendala asli. Tentu saja, solusi untuk masalah tanpa kendala (solusi yang diperkirakan) mungkin tidak sama persis dengan solusi untuk masalah yang dibatasi (solusi yang sebenarnya). Apakah solusi untuk masalah tanpa kendala adalah pendekatan yang baik untuk solusi sebenarnya tergantung pada parameter penalti λ dan fungsi penalti P . Kita berharap bahwa semakin besar nilai parameter penalti λ , semakin dekat solusi yang didekati. untuk solusi yang benar, karena titik yang melanggar batasan di penalty

lebih berat. Idealnya, dalam limit $\gamma \rightarrow \infty$, metode penalti harus menghasilkan solusi yang benar untuk masalah terkendala. Di sisa bagian ini, kami menganalisis sifat ini dari metode fungsi penalti.

Dalam analisis kami tentang metode penalti, kami mengadopsi teknik berikut. Ingatlah bahwa masalah optimasi terkendala asli adalah :

$$\min f(x)$$

Subject to

$$\begin{aligned} g_1(x) &\leq 0 \\ g_2(x) &\leq 0 \\ &\vdots \\ g_p(x) &\leq 0 \end{aligned}$$

Dilambangkan dengan x^* sebuah solusi (peminimum global) untuk masalah di atas. misalkan P menjadi fungsi penalti untuk masalah. Untuk setiap $i = 1, 2, \dots$, misalkan $\gamma_k \in \mathbb{R}$ menjadi konstanta positif yang diberikan. Tentukan fungsi terkait $q(\gamma_k, \cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ dengan

$$q(\gamma_k, x) = f(x) + \gamma_k P(x)$$

Untuk setiap k , kita dapat menulis masalah optimasi tanpa kendala terkait berikut ini :

$$\min q(\gamma_k, x)$$

Dilambangkan dengan $x^{(k)}$ suatu peminimum dari $q(\gamma_k, x)$.

Contoh 10.2.2

$$\min f(x) = x_1^2 + x_2^2 + x_3^2$$

Subject to

$$h(x) = x_1 + x_2 + x_3 = 1$$

dengan titik layak/fisibel awal $x^0 = [1, 0, 0]^T$

interpretasi alternatif masalah di atas adalah menentukan jarak terpendek dari titik asal ke bidang $x_1 + x_2 + x_3 = 1$ menggunakan metode penalty. Perhatikan masalahnya, fungsi penalti yang sesuai adalah $P = x_1^2 + x_2^2 + x_3^2 + \rho(x_1 + x_2 + x_3 - 1)^2$.

Kondisi yang diperlukan pada sebuah minimum tanpa kendala dari P adalah

$$\frac{\partial P}{\partial x_1} = 2x_1 + 2\rho(x_1 + x_2 + x_3 - 1) = 0$$

$$\frac{\partial P}{\partial x_2} = 2x_2 + 2\rho(x_1 + x_2 + x_3 - 1) = 0$$

$$\frac{\partial P}{\partial x_3} = 2x_3 + 2\rho(x_1 + x_2 + x_3 - 1) = 0$$

Jelas bahwa $x_1 = x_2 = x_3$, dan itu mengikuti $x_1 = -\rho(3x_1 - 1)$, yakni

$$x_1(\rho) = \frac{\rho}{1 + 3\rho} = \frac{1}{3 + \frac{1}{\rho}}$$

dan

$$\lim_{\rho \rightarrow \infty} x_1(\rho) = \frac{1}{3}$$

Maka jarak terpendek adalah

$$\sqrt{x_1^2 + x_2^2 + x_3^2} = \sqrt{3\left(\frac{1}{3}\right)^2} = \frac{1}{\sqrt{3}}$$

Contoh 10.2.3 Terapkan metode fungsi penalti untuk masalah :

$$\min f(x) = (x_1 - 1)^2 + (x_2 - 2)^2$$

Subject to

$$h(x) = x_2 - x_1 - 1 = 0,$$

$$g(x) = x_1 + x_2 - 2 \leq 0,$$

$$-x_1 \leq 0,$$

$$-x_2 \leq 0$$

fungsi penalti yang sesuai adalah

$$P = (x_1 - 1)^2 + (x_2 - 2)^2 + \rho(x_2 - x_1 - 1)^2 + \beta_1(x_1 + x_2 - 2) + \beta_2 x_1^2 + \beta_3 x_2^2$$

dimana $\rho \gg 0$ dan $\beta_j = \rho$ jika batasan berupa pertidaksamaan yang sesuai dilanggar, sebaliknya $\beta_j = 0$

Jelas minimum tanpa kendala $[1,2]^T$ melanggar batasan $g(x) \leq 0$, oleh karena itu asumsikan bahwa x berada di kuadran pertama tetapi di luar daerah fisibel, yakni $\beta_1 = \rho$. Fungsi penalti kemudian menjadi

$$P = (x_1 - 1)^2 + (x_2 - 2)^2 + \rho(x_2 - x_1 - 1)^2 + \rho(x_1 + x_2 - 2)^2$$

Kondisi yang diperlukan pada sebuah minimum tanpa kendala dari P adalah :

$$\frac{\partial P}{\partial x_1} = 2(x_1 - 1) - 2\rho(x_2 - x_1 - 1) + 2\rho(x_2 + x_1 - 2) = 0$$

$$\frac{\partial P}{\partial x_2} = 2(x_2 - 2) - 2\rho(x_2 - x_1 - 1) + 2\rho(x_2 + x_1 - 2) = 0$$

Kondisi pertama adalah $x_1(2 + 4\rho) - 2 - 2\rho = 0$, yang mana mengikuti

$$x_1(\rho) = \frac{2\rho + 2}{4\rho + 2} = \frac{2 + \frac{2}{\rho}}{4 + \frac{2}{\rho}}$$

dan

$$\lim_{\rho \rightarrow \infty} x_1(\rho) = \frac{1}{2}$$

Kondisi kedua adalah $x_2(2 + 4\rho) - 4 - 6\rho = 0$, yang mana diberikan

$$x_2(\rho) = \frac{6\rho + 4}{4\rho + 2} = \frac{6 + \frac{4}{\rho}}{4 + \frac{2}{\rho}}$$

dan

$$\lim_{\rho \rightarrow \infty} x_2(\rho) = \frac{3}{2}$$

Oleh karena itu yang optimal adalah $x^* = \left[\frac{1}{2}, \frac{3}{2} \right]^T$

Contoh 10.2.4 Terapkan metode fungsi penalti untuk masalah :

$$\min f(x) = x_1^2 + 2x_2^2$$

Subject to

$$g(x) = 1 - x_1 - x_2 \leq 0$$

Fungsi penaltynya adalah $P = x_1^2 + 2x_2^2 + \beta(1 - x_1 - x_2)^2$. Sekali lagi minimum tanpa kendala jelas melanggar kendala, oleh karena itu asumsi kendala dilanggar dalam fungsi penalti, yaitu $\beta = \rho$.

Kondisi yang diperlukan pada sebuah minimum tanpa kendala dari P adalah :

$$\frac{\partial P}{\partial x_1} = 2x_1 - 2\rho(1 - x_1 - x_2) = 0$$

$$\frac{\partial P}{\partial x_2} = 4x_2 - 2\rho(1 - x_1 - x_2) = 0$$

Kondisi ini memberikan $x_1 = 2x_2$, dan memecahkan hasil lebih lanjut

$$x_2(\rho) = \frac{2\rho}{4 + 6\rho} = \frac{2\rho}{\rho \left(\frac{4}{\rho} + 6 \right)}$$

dan

$$\lim_{\rho \rightarrow \infty} x_2(\rho) = x_2^* = \frac{1}{3}$$

Dan juga

$$x_1^* = \frac{2}{3}$$

Contoh 10.2.5

$$f(x) = 2x_1^2 + x_2^2$$

Subject to

$$g(x) = 5 - x_1 + 3x_2 \leq 0$$

Melalui pendekatan fungsi penalti.

Solusi tanpa kendala, $x_1^* = x_2^* = 0$, melanggar batasan, oleh karena itu aktif dan P menjadi $P = 2x_1^2 + x_2^2 + \rho(5 - x_1 + 3x_2)^2$ dengan kondisi yang diperlukan untuk minimum tanpa kendala :

$$\frac{\partial P}{\partial x_1} = 4x_1 - 2\rho(5 - x_1 + 3x_2) = 0$$

$$\frac{\partial P}{\partial x_2} = 2x_2 + 6\rho(5 - x_1 + 3x_2) = 0$$

Oleh karena itu $x_2 = -6x_1$ dan mensubstitusikan ke dalam kondisi pertama menghasilkan

$$x_1(\rho) = \frac{10\rho}{4 + 38\rho} = \frac{10\rho}{\rho\left(\frac{4}{\rho} + 38\right)}$$

dan

$$\lim_{\rho \rightarrow \infty} x_1(\rho) = x_1^* = \frac{10}{38} = 0.2632$$

Ini memberikan $x_2^* = -1.5789$ dengan $f(x^*) = 2.6316$

BAB 11

Masalah Optimasi dengan Kendala Berupa Pertidaksamaan Kondisi Karush-Kuhn-Tucker

Teorema 11.1 Diberikan suatu masalah primal (primal problem) :

$$\text{minimize } f(x)$$

Subject to

$$g_j(x) \leq 0, \quad j = 1, 2, \dots, m$$

Definisikan Lagrangian :

$$L(x, \lambda) = f(x) + \sum_{j=1}^m \lambda_j g_j(x)$$

Dalam masalah primal diatas, misalkan f dan $g_i \in C^1$, dan diberikan keberadaan pengali Lagrange λ^* , maka kondisi berikut harus memenuhi pada titik x^* yang sesuai dengan solusi masalah primal :

$$\frac{\partial f}{\partial x_j}(x^*) + \sum_{i=1}^m \lambda_i^* \left[\frac{\partial g_i}{\partial x_j} \right](x^*) = 0, \quad j = 1, 2, \dots, n$$

$$g_i(x^*) \leq 0, \quad i = 1, 2, \dots, m$$

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, 2, \dots, m$$

$$\lambda_i^* \geq 0, \quad i = 1, 2, \dots, m$$

atau dalam notasi yang lebih ringkas :

$$\nabla_x L(x^*, \lambda^*) = 0$$

$$\nabla_\lambda L(x^*, \lambda^*) \leq 0.$$

$$\lambda^{*T} g(x^*) = 0$$

$$\lambda^* \geq 0$$

Kondisi ini dikenal sebagai kondisi stasioner Karush-Kuhn-Tucker (KKT).

Bukti. Pertama ubah ketaksamaan batasan dalam masalah primal diatas menjadi kesamaan batasan dengan menggunakan variabel slack s_i :

$$g_i(x) + s_i = 0, \quad i = 1, 2, \dots, m$$

$$s_i \geq 0.$$

Tentukan fungsi Lagrange yang sesuai

$$L(x, s, \lambda) = f(x) + \sum_{i=1}^m \lambda_i (g_i(x) + s_i)$$

Asumsikan bahwa solusi untuk masalah primal dengan kendala :

$$g_i(x) + s_i = 0, \quad i = 1, 2, \dots, m$$

$$s_i \geq 0.$$

diberikan oleh x^*, s^* .

Sekarang bedakan antara dua kemungkinan:

- (i) Misalkan $s_i^* > 0$ untuk semua i . Dalam hal ini masalahnya identik dengan masalah minimisasi interior dengan kendala kesetaraan yaitu diselesaikan dengan menggunakan

pengali Lagrange. Di sini ada m tambahan variabel s_1, s_2, \dots, s_m . Oleh karena itu kondisi yang diperlukan untuk minimum adalah

$$\frac{\partial L}{\partial x_j}(x^*, s^*, \lambda^*) = \frac{\partial f}{\partial x_j}(x^*) + \sum_{i=1}^m \lambda_i^* \frac{\partial g_i}{\partial x_j}(x^*) = 0, j = 1, \dots, n$$

$$\frac{\partial L}{\partial s_i}(x^*, s^*, \lambda^*) = \lambda_i^* = 0, i = 1, \dots, m$$

Ketika $s_i^* > 0$, maka $g_i(x^*) < 0$ dan dengan bukti bahwa $\lambda_i^* = 0$ sehingga

$$\lambda_i^* g_i(x^*) = 0$$

Akibatnya semua kondisi teorema berlaku untuk kasus ini s_i^* untuk semua i .

(ii) Diberikan $s_i^* = 0, i = 1, 2, \dots, p$ dan $s_i^* > 0$ untuk $i = p + 1, \dots, m$

Di kasus ini solusi mungkin untuk solusi minimasi ekuivalen dengan mengikuti persamaan constraint :

$$g_i(x) = 0, i = 1, 2, \dots, p$$

$$g_i(x) + s_i = 0, i = p + 1, \dots, m$$

Lalu gunakan teori Lagrange :

$$\frac{\partial L}{\partial x_j}(x^*, s^*, \lambda^*) = \frac{\partial f}{\partial x_j}(x^*) + \sum_{i=1}^m \lambda_i^* \frac{\partial g_i}{\partial x_j}(x^*) = 0, j = 1, \dots, n$$

$$\frac{\partial L}{\partial s_i}(x^*, s^*, \lambda^*) = \lambda_i^* = 0, i = p + 1, \dots, m$$

Ketika $g_i(x^*) = 0$ untuk $i = 1, 2, \dots, p$ adalah sebagai berikut

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, 2, \dots, m.$$

Jelas sekali bahwa

$$g_i(x^*) < 0, \quad i = p + 1, \dots, m$$

Dan saat $g_i(x^*) = 0$ untuk $i = 1, 2, \dots, p$ adalah sebagai berikut

$$g_i(x^*) \leq 0, \quad i = 1, 2, \dots, m$$

Namun, tidak ada informasi mengenai λ_i^* , $i = 1, \dots, p$ yang tersedia. Informasi ini diperoleh dari argumen tambahan berikut

Misalkan perubahan yang layak dari x^*, s^* di semua variabel $x_1, \dots, x_n, s_1, \dots, s_m$. Sekali lagi perhatikan m sebagai variabel dependen dan n sisanya sebagai variabel independen.

Jika $p \leq n$ maka s_1, s_2, \dots, s_p selalu dapat di ikutkan dalam himpunan variabel

bebas/independen. (Temukan λ^* dengan menggunakan turunan parsial dari L di x^*, s^* dengan memperhatikan variabel dependen, sama dengan nol dan selesaikan λ^*

Ketika $ds_i > 0$ ($s_i^* = 0$) harus berlaku untuk perubahan yang layak pada variabel independen s_1, s_2, \dots, s_p , yang mengikuti perubahan yang konsisten dengan persamaan kendala, yakni

$$df \geq 0$$

Untuk perubahan yang melibatkan s_1, \dots, s_p . Jadi jika variabel independen ini dapat divariasikan satu per satu, karena semua turunan parsial dari L dengan memperhatikan variabel dependen harus sama dengan nol, yakni

$$df = \frac{\partial L}{\partial s_i}(x^*, \lambda^*) ds_i = \lambda_i^* ds_i \geq 0, \quad i = 1, 2, \dots, p$$

Ketika $ds_i > 0$ mengikuti $\lambda_i \geq 0$, $i = 1, 2, \dots, p$. Dengan demikian, karena sudah terbukti bahwa $\lambda_i^* = 0$, $i = p + 1, \dots, m$, maka $\lambda_i^* \geq 0$, $i = 1, 2, \dots, m$.

Contoh 1. Karush-Kuhn-Tucker theory (KKT)

$$\min f(x) = (x_1 - 1)^2 + (x_2 - 2)^2$$

Subject to

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_1 + x_2 \leq 2, \quad x_2 - x_1 = 1$$

Bentuk standar kendala adalah :

$$-x_1 \leq 0, \quad -x_2 \leq 0, \quad x_1 + x_2 - 2 \leq 0, \quad x_2 - x_1 - 1 = 0$$

Dan Lagrangian, sekarang termasuk persamaan kendala dengan Pengali Lagrange μ demikian juga :

$$L(x, \lambda, \mu) = (x_1 - 1)^2 + (x_2 - 2)^2 + \lambda_3(x_1 + x_2 - 2) - \lambda_1 x_1 - \lambda_2 x_2 + \mu(x_2 - x_1 - 1)$$

Kondisi KKT adalah sebagai berikut

$$2(x_1 - 1) + \lambda_3 - \lambda_1 - \mu = 0; \quad 2(x_2 - 2) + \lambda_3 - \lambda_2 + \mu = 0$$

$$-x_1 \leq 0; \quad -x_2 \leq 0; \quad x_1 + x_2 - 2 \leq 0; \quad x_2 - x_1 - 1 = 0$$

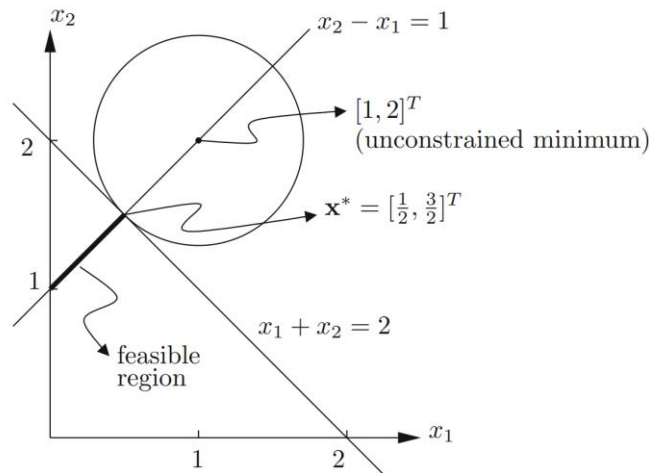
$$\lambda_3(x_1 + x_2 - 2) = 0; \lambda_1 x_1 = 0; \lambda_2 x_2 = 0$$

$$\lambda_1 \geq 0; \lambda_2 \geq 0; \lambda_3 \geq 0$$

Secara umum, pendekatan dari solusi adalah kombinatorial. Coba kemungkinan berbeda dan menguji kontradiksi. Salah satu pilihan yang mungkin adalah $\lambda_3 \neq 0$ yang mengimplikasikan $x_1 + x_2 - 2 = 0$. Ini bersama-sama dengan persamaan kendala

$$x_1^* = \frac{1}{2}, \quad x_2^* = \frac{3}{2}, \quad \lambda_3^* = 1, \quad \lambda_1^* = \lambda_2^* = 0, \quad \mu^* = 0, \quad f(x^*) = \frac{1}{2}$$

Solusi kandidat ini memenuhi semua kondisi KKT dan memang solusi optimal. Mengapa? Solusi grafis digambarkan sebagai berikut



Contoh 2. Karush-Kuhn-Tucker theory

Maksimumkan :

$$f(x_1, x_2, x_3) = -x_1^2 - 2x_2^2 + 27x_1 + 45x_2 - 10x_3$$

Dengan kendala:

$$g_1(x_1, x_2, x_3) = x_1 + x_2 + x_3 \leq 0 \quad g_2(x_1, x_2, x_3) = x_3 \leq 17.25$$

Penyelesaian:

Syarat Karush-Kuhn-Tucker

Pertama turunkan terhadap masing-masing variable sebagai berikut:

$$\frac{\partial f}{\partial x_1} + \sum_{j=1}^2 \lambda_j \frac{\partial g_j}{\partial x_i} = 0; i = 1,2,3$$

$$-2x_1 + 27 + \lambda_1 = 0 \quad (1)$$

$$-4x_2 + 45 + \lambda_1 = 0 \quad (2)$$

$$-10 - \lambda_1 + \lambda_2 = 0 \quad (3)$$

Kemudian $\lambda_j g_j = 0$

$$\lambda_1(x_1 + x_2 - x_3) = 0 \quad (4)$$

$$\lambda_2(x_3 - 17,25) = 0 \quad (5)$$

$$g_j \leq 0$$

$$x_1 + x_2 - x_3 \leq 0 \quad (6)$$

$$x_1 + x_2 - x_3 \quad (7)$$

$$\lambda_j \leq 0$$

$$\lambda_1 \leq 0 \quad (8)$$

$$\lambda_2 \leq 0 \quad (9)$$

Kasus (I): Jika $\lambda_2 = 0$ dan $\lambda_1 \neq 0$

Substitusi $\lambda_2 = 0$ ke persamaan (3), didapat

$$-10 - \lambda_1 + \lambda_2 = 0 \quad -10 - \lambda_1 + 0 = 0 \quad \lambda_1 = -10$$

Setelah memperoleh nilai λ_1 lalu substitusikan λ_1 ke persamaan (1), didapat

$$-2x_1 + 27 + \lambda_1 = 0 \quad -2x_1 + 27 - 10 = 0 \quad -2x_1 + 17 = 0 \quad x_1 = \frac{17}{2}$$

Setelah memperoleh nilai λ_1 lalu substitusi λ_1 ke persamaan (2), didapat

$$-4x_2 + 45 + \lambda_1 = 0 - 4x_2 + 45 - 10 = 0 - 4x_2 + 35 = 0x_2 = \frac{35}{4}$$

Selanjutnya substitusi x_1, x_2 , dan λ_1 ke persamaan (4), didapat

$$\lambda_1(x_1 + x_2 - x_3) = 0 - 10\left(\frac{17}{2} + \frac{35}{4} - x_3\right) = 0 - 10\left(\frac{69}{4} - x_3\right) = 0\frac{69}{4} - x_3 = 0x_3 = \frac{69}{4}$$

Selanjutnya menguji nilai $g_j \geq 0$

$$x_1 + x_2 - x_3 \leq 0 \quad (\text{Pers. 6})$$

$$\frac{17}{2} + \frac{35}{4} - \frac{69}{4} \leq 0 = 0$$

Selanjutnya menguji nilai lamda ($\lambda_j \leq 0$)

$$\lambda_1 = -10 \text{ (Benar)}$$

$$\lambda_1 = 0 \text{ (Benar)}$$

Karena λ_1 dan λ_2 memenuhi syarat maka $f(x_1, x_2)$ optimum pada $x_1 = \frac{17}{2}$, $x_2 = \frac{35}{4}$ dan $x_3 = \frac{69}{4}$

dengan nilai $f(x_1, x_2, x_3)$ sebagai berikut :

$$\begin{aligned} f(x_1, x_2, x_3) &= -x_1^2 - 2x_2^2 + 27x_1 + 45x_2 - 10x_3 \\ &= -\left(\frac{17}{2}\right)^2 - 2\left(\frac{35}{4}\right)^2 + 27\left(\frac{17}{2}\right) + 45\left(\frac{35}{4}\right) - 10\left(\frac{69}{4}\right) = 225\frac{3}{8} \end{aligned}$$

Daftar Pustaka

- Dorigo, M. (1992). *Opimization, learning and natural algorithms*. PhD thesis Politecnico di Milano, Italy.
- Engelbrecht, A.P. (2005). *Fundamentals of Computational Swarm Intelligence*. Hoboken, NJ: Wiley.
- Geem, ZW., Kim, J.H., and Loganathan, G.V. (2001). A new heuristic optimization algorithm: harmony search. *Simulation* 76 (1): 60-68.
- Gutowski, M. (2001). Lévy flights as an underlying mechanism for global optimization algorithms. *ArXiv Mathematical Physics e-Prints* <https://arxiv.org/abs/math-ph/0106003> (accessed 28 February 2018).
- He, X.S., Yang, X.S., Karamanoglu, M. et al. (2017). Global convergence analysis of the flower pollination algorithm: a discrete-time Markov chain approach. *Procedia Computer Science* 108 (1): 1354-1363.
- Karaboga, D. and Basturk, B. (2008). On the performance of artificial bee colony (ABC) algorithm. *Applied Soft Computing* 8 (1): 687-697.
- Kaveh, A. and Talatahari, S. (2010). A novel heuristic optimization method: charged system search. *Acta Mechanica* 213 (3-4): 267-289.
- Kennedy, J. and Eberhart, R.C. (1995). Particle swarm optimization. In: *Proceedings of IEEE International Conference on Neural Networks, 1942-1948* Piscataway, NJ: IEEE.
- Kennedy, J. and Eberhart, R.C. (1997). A discrete binary version of the particle swarm algorithm. *International Conference on Systems, Man, and Cybernetics*. Orlando, FL (12-15 October 1997), vol. 5, 4104-4109. Piscataway, NJ: IEEE Publications.
- Kennedy, J., Eberhart, E.C. and Shi, Y. (2001). *Swarm Intelligence*. London: Academic Press.
- Mantegna, R.N. (1994). Fast, accurate algorithm for numerical simulation of Lévy stable stochastic processes. *Physical Review E* 49: 4677-4683.

- Moritz, R.E. and Southwick, E. (1992). *Bees as Superorganisms*. Berlin: Springer.
- Nolan, J.P. (2009). *Stable Distributions: Models for Heavy-Tailed Data*. Washington, DC: American University.
- Pavlyukevich, L (2007). Lévy flights, non-local search and simulated annealing. *Journal of Computational Physics* 226 (2): 1830-1844.
- Reynolds, A.M. and Frye, M.A. (2007). Free-flight odor tracking in drosophila is consistent with an optimal intermittent scale-free search. *PLoS One* 2 (4): e354-e363.
- Richardson, P. (2008). *Bats*. London: Natural History Museum.
- Waser, N.M. (1986). Flower constancy: definition, cause and measurement. *The American Naturalist* 127 (5): 596-603.
- Yang, X.-S. (2008), *Nature-Inspired Metaheuristic Algorithms*. Frome, UK: Luniver Press
- Yang, X.S. (2010a). Firefly algorithm, stochastic test functions and design optimisation. *International Journal of Bio-Inspired Computation* 2 (2): 78-84.
- K. G. Binmore, *Calculus*. Cambridge, England: Cambridge University Press, 1986.
- R. G. Bland, D. Goldfarb, and M. J. Todd, "The ellipsoid method: A survey," *Operations Research*, vol. 29, pp. 1039-1091, 1981.
- V. G. Boltyanskii, *Mathematical Methods of Optimal Control*. New York: Holt, Rinehart and Winston, 1971.
- S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia: SIAM, 1994.
- L. Brickman, *Mathematical Introduction to Linear Programming and Game Theory*. New York: Springer-Verlag, 1989.
- C. G. Broyden, "Quasi-newton methods," in *Optimization Methods in Electronics and Communications* (K. W. Cattermole and J. J. O'Reilly, eds.), vol. 1 of *Mathematical Topics in Telecommunications*, (New York), pp. 105-110, Wiley. 1984.

- A. E. Bryson and Y.-C. Ho, *Applied Optimal Control; Optimization, Estimation, and Control*. Washington, DC: Revised Printing, Hemisphere Publishing Corporation, 1975.
- B. D. Bunday, *Basic Optimization Methods*. London: Edward Arnold, 1984.
- J. Campbell, *The Improbable Machine*. New York: Simon and Schuster, 1989.
- S. L. Campbell and C. D. Meyer, Jr., *Generalized Inverses of Linear Transformations*. New York: Dover Publications, 1991.
- A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. Chichester: Wiley, 1993.
- S. D. Conte and C. de Boor, *Elementary Numerical Analysis: An Algorithmic Approach*. New York: McGraw-Hill Book Co., third ed., 1980.
- M. A. Dahleh and I. J. Diaz-Bobillo, *Control of Uncertain Systems: A Linear Programming Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- G. B. Dantzig, *Linear Programming and Extensions*. Princeton, NJ: Princeton University Press, 1963.
- G. B. Dantzig and M. N. Thapa, *Linear Programming, 1: Introduction*. New York: Springer-Verlag, 1997.
- L. Davis, ed., *Genetic Algorithms and Simulated Annealing*, Research Notes in Artificial Intelligence, (London), Pitman, 1987.
- V.F. Dem'yanov and L. V. Vasilev, *Nondifferentiable Optimization*. New York: Optimization Software, Inc., Publications Division, 1985.
- D.P. Bertsekas. *Convex Optimization Algorithms*. Athena Scientific, first edition, 2015
- F.H. Branin. Widely used convergent methods for finding multiple solutions of simultaneous equations. *IBM Journal of Research and Development*, 16:504-522, 1972.

- F.H. Branin and S.K. Hoo. A method for finding multiple extrema of a function of n variables. In F.A. Lootsma, editor, *Numerical Methods of Nonlinear Optimization*, pages 231-237, London, 1972. Academic Press
- R.L. Burden, J.D. Faires, and A.M. Burden. *Numerical Analysis*: Cengage Learning, 2015.
- H.-S. Chung and J.J. Alonso. Using gradients to construct response surface models for high-dimensional design optimization problems. In *39th AIAA Aerospace Sciences Meeting and Exhibit*, 2001.
- V. Chvatal, *Linear Programming*. W.H. Freeman & Co Ltd, New York, 1983.
- AR. Conn, NIM. Gould, and Ph.L. Toint. Convergence of quasi Newton matrices generated by the symmetric rank one update. *Mathematical Programming Series B*, 50:177-195, 3 1991.
- G. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, 1963.
- W.C. Davidon. Variable metric method for minimization. R&D Report ANL-5990 (Rev.), US Atomic Energy Commission, Argonne National Laboratories, 1959.
- E. De Klerk and J.A. Snyman. A feasible descent cone method for linearly constrained minimization problems. *Computers and Mathematics with Applications*, 28:33-44, 1994.
- L.C.W. Dixon and G.P. Szego. The global optimization problem: An introduction. In L.C.W. Dixon and G.P. Szegö, editors, *Towards Optimization 2*, pages 1-15, Amsterdam, 1978. North-Holland.
- L.C.W. Dixon, J. Gomulka, and G.P. Szego. Towards a global optimization technique. In L.C.W. Dixon and G.P. Szego, editors. *Towards Global Optimization*, pages 29-54, Amsterdam, 1975. North-Holland.

Simulasi Menggunakan Python

Metode Newton Raphson

```
atplotlib inline
port matplotlib.pyplot as plt
t.style.use('seaborn-white')
port numpy as np
om mpl_toolkits import mplot3d

def Rosenbrock(x1,x2):
    return 100*(x2-x1**2)**2+(1-x1)**2
def Grad_Rosenbrock(x1,x2):
    g1 = 400*x1*(x1**2-x2)+2*x1-2
    g2 = -200*x1**2+200*x2
    return np.array([g1,g2])
def Hessian_Rosenbrock(x1,x2):
    h11 = 1200*x1**2 -400*x2+ 2
    h12 = -400 * x1
    h21 = -400 * x1
    h22 = 200
    return np.array([[h11,h12],[h21,h22]])

def Newton_Raphson_Optimize(Grad, Hess, x1,x2, epsilon=0.000001, nMax = 200):
    #Initialization
    i = 0
    iter_x1, iter_x2, iter_count = np.empty(0),np.empty(0), np.empty(0)
    error = 10
    X = np.array([x1,x2])
    #Looping as long as error is greater than epsilon
    while np.linalg.norm(error) > epsilon and i < nMax:
        i +=1
        iter_x1 = np.append(iter_x1,x1)
        iter_x2= np.append(iter_x2,x2)
        iter_count = np.append(iter_count ,i)
        print(X)
        X_prev = X
        X = X - np.linalg.inv(Hess(x1,x2)) @ Grad(x1,x2)
        error = X - X_prev
        x1,x2 = X[0], X[1]
    return X, iter_x1,iter_x2, iter_coun

root,iter_x1,iter_x2, iter_count = Newton_Raphson_Optimize(Grad_Rosenbrock,Hessi
an_Rosenbrock,-1.2,1.0)
```

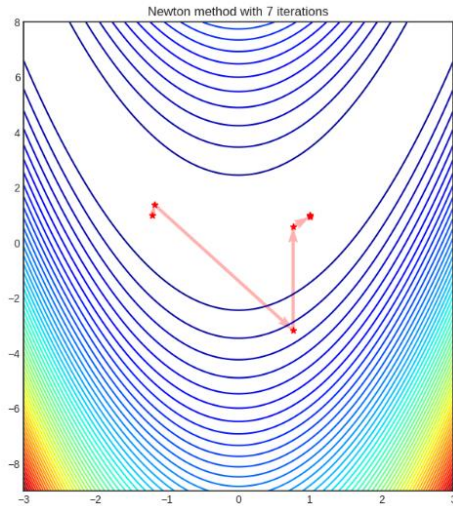
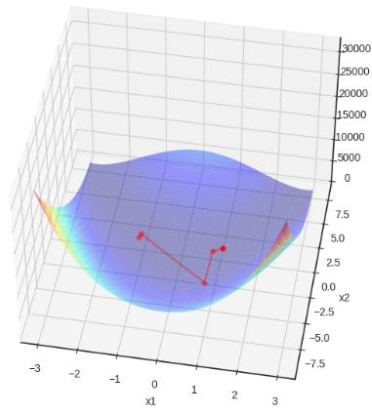
Hasil

Iterasi	Gradien Rosenbrock	Hessian Rosenbrock
0	-1.2	1
1	-1.1752809	1.38067416
2	0.76311487	-3.17503385
3	0.76342968	0.58282478
4	0.99999531	0.94402732
5	0.9999957	0.99999139
6	1	1

Dari $x^0 = [-1.2 \ 1]^T$ menggunakan metode ini, diketahui untuk mendapatkan nilai x^* diperlukan 6 iterasi dan diperoleh $[x_1 \ x_2]^T = [1 \ 1]^T$.

Grafik

```
x = np.linspace(-3,3,250)
y = np.linspace(-9,8,350)
X, Y = np.meshgrid(x, y)
Z = Rosenbrock(X, Y)
#Angles needed for quiver plot
anglesx = iter_x1[1:] - iter_x1[:-1]
anglesy = iter_x2[1:] - iter_x2[:-1]
%matplotlib inline
fig = plt.figure(figsize = (16,8))
#Surface plot
ax = fig.add_subplot(1, 2, 1, projection='3d')
ax.plot_surface(X,Y,Z,rstride = 5, cstride = 5, cmap = 'jet', alpha = .4, edgecolor = 'none' )
ax.plot(iter_x1,iter_x2, Rosenbrock(iter_x1,iter_x2),color = 'r', marker = '*', alpha = .4)
#Rotate the initialization to help viewing the graph
ax.view_init(45, 280)
ax.set_xlabel('x1')
ax.set_ylabel('x2')
#Contour plot
ax = fig.add_subplot(1, 2, 2)
ax.contour(X,Y,Z, 60, cmap = 'jet')
#Plotting the iterations and intermediate values
ax.scatter(iter_x1,iter_x2,color = 'r', marker = '*')
ax.quiver(iter_x1[:-1], iter_x2[:-1], anglesx, anglesy, scale_units = 'xy', angles = 'xy', scale = 1, color = 'r', alpha = .3)
ax.set_title('Newton method with {} iterations'.format(len(iter_count)))
plt.show()
```



Metode Fletcher-Reeves

```
import numpy as np
import pandas as pd

def f(x):
    return 100*(x[1]-x[0]**2)**2 + (1-x[0])**2

def grad_f(x):
    dfdx1 = -400*x[0]*(x[1]-x[0]**2) - 2*(1-x[0])
    dfdx2 = 200*(x[1]-x[0]**2)
    return np.array([dfdx1, dfdx2])

def FR_method(x0, eps=1e-6, max_iter=100):
    x = x0
    g = grad_f(x)
    p = -g
    k = 0
    data = [[k, x[0], x[1], f(x), np.linalg.norm(g)]]
    while np.linalg.norm(g) >= eps and k < max_iter:
        a = backtracking_line_search(x, p)
        x_new = x + a*p
        g_new = grad_f(x_new)
        beta = (g_new @ g_new) / (g @ g)
        p_new = -g_new + beta*p
        x, g, p = x_new, g_new, p_new
        k += 1
        data.append([k, x[0], x[1], f(x), np.linalg.norm(g)])
    headers = ['Iterasi', 'x1', 'x2', 'f(x)', '||grad f(x)||']
    df = pd.DataFrame(data, columns=headers)
    return x, f(x), k, df

def backtracking_line_search(x, p, alpha=1, beta=0.5, rho=0.5):
    t = alpha
```

```

while f(x+t*p) > f(x) + rho*t*(grad_f(x) @ p):
    t *= beta
return t

# contoh penggunaan
x0 = np.array([-1.2, 1])
x_opt, f_opt, n_iter, df = FR_method(x0, max_iter=100)

print('Solusi optimum: x* =', x_opt)
print('Nilai fungsi: f(x*) =', f_opt)
print('Jumlah iterasi:', n_iter)
print('\nTabel iterasi:')
print(df)

```

Hasil

Solusi optimum: $x^* = [0.99999971 \quad 0.99999943]$

Nilai fungsi: $f(x^*) = 8.253862309970111 \times 10^{-14}$

Jumlah iterasi: 81

Tabel iterasi

Iterasi	X_1	X_2	$f(x)$	$grad f(x)$
0	-1.200000	1.000000	$2.420000e + 01$	$2.328677e + 02$
1	-1.094727	1.042969	$6.804583e + 00$	$7.866765e + 01$
2	-1.047428	1.063054	$4.307909e + 00$	$1.958361e + 01$
3	-1.035531	1.067624	$4.145597e + 00$	$6.091230e + 00$
4	-1.031442	1.068525	$4.128920e + 00$	$2.336418e + 00$
⋮	⋮	⋮	⋮	⋮
77	0.999999	0.999998	$2.595215e - 12$	$5.808003e - 05$
78	1.000000	0.999999	$6.458664e - 13$	$3.018069e - 05$
79	1.000000	0.999999	$1.195448e - 13$	$7.891060e - 06$
80	1.000000	0.999999	$8.600065e - 14$	$2.518832e - 06$
81	1.000000	0.999999	$8.253862e - 14$	$7.755664e - 07$

Nilai awal $x^0 = \begin{bmatrix} -1.2 \\ 1 \end{bmatrix}$

Dari metode ini diperoleh $x^* = \begin{bmatrix} 1 \\ 0.999999 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Metode Polak Ribiere

```
from sympy import symbols
x=symbols('x')
def func(x): # Objective function
    return 100*(x[1] - x[0]**2)**2 + (1 - x[0])**2
def Df(x):
    g1= -400*x[0]*(x[1]-x[0]**2) - 2*(1-x[0])
    g2= 200*(x[1]-x[0]**2)
    return np.array([g1,g2])
from scipy.optimize import minimize
res=minimize(fun=func, x0=np.array([-
1.2,1.]), jac=Df, method='CG', options={'gtol':10**-
6, 'disp':True, 'return_all':True})
for i in res.allvecs: print(i)
```

Hasil

Optimization terminated successfully.

Current function value: 0.000000

Iterations: 37

Function evaluations: 80

Gradient evaluations: 79

Iterasi	X_1	X_2
0	-1.2	1
1	-1.02827682	1.07009109
2	-0.91547178	0.80713404
3	-0.47778908	0.19985726
4	-0.4540153	0.21596999
5	-0.28621477	0.05819617
⋮	⋮	⋮
33	0.99981857	0.99966631
34	0.99987932	0.99976405
35	0.99988172	0.99976297
36	1.00000001	1
37	1	1

Untuk nilai awal $x^0 = [-1.2 \ 1]^T$. Dari metode di atas diperlukan 37 iterasi untuk memperoleh x^* , di mana $x^* = [x_1 \ x_2]^T = [1 \ 1]^T$.

Metode Kuadratik

```
import numpy as np
def quadratic(x, A, b, c):
    return 1/2 * np.dot(x.T, np.dot(A, x)) + np.dot(b.T, x) + c
def gradient_quadratic(x, A, b, c):
    return np.dot(A, x) + b
def hessian_quadratic(x, A, b, c):
    return A
def newton_quadratic(x0, A, b, c, epsilon, max_iter):
    x = x0
    for i in range(max_iter):
        grad = gradient_quadratic(x, A, b, c)
        hess = hessian_quadratic(x, A, b, c)
        if np.linalg.norm(grad) <= epsilon:
            break
        p = -np.linalg.solve(hess, grad)
        x += p
    return x
```

Hasil

```
print(newton_quadratic([0,0], [[10,8],[8,10]], [-34,-38], [0,0], 1e-6,1))
[1. 3.]
```

Nilai awal $x^0 = [0 \ 0]^T$.

Dari metode ini diperoleh minimal $x^* = [x_1 \ x_2]^T = [1. \ 3.]^T$

Metode Quasi-Newton

```
import numpy as np
from scipy.optimize import minimize
# Define the objective function
def f(x):
    theta = 0.01
    return 100 * ((x[2] - 10*theta)**2) + (np.sqrt(x[0]**2 + x[1]**2) -
1)**2 + x[2]**2
# Define the gradient of the objective function
def grad_f(x):
    theta = 0.01
    g1 = 400 * (x[2] - 10*theta) * (x[1]**2 + x[0]**2)**(-0.5) * x[0]
    g2 = 400 * (x[2] - 10*theta) * (x[1]**2 + x[0]**2)**(-0.5) * x[1]
    g3 = 200 * (x[2] - 10*theta) + 2 * x[2]
    return np.array([g1, g2, g3])
# Set the initial point and the bounds of x
x0 = np.array([1, 0, 0])
bounds = ((-100, 100), (-100, 100), (-100, 100))
# Minimize the function using L-BFGS-B method with bounds
res = minimize(f, x0, method='L-BFGS-
B', jac=grad_f, bounds=bounds, options={'disp': True})
# Print the results
print(res)
```

Hasil

message: ABNORMAL_TERMINATION_IN_LNSRCH
success: False
status: 2
fun: 0.04990194341052509
x: $[1.200 \times 10^0 \quad 0.000 \times 10^0 \quad 9.901 \times 10^2]$
nit: 2
jac: $[-3.979 \times 10^{-1} \quad -0.000 \times 10^0 \quad -9.511 \times 10^{-4}]$
nfev: 42
njev: 42

Diketahui nilai awal $x^0 = [-1 \quad 0 \quad 0]^T$.

Dari metode ini diperoleh nilai minimal untuk $x^* = \begin{bmatrix} 1.2000 \\ 0.0000 \\ 0.09901 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

Metode Broyden

```
import numpy as np
from scipy.optimize import minimize

def f(x):
    if x[0]>0:
        t = np.arctan(x[1]/x[0])/2*np.pi
    else:
        t = (np.pi + np.arctan(x[1]/x[0]))/2*np.pi
    return 100*((x[2] - 10*t)**2 + (np.sqrt(x[0]**2 + x[1]**2) -
1)**2) + x[2]**2

x0 = [-1, 0, 0] # initial guess

res = minimize(f, x0, method='BFGS', options={'gtol': 1e-
6, 'disp': True})

print(res.x)
print(res.fun)
```

Hasil

Warning: Desired error not necessarily achieved due to precision loss.
Current function value: 0.000000
Iterations: 66
Function evaluations: 452

Gradient evaluations: 110

$[1.00000005 \times 10^0 \quad -7.70286032 \times 10^{-7} \quad -1.19897240 \times 10^{-5}]$
 $1.4516553139770731 \times 10^{-10}$

Metode Gradient Descent

```
import numpy as np

# fungsi yang ingin dioptimasi
def func(x):
    return -(1 / (1 + (x[0] - x[1])**2)) - np.sin((1/2) * np.pi * x[1] * x[2]) -
    np.exp(-((x[0] + x[2]) / x[1] )- 2)**2)
# turunan dari fungsi func
def grad_func(x):
    dfdx0 = 2 * (x[0] - x[1]) / ((x[0] -
    x[1])**2 + 1)**2 + 2 * (x[0] + x[2] / x[1] - 2) * np.exp(-
    ((x[0] + x[2]) / x[1] - 2)**2) / x[1]**2
    dfdx1 = -2 * (x[0] - x[1]) / ((x[0] -
    x[1])**2 + 1)**2 + 1 / 2 * np.pi * x[2] * np.cos(1 / 2 * np.pi * x[1] * x[2]) /
    x[1] + 2 * (x[0] + x[2] / x[1] - 2) * np.exp(-((x[0] + x[2]) / x[1] -
    2)**2) * (-x[2] / x[1]**2) / x[1]**2
    dfdx2 = 1 / 2 * np.pi * x[1] * np.cos(1 / 2 * np.pi * x[1] * x[2]) * (-
    x[0] / x[1]) - 2 * (x[0] + x[2] / x[1] - 2) * np.exp(-((x[0] + x[2]) / x[1] -
    2)**2) * (1 / x[1]) / x[1]
    return np.array([dfdx0, dfdx1, dfdx2])
# metode Gradien Turun
# metode Gradien Turun
def gradient_descent(func, grad_func, x0, alpha=0.0075, epsilon=1e-
9, max_iter=1000):
    x = x0.copy()
    for i in range(max_iter):
        grad = grad_func(x)
        if np.linalg.norm(grad) < epsilon:
            break
        x -= alpha * grad
    return x, func(x)
x0 = np.array([0.0, 1.0, 2.0])
x_opt, f_opt = gradient_descent(func, grad_func, x0)
print("x_opt =", x_opt)
print("f_opt =", f_opt)
```

Hasil

$x_{opt} = [1.49812516 \quad 1.48404627 \quad 0.6680126]$

$f_{opt} = -2.746467239805014$

Diketahui nilai awal $x^0 = [0 \quad 1 \quad 2]^T$.

Dari metode ini diketahui bahwa $x^* = [x_1 \ x_2 \ x_3]^T = [1.499812516 \ 1.48404627 \ 0.6680126]^T$.

Metode Steepest Descent

```
import numpy as np
from scipy.optimize import minimize_scalar

def steepest_descent(f, grad, x0, tol=1e-6, maxiter=100):
    x = x0
    for i in range(maxiter):
        g = grad(x)
        alpha = minimize_scalar(lambda a: f(x - a*g)).x
        x = x - alpha*g
        if np.linalg.norm(g) < tol:
            break
    return x

def freudenstein_roth(x):
    return (-13 + x[0] + ((5 - x[1])*x[1] - 2)*x[1])**2 + \
           (-29 + x[0] + ((x[1] + 1)*x[1] - 14)*x[1])**2

from scipy.optimize import minimize
x0 = [0.5, -2] # initial guess
res = minimize(freudenstein_roth, x0, method='Nelder-Mead')
print(res.x)
```

Hasil

[11.41278135 -0.89680399]

Untuk nilai awal $x^0 = [0.5 \ -2]^T$.

Dari metode ini diketahui penyelesaian untuk $x^* = \begin{bmatrix} 11.41278135 \\ -0.89680399 \end{bmatrix}$.

Metode Conjugate Gradient

```
import numpy as np
from scipy.optimize import minimize

def freudenstein_roth(x):
    x1 = x[0]
    x2 = x[1]
    term1 = (-13 + x1 + ((5 - x2) * x2 - 2) * x2)**2
    term2 = (-29 + x1 + ((x2 + 1) * x2 - 14) * x2)**2
    return term1 + term2

x0 = np.array([0.5, -2])
result = minimize(freudenstein_roth, x0, method='CG', options={'gtol': 1e-
```

```
6, 'disp': True})
print("Solution:", result.x)
print("Iterations:", result.nit)
```

Hasil

Warning: Disired error not necessarily achieved due to precision loss.

Iterations: 14

Function evaluations: 221

Gradient evaluations: 70

Solution: [11.41276833 -0.89680588]

Iterations: 14

Untuk nilai awal $x^0 = [0.5 \quad -2]^T$.

Dari metode ini dibutuhkan 14 iterasi dengan hasil $x^* = \begin{bmatrix} 11.41276833 \\ -0.89680588 \end{bmatrix}$

Metode Broyden-Fletcher-Goldfarb-Shanno

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import minimize
from mpl_toolkits.mplot3d import Axes3D

# Fungsi Brown's badly scaled function
def brown_badly_scaled(x):
    x1 = x[0]
    x2 = x[1]
    return (x1-1e6)**2 + (x2-2e-6)**2 + (x1*x2-2)**2
def grad_f(x):
    x1 = x[0]
    x2 = x[1]
    df_dx1 = 2*(x1-1e6) + 2*(x1*x2-2)*x2
    df_dx2 = 2*(x2-2e-6) + 2*(x1*x2-2)*x1
    return np.array([df_dx1,df_dx2])
x0 = np.array([1, 1]) # Titik Awal
x_star = np.array([1e6, 2e-6]) # Titik Target
result=minimize(fun=brown_badly_scaled, x0=np.array([1,1]), jac=grad_f, method='
BFGS', options={'gtol':10**-6, 'disp':True, 'return_all':True})
for i in result.allvecs: print(i)
```

Hasil

Optimization terminated successfully.

Current function value: 0.000000

Iterations: 16

Function evaluations: 27

Gradient evaluations: 27

Iterasi	X_1	X_2
0	1	1
1	$8.82558100e + 04$	$1.00000018e + 00$
2	$1.57455864e + 05$	$-5.68187534e - 01$
3	$3.18051689e + 05$	$-9.05536076e - 01$
4	$5.26750162e + 05$	$8.07461972e - 03$
5	$6.42356303e + 05$	$2.46295024e - 01$
⋮	⋮	⋮
12	$9.99999895e + 05$	$1.96538988e - 06$
13	$1.00000000e + 06$	$1.99996817e - 06$
14	$1.0000000e + 06$	$2.0000001e - 06$
15	$1.e + 06$	$2.e - 06$
16	$1.e + 06$	$2.e - 06$

Untuk nilai awal $x^0 = [1 \ 1]^T$.

Kode ini membutuhkan 16 iterasi untuk memperoleh nilai minimum $x^* = \begin{bmatrix} 1 \times 10^6 \\ 2 \times 10^{-6} \end{bmatrix}$