06: The Hessian Matrix

Chapter Goal: To introduce the "second level" of multivariate derivatives: the **Hessian**. If the [Gradient] is about the "slope" of the landscape, the Hessian is about its **"curvature"**.

---

1. Core Idea: The Second Derivative of a Multi-variable Function

- **Recap:**
  - **First Derivative (Gradient/Jacobian):** Collects all the first partial derivatives ($\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \ldots$) into a **vector**.
- **New Idea (Hessian):**
  - Collects all the **second** partial derivatives ($\frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x \partial y}, \frac{\partial^2 f}{\partial y \partial x}, \ldots$) into a **MATRIX**.
- **Definition of the Hessian Matrix `H`:**
  For a function $f(x, y)$, the Hessian is a 2x2 matrix:

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

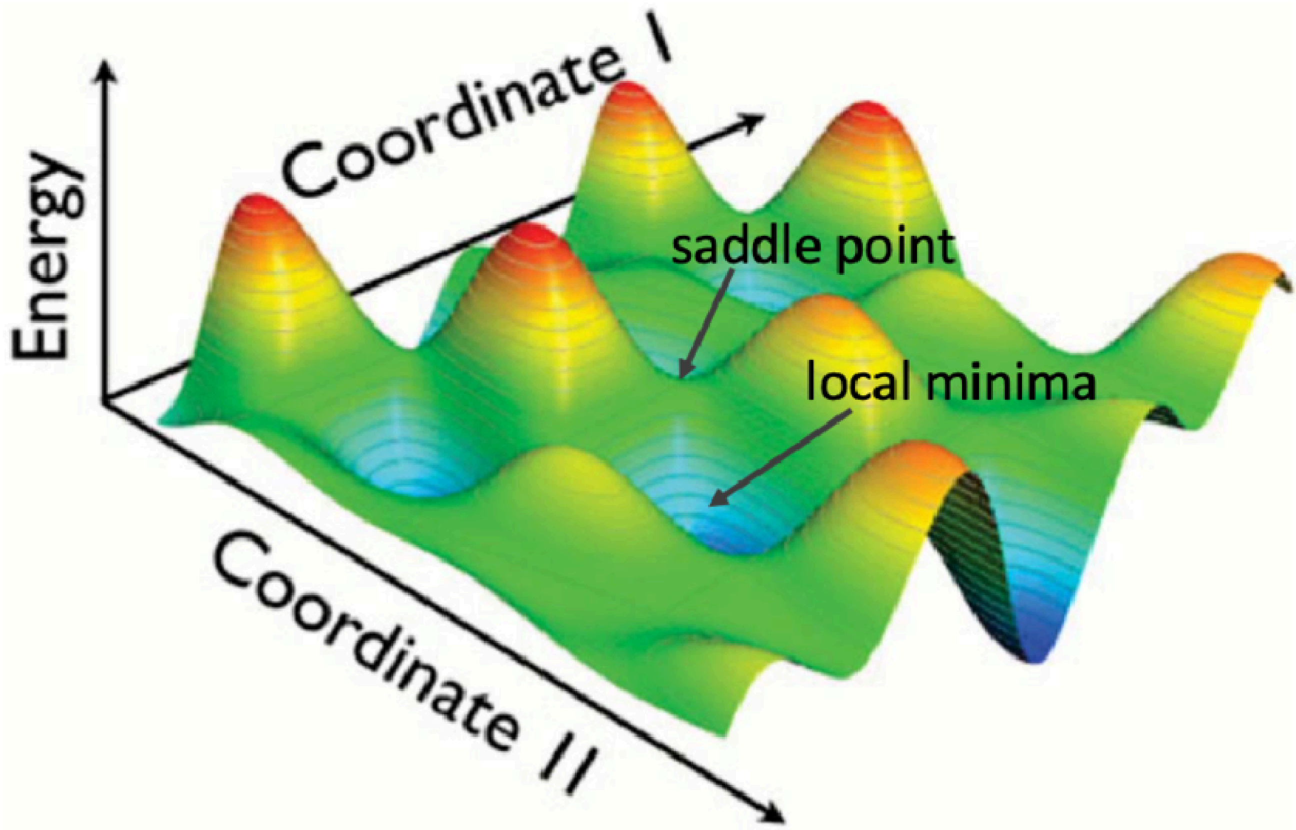
  (Note: $\frac{\partial^2 f}{\partial x \partial y}$ means: first differentiate `f` with respect to `x`, then differentiate the result with respect to `y`).
- **Important Property (Symmetry):** For "well-behaved" (continuous) functions, the Hessian will always be **symmetric** across its main diagonal ($\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$).

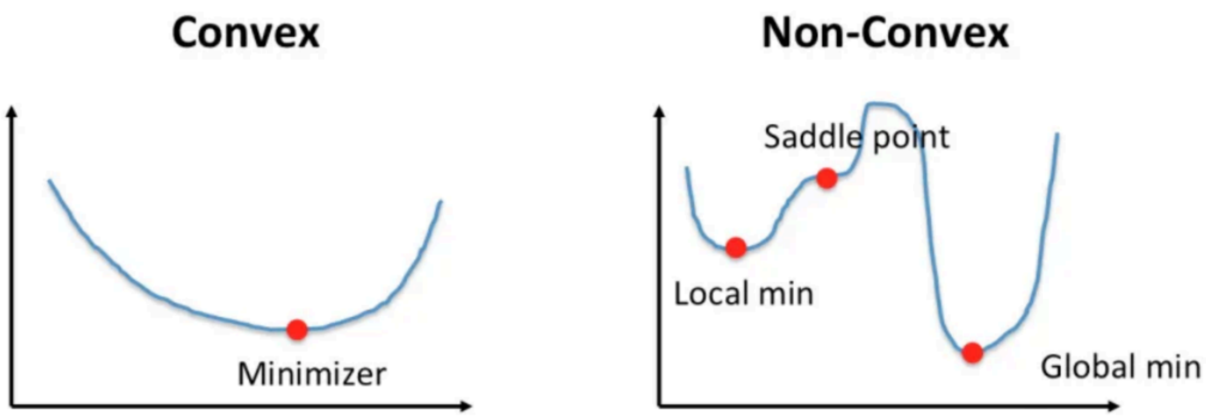
---

2. What is the Geometric Meaning of the Hessian?

> If the **Gradient** tells us about the **slope** of the landscape...
> ...then the **Hessian** tells us about the **curvature** of the landscape around a single point.

- **1D Analogy:** The second derivative tells us if a curve is "smiling" (curving up, concave up) or "frowning" (curving down, concave down).
- **Multi-dimensional Analogy:** The Hessian tells us if the landscape around a point is shaped like a **"bowl facing up"**, a **"hilltop"**, or a **"saddle"**.





---

3. The Main Use: The "Second Derivative Test" for Critical Points

- **The Problem:** We've used the Gradient to find a point where $\nabla f = 0$. We know this is a "flat" point. But is it a **peak (maximum)**, a **valley (minimum)**, or a **saddle point**?
- The **Hessian** is the tool to tell them apart!
- **The "Second Derivative Test" Recipe:**
    1. Calculate the Hessian Matrix `H` at the critical point.
    2. Calculate the **Determinant** of the Hessian, $\det(H)$.

3. Look at the signs of $\det(H)$ and the first element of the Hessian ($H_{11} = \frac{\partial^2 f}{\partial x^2}$).

- **Case 1:** `det(H) > 0` **(Shaped like a Bowl/Hill)**
  - We know it's definitely a local minimum or maximum.
  - To distinguish, look at H_{11}:
    - If $H_{11} > 0$ (curving **up** in the x-direction), it's a **Local Minimum** (a "smiling" bowl).
    - If $H_{11} < 0$ (curving **down** in the x-direction), it's a **Local Maximum** (a "frowning" hilltop).
- **Case 2:** `det(H) < 0` **(Shaped like a Saddle)**
  - This is a **Saddle Point**.
  - **Meaning:** In one direction the landscape curves up, but in another direction it curves down. It is neither a minimum nor a maximum.
- **Case 3:** `det(H) = 0` **(Inconclusive)**
  - The test provides insufficient information. The point could be anything.

---

4. Example: `f(x,y) = x² + y²` (The Perfect Bowl)

- **Gradient:** $\nabla f = [2x, 2y]$. At `(0,0)`, the gradient is `[0,0]`. This is a critical point.
- **Hessian:**
  - $\frac{\partial^2 f}{\partial x^2} = 2$
  - $\frac{\partial^2 f}{\partial y^2} = 2$
  - $\frac{\partial^2 f}{\partial x \partial y} = 0$

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{(constant everywhere)}$$

- **Analysis at `(0,0)`:**
  1. $\det(H) = (2 \cdot 2) - (0 \cdot 0) = 4$. (Positive $\implies$ it's a min/max).
  2. $H_{11} = 2$. (Positive $\implies$ it's a minimum).
- **Conclusion:** The point `(0,0)` is a **Local Minimum**.

---

5. Key Message

- The Hessian is an advanced "diagnostic tool". After the Gradient has led us to a flat point, the Hessian helps us understand the **type** of flat point it is (peak, valley, or saddle).
- This is crucial in more advanced optimization algorithms.

> ⚠ **Warning**
>
> Ini menurut saya agak sulit dipahami, sehingga rasanya butuh penjelasan lebih detail dan pelan pelan

Bagian 1: Keterbatasan "Kompas" Gradien

> **Teks:** "In this video, I'm going to briefly introduce a kind of gradient playground to help you further develop your intuition on the Jacobian, which will also set you up for the following exercises."

Mengingat Kembali:

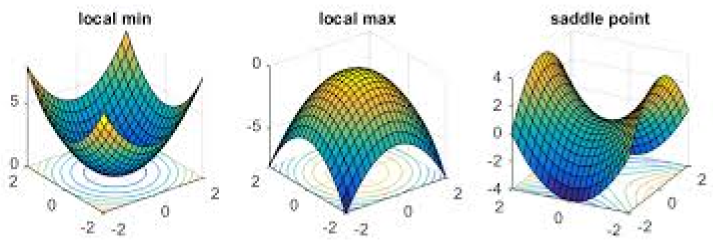
- Kita sudah punya "alat" yang hebat bernama **Gradien (** `∇f` **)**.
- Gradien adalah **"Kompas"** kita. `∇f` menunjuk ke arah tanjakan tercepat, dan `-∇f` menunjuk ke arah turunan tercepat.
- Algoritma Gradient Descent adalah seperti pendaki yang "bodoh": "Ikuti saja arah kompas `-∇f` selangkah demi selangkah."

Masalah Baru:
Bayangkan pendaki kita akhirnya sampai di sebuah tempat di mana tanahnya **datar sempurna**.

- Dia menyalakan kompas Gradien-nya.
- Karena tanahnya datar, tidak ada arah "tanjakan".
- Kompasnya akan menunjukkan `[0, 0]`.
- Pendaki kita akan berkata, "Oke, Gradiennya nol. Aku sudah sampai. Misiku selesai." dan dia **berhenti**.

TAPI... di mana sebenarnya dia berhenti? Ada tiga kemungkinan tempat datar di sebuah lanskap:

1. **Dasar Lembah (Local Minimum):** Misi berhasil! Ini yang kita cari.
2. **Puncak Gunung (Local Maximum):** Ini adalah **solusi terburuk!** Algoritma kita salah berhenti di tempat tertinggi.
3. **Titik Pelana Kuda (Saddle Point):** Ini adalah titik yang aneh, seperti celah di antara dua gunung. Di satu arah (depan-belakang) jalannya menanjak, tapi di arah lain (kiri-kanan) jalannya menurun. Ini juga bukan solusi yang kita inginkan.



Keterbatasan Gradien:

> Gradien hanya memberitahu kita tentang **"kecuraman"**. Dia **tidak bisa membedakan** antara dasar lembah, puncak gunung, atau pelana kuda. Baginya, semua yang "datar" terlihat sama.

Kita butuh "alat" baru yang lebih canggih untuk menganalisis titik-titik datar ini.

Cek Pemahaman (Berhenti di Sini Dulu):

1. Apakah kamu paham bahwa algoritma Gradient Descent akan **berhenti** saat Gradiennya menjadi nol?
2. Apakah kamu bisa memvisualisasikan tiga jenis "titik datar" yang berbeda: lembah, puncak, dan pelana kuda?
3. Apakah kamu melihat **masalahnya**: bahwa Gradien saja tidak cukup untuk memberitahu kita jenis titik datar mana yang telah kita temukan?

Bagaimana perasaanmu tentang keterbatasan dari "kompas" Gradien ini?

Bagus sekali! Kita sudah sepakat bahwa "Kompas" Gradien (turunan pertama) tidak cukup. Kita butuh alat baru untuk menganalisis titik datar.

Sekarang, mari kita kembali ke dunia 1D (kalkulus biasa) untuk mengingat kembali alat apa yang kita gunakan di sana.

Bagian 2: Mengingat Kembali "Tes Turunan Kedua" di 1D

Bayangkan sebuah grafik 1D, `y = f(x)`.

- **Turunan Pertama (`f'(x)`):** Memberitahu kita **kemiringan (slope)**.
  - Saat `f'(x) = 0`, kita tahu kita berada di titik datar (puncak atau lembah).
- **Turunan Kedua (`f''(x)`):**
  - Ingat dari Chapter 10, turunan kedua memberitahu kita tentang **KELENGKUNGAN (CURVATURE)**.
  - `f''(x) > 0` **(Positif):** Artinya kurvanya **melengkung ke atas** (seperti mangkok "tersenyum"). Ini adalah **LEMBAH (Minimum)**.
  - `f''(x) < 0` **(Negatif):** Artinya kurvanya **melengkung ke bawah** (seperti mangkok "cemberut"). Ini adalah **PUNCAK (Maksimum)**.

"Aha!" Moment di 1D:
Di dunia 1D, **turunan kedua** adalah alat diagnosis yang sempurna untuk membedakan antara lembah dan puncak setelah kita menemukan titik datar.

Bagian 3: Membawa Ide Ini ke Multi-Dimensi (Lahirnya Hessian)

> **Teks:** "For the Jacobian, we collected together all of the first order derivatives... Now, we're going to collect all of the **second order derivatives** together into a **matrix**, which... look like this."
> *(Untuk Jacobian, kita mengumpulkan semua turunan pertama... Sekarang, kita akan mengumpulkan semua **turunan kedua** ke dalam sebuah **matriks**, yang... terlihat seperti ini.)*

Pertanyaan Logis:
Jika turunan kedua di 1D sangat berguna, bisakah kita membuat "turunan kedua" untuk lanskap multi-dimensi kita?

Masalahnya:
Di multi-dimensi, "kelengkungan" menjadi lebih rumit.

- Di satu titik, lanskapnya bisa saja melengkung ke atas di arah Timur-Barat...
- ...tapi melengkung ke bawah di arah Utara-Selatan (ini adalah "pelana kuda").
- Ada juga "kelengkungan diagonal" ($\partial^2 f/\partial x\partial y$).

Solusinya:
Kita tidak bisa lagi hanya menggunakan satu angka. Kita butuh sebuah "wadah" yang lebih besar untuk menyimpan semua informasi kelengkungan ini. "Wadah" itu adalah sebuah **MATRIKS**.

Definisi Matriks Hessian (H):

> **Hessian** adalah **"Gradien dari Gradien"**. Ia adalah sebuah matriks yang mengumpulkan **SEMUA KEMUNGKINAN TURUNAN KEDUA** dari sebuah fungsi.

Bagaimana cara membuatnya (untuk $f(x,y)$):

1. Mulai dengan **Gradien:** $\nabla f = [\ \partial f/\partial x,\ \partial f/\partial y\ ]$.
2. Sekarang, turunkan **setiap komponen** dari Gradien ini terhadap **setiap variabel input** lagi.
   - **Baris 1 (Ambil $\partial f/\partial x$):**
     - Turunkan lagi terhadap x → $\partial^2 f/\partial x^2$.
     - Turunkan lagi terhadap y → $\partial^2 f/\partial x\partial y$.
   - **Baris 2 (Ambil $\partial f/\partial y$):**
     - Turunkan lagi terhadap x → $\partial^2 f/\partial y\partial x$.
     - Turunkan lagi terhadap y → $\partial^2 f/\partial y^2$.
3. **Susun dalam Matriks:**

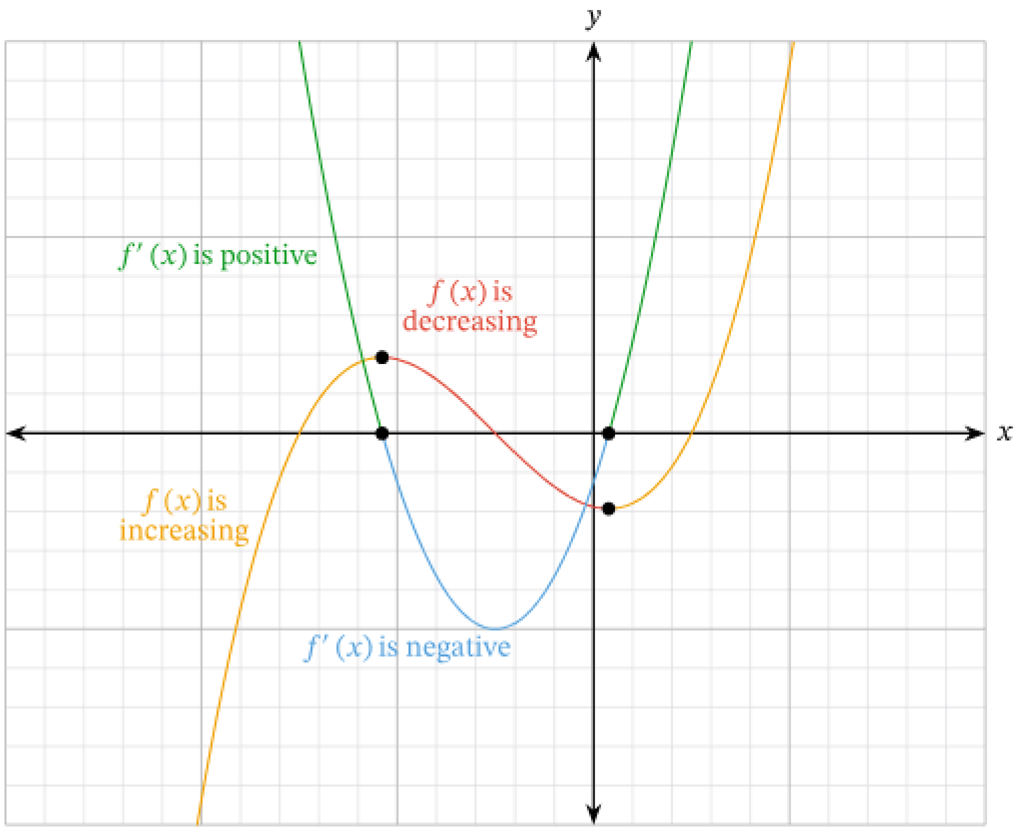
```
H = [ [ ∂²f/∂x²,  ∂²f/∂x∂y ],
      [ ∂²f/∂y∂x,  ∂²f/∂y² ] ]
```

Kesimpulan:
Hessian adalah **generalisasi dari konsep "turunan kedua"** ke dunia multi-dimensi. Ia bukan lagi sebuah angka, melainkan sebuah **matriks** yang menangkap informasi kelengkungan di semua arah.

Cek Pemahaman (Berhenti di Sini Dulu):

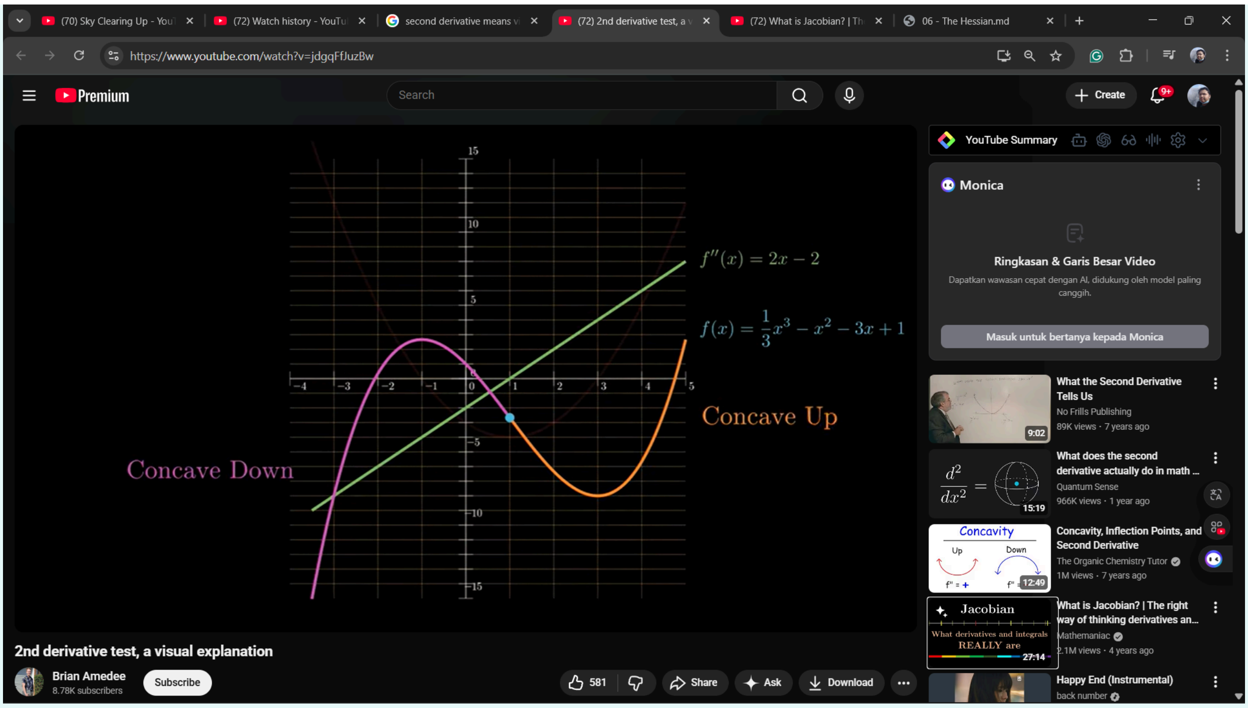
1. Apakah kamu ingat bagaimana "tes turunan kedua" (`f''`) bekerja di kalkulus 1D untuk membedakan puncak dan lembah?
2. Apakah kamu bisa menerima ide bahwa untuk lanskap multi-dimensi, kita butuh "wadah" yang lebih besar (sebuah matriks) untuk menyimpan semua informasi kelengkungan?
3. Apakah kamu melihat bagaimana Matriks Hessian ini dibangun dengan cara menurunkan Gradien sekali lagi terhadap semua variabel?

Bagaimana perasaanmu tentang Hessian sebagai "turunan kedua versi multi-dimensi"?



💧 **Tip**

Watch this video :
https://www.youtube.com/watch?v=jdgqFfJuzBw



Tags: #mml-specialization #multivariate-calculus #hessian-matrix #second-derivative-test #optimization