10.007: Modeling the Systems World

Week 02 - Cohort 1

Taylor Series for Multivariate Functions

Term 3, 2017

Singapore University of
Technology & Design

Agenda

- Review of Taylor series for univariate functions.
- Taylor series for multivariate functions.
- Understanding the optimality conditions better using Taylor series.
- Solving unconstrained optimization problem using optimality conditions.

Activity 1 (10 minutes)

Consider a function $f : \mathbb{R} \to \mathbb{R}$. The first-order derivative $f'(\bar{x})$ represents the slope of the graph of f at $\bar{x}$. Then the "best" linear (first-order) approximation to f at $\bar{x}$ is:

$$\hat{f}_{\bar{x}}(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}).$$

1. Compute the first-order approximation of

   $$f(x) = (x - 2)^3 - x^2,$$

   at the point $\bar{x} = 2$.
2. Draw a sketch of the function and its first-order approximation over $[0, 5]$.

A better approximation? (1D)

If we want a better approximation, we need a second-degree one. This is the second-degree Taylor polynomial:

$$\hat{\hat{f}}_{\bar{x}}(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + \frac{1}{2}f''(\bar{x})(x - \bar{x})^2.$$

This function matches the first- and second-order derivatives of $f(x)$ at the point $\bar{x}$.

Activity 2 (10 minutes)

Recall: $f(x) = (x-2)^3 - x^2$ has $\hat{f}_{\bar{x}=2}(x) = 4 - 4x$.

1. Compute the second-order approximation of

$$f(x) = (x-2)^3 - x^2,$$

   at the point $\bar{x} = 2$.

2. Draw a sketch of the function and its second-order approximation over $[0, 5]$.

Analysis of critical points in $\mathbb{R}$

Compute the second-degree approximation at critical point $\bar{x}$ of the function f, i.e. $f'(\bar{x}) = 0$. We obtain:

$$f(x) \approx \hat{\hat{f}}_{\bar{x}}(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + \frac{1}{2} f''(\bar{x})(x - \bar{x})^2$$
$$= f(\bar{x}) + \frac{1}{2} f''(\bar{x})(x - \bar{x})^2.$$

Therefore:

$$f(x) - f(\bar{x}) \approx \frac{1}{2} f''(\bar{x})(x - \bar{x})^2.$$

The difference in function value with respect to $\bar{x}$ depends on the sign of f'' (up to second order).

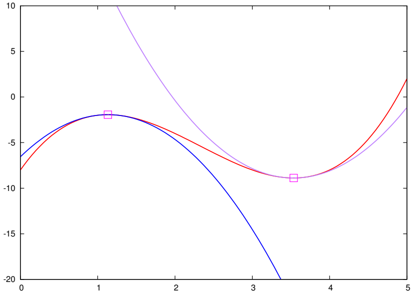
- If $f''(\bar{x}) > 0$, for all points x around $\bar{x}$ the function has higher values.
- If $f''(\bar{x}) < 0$, for all points x around $\bar{x}$ the function has lower values.

Importance of the second derivative

$f(x) = (x-2)^3 - x^2$ has two critical points.

$$f'(1.131482) = 0 \qquad \text{and} \qquad f'(3.53518) = 0.$$

$$f''(1.131482) = -7.211107 \qquad \text{and} \qquad f''(3.53518) = 7.211107.$$



The local minimum and the local maximum is identified using the sign of the second derivative.

First-order approximation (2D)

Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$. Then the “best” linear approximation to f at $\bar{x}$ is:

$$\hat{f}_{\bar{x}}(x) = f(\bar{x}) + \frac{\partial f(\bar{x})}{\partial x_1}(x_1 - \bar{x}_1) + \frac{\partial f(\bar{x})}{\partial x_2}(x_2 - \bar{x}_2)$$

This can be rewritten as:

$$\hat{f}_{\bar{x}}(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}).$$

The gradient gives the slopes of the tangent plane to the surface represented by f.

Activity 3 (10 minutes)

Consider the function:

$$f(x_1, x_2) = (x_1 - 2)^4 - 4(x_1 - 2)^2 - (x_2 - 4)^3.$$

Compute its first-order approximation at the point $(3.5, 2.5)$.

Second-order approximation (2D)

As in the univariate case, if we want a better approximation we need to use a second-degree polynomial (Taylor series):

$$\hat{\hat{f}}_{\bar{x}}(x) = f(\bar{x}) + \sum_{i=1}^{2} \frac{\partial f(\bar{x})}{\partial x_i}(x_i - \bar{x}_i) + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\partial^2 f(\bar{x})}{\partial x_j \partial x_i}(x_i - \bar{x}_i)(x_j - \bar{x}_j)$$

This can be rewritten as:

$$\hat{\hat{f}}_{\bar{x}}(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \begin{pmatrix} \frac{\partial^2 f(\bar{x})}{\partial x_1^2} & \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(\bar{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\bar{x})}{\partial x_2^2} \end{pmatrix} (x - \bar{x}),$$

and the 2×2 matrix is simply $H_f(\bar{x})$.
Notice the analogy with the case where $f : \mathbb{R} \to \mathbb{R}$:

$$\hat{\hat{f}}_{\bar{x}}(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + \frac{1}{2} f''(\bar{x})(x - \bar{x})^2.$$

Example in 2D

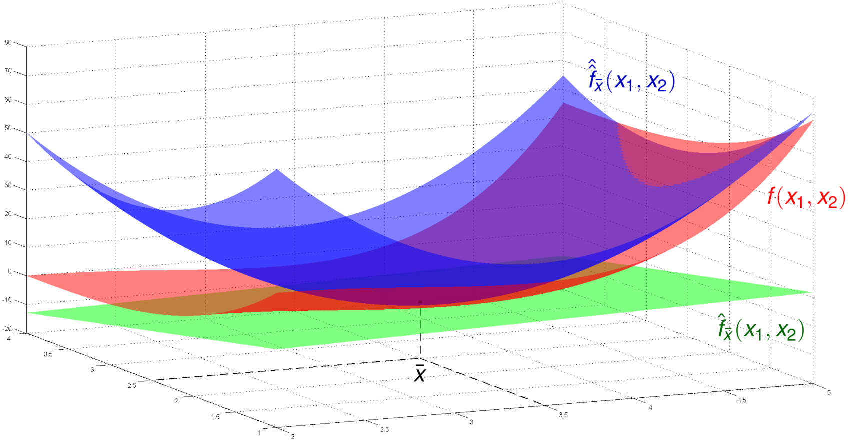
Consider the function:

$$f(x_1, x_2) = (x_1 - 2)^4 - 4(x_1 - 2)^2 - (x_2 - 4)^3.$$

The second-order approximation at the point $(3.5, 2.5)$ is:

$$\hat{\hat{f}}_{\bar{x}}(x) = f(3.5, 2.5) + \nabla f(3.5, 2.5)^T \begin{pmatrix} x_1 - 3.5 \\ x_2 - 2.5 \end{pmatrix}$$
$$+ \frac{1}{2}(x_1 - 3.5, x_2 - 2.5) H_f(3.5, 2.5) \begin{pmatrix} x_1 - 3.5 \\ x_2 - 2.5 \end{pmatrix}$$

Example in 2D (cont'd)



Taylor series

Taylor expansion at $\bar{x}$ up to second-order:

- One-variable, $\bar{x} \in \mathbb{R}$, $\alpha \in \mathbb{R}$:

$$f(\bar{x}+\alpha) = f(\bar{x}) + \alpha f'(\bar{x}) + \frac{\alpha^2}{2} f''(\bar{x}) + r(\alpha^2)$$

- Multi-variable, $\bar{x}, u \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$:

$$f(\bar{x}+\alpha u) = f(\bar{x}) + \alpha \nabla f(\bar{x})^T u + \frac{\alpha^2}{2} u^T H_f(\bar{x}) u + r(\alpha^2)$$

Here, we assume that $\|u\| = 1$: u is a unit vector that represents the direction of displacement from $\bar{x}$.

In general, the Taylor series up to k-th order has an error term $r(\alpha^k)$ with the property that $\lim_{\alpha \to 0} r(\alpha^k)/\alpha^k = 0$.

The Taylor series at a critical point

Suppose now that we have a critical point: $\nabla f(\bar{x}) = \vec{0}$. Then:

$$\begin{aligned}\hat{\hat{f}}_{\bar{x}}(x) &= f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T H_f(\bar{x})(x - \bar{x}) \\ &= f(\bar{x}) + \frac{1}{2}(x - \bar{x})^T H_f(\bar{x})(x - \bar{x}).\end{aligned}$$

For simplicity, we assume $\bar{x} = \vec{0}$. We can always translate the origin to verify this condition. We obtain:

$$\hat{\hat{f}}_{\bar{x}=\vec{0}}(x) - f(\vec{0}) = \frac{1}{2} x^T H_f(\vec{0}) x.$$

The difference in function value with respect to $\bar{x} = \vec{0}$ depends on the Hessian (up to second order).

Connection to positive definiteness

Equivalent definitions of positive definiteness

A symmetric matrix $H \in \mathbb{R}^{n\times n}$ is positive definite if $x^T Hx > 0$ for all nonzero vectors $x \in \mathbb{R}^n$. Equivalently, H is positive definite if all its eigenvalues are > 0.

Consider the function $f(x_1, x_2) = 2x_1^2 + 2x_2^2 - x_1 x_2$. The point $\bar{x} = (0, 0)$ is a critical point. The Hessian is $H_f(\vec{0}) = \begin{pmatrix} 4 & -1 \\ -1 & 4 \end{pmatrix}$. $H_f(\vec{0})$ is pd, therefore:

$$\hat{\hat{f}}_{\bar{x}}(x) - f(\vec{0}) = \frac{1}{2}(x_1, x_2)H_f(\vec{0})(x_1, x_2)^T > 0 \qquad \text{for all } x_1, x_2.$$

The point $(0, 0)$ is a minimum because the second-order approximation tells us that f increases along all directions when moving away from $(0, 0)$.

A sufficient condition

Proposition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Suppose $x^* \in S$ satisfies:

$$\nabla f(x^*) = \vec{0}, \qquad H_f(x^*) \text{ is pd.}$$

Then x^* is a (strict) unconstrained local minimum of f.

A sufficient condition

Proposition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Suppose $x^* \in S$ satisfies:

$$\nabla f(x^*) = \vec{0}, \qquad H_f(x^*) \text{ is pd.}$$

Then x^* is a (strict) unconstrained local minimum of f.

Previous discussion on the eigenvalues of the Hessian also proves this proposition.

We will take a look at the necessary condition together next.

Activity 4 (20 minutes)

Proposition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function.
If x^* is an unconstrained local minimum of f, then $H_f(x^*)$ is positive semidefinite.

Fill in the blanks in the proof of this proposition:

Choose any $u \in \mathbb{R}^n$ with $\|u\| = 1$. For all $\alpha \in \mathbb{R}$, $\alpha \geqslant 0$, the Taylor expansion yields:

$$f(x^* + \alpha u) = f(x^*) + \cdots + \cdots + \ldots$$

Bring $f(x^*)$ to the left-hand side and recall that $\nabla f(x^*) = \vec{0}$ (first-order optimality condition).

$$f(x^* + \alpha u) - f(x^*) = \cdots + \ldots$$

By definition of local optimum, there exists $\varepsilon > 0$ such that for all $\alpha \leqslant \varepsilon$ we have $f(x^* + \alpha u) \geqslant f(x^*)$, so for small α>0:

$$\cdots \leqslant \frac{f(x^* + \alpha u) - f(x^*)}{\alpha^2} = \cdots + \ldots$$

Take the limit for $\alpha \to 0$:

$$\ldots \leqslant \lim_{\alpha \to 0} \frac{f(x^* + \alpha u) - f(x^*)}{\alpha^2} = \frac{1}{2} u^T H_f(x^*) u + \lim_{\alpha \to 0} \frac{r(\alpha^2)}{\alpha^2} = \ldots$$

Because this is true for all $u \in \mathbb{R}^n$, $H_f(x^*)$ is ... by definition.

A simple optimization procedure

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable.

Based on the three propositions above, we have a method to find the minimum of f .

1. Find all points $z \in \mathbb{R}^n$ such that $\nabla f(z) = \vec{0}$.
2. Among those, find all points where $H_f(z)$0 is psd.
   - 2.1 Among those, all points where $H_f(z)$ is pd are local minima.
   - 2.2 Further investigate points where $H_f(z)$ is psd and $\det H_f(z) = 0$.

A simple optimization procedure

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable.

Based on the three propositions above, we have a method to find the minimum of f .

1. Find all points $z \in \mathbb{R}^n$ such that $\nabla f(z) = \vec{0}$.
2. Among those, find all points where $H_f(z)$0 is psd.
    2.1 Among those, all points where $H_f(z)$ is pd are local minima.
    2.2 Further investigate points where $H_f(z)$ is psd and $\det H_f(z) = 0$.

If there are local minima, they will be identified in Step 2. If there are no points left after Step 2, there is neither local nor global minimum.

The problem is either unbounded or the minimum is not attained!

Summary

- Taylor series in two dimensions.
- The Hessian function in two dimensions.
- Understanding the optimality conditions better using the Taylor series.

Derivation of second order Taylor series in 2D

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$. Denote by $F(\alpha)$:

$$F(\alpha) = f(m_0 + \alpha\, u),$$

where $m_0 = (x_0, y_0)^T$, $u = (u_1, u_2)^T$.

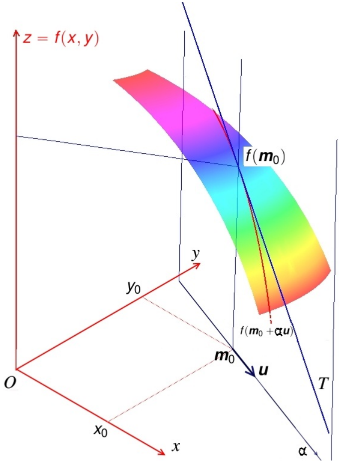
The graph of $F(\alpha)$ is the cross-section of the surface $z = f(x, y)$ at the point (x_0, y_0) along the direction u.

Recall the Taylor series for a function of one variable $F(\alpha)$ at the origin:

$$F(\alpha) = F(0) + \alpha\, F'(0) + \frac{\alpha^2}{2} F''(0) + r(\alpha^2).$$

We now compute the Taylor series of $F(\alpha)$ as defined above in terms of the function $f(x, y)$.

Derivation of second order Taylor series in 2D



Derivation of second order Taylor series in 2D

We want to use the formula:

$$F(\alpha) = F(0) + \alpha\, F'(0) + \frac{\alpha^2}{2} F''(0) + r(\alpha^2).$$

- Since $F(\alpha) = f(m_0 + \alpha\, u)$, then $F(0) = f(m_0) = f(x_0, y_0)$.
- Computing $F'(0)$:

$$F'(0) = \left.\frac{\mathrm{d}F}{\mathrm{d}\alpha}\right|_{\alpha=0} = \left.\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(f(x_0 + \alpha\, u_1, y_0 + \alpha\, u_2)\right)\right|_{\alpha=0}$$

$$= \left[\frac{\partial f(x_0 + \alpha\, u_1, y_0 + \alpha\, u_2)}{\partial x}\frac{\mathrm{d}(x_0 + \alpha\, u_1)}{\mathrm{d}\alpha} + \frac{\partial f(x_0 + \alpha\, u_1, y_0 + \alpha\, u_2)}{\partial y}\frac{\mathrm{d}(y_0 + \alpha\, u_2)}{\mathrm{d}\alpha}\right]_{\alpha=0}.$$

$$= \frac{\partial f(x_0, y_0)}{\partial x} u_1 + \frac{\partial f(x_0, y_0)}{\partial y} u_2 = \nabla f(x_0, y_0)^T u.$$

Derivation of second order Taylor series in 2D

- Computing $F''(0)$:

$$F''(0) = \left.\frac{\mathrm{d}^2 F}{\mathrm{d}\alpha^2}\right|_{\alpha=0} = \frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\big(f(x_0 + \alpha\, u_1, y_0 + \alpha\, u_2)\big)\Big|_{\alpha=0}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\alpha}\left[\frac{\partial f(x_0 + \alpha\, u_1, y_0 + \alpha\, u_2)}{\partial x}\frac{\mathrm{d}(x_0 + \alpha\, u_1)}{\mathrm{d}\alpha} + \frac{\partial f(x_0 + \alpha\, u_1, y_0 + \alpha\, u_2)}{\partial y}\frac{\mathrm{d}(y_0 + \alpha\, u_2)}{\mathrm{d}\alpha}\right]_{\alpha}$$

Now $\frac{\mathrm{d}(x_0+\alpha\, u_1)}{\mathrm{d}\alpha} = u_1$ and $\frac{\mathrm{d}(y_0+\alpha\, u_2)}{\mathrm{d}\alpha} = u_2$ so that

$$= u_1\left(\frac{\partial^2 f(x_0, y_0)}{\partial x^2}u_1 + \frac{\partial^2 f(x_0, y_0)}{\partial y \partial x}u_2\right) + u_2\left(\frac{\partial^2 f(x_0, y_0)}{\partial x \partial y}u_1 + \frac{\partial^2 f(x_0, y_0)}{\partial y^2}u_2\right).$$

In matrix form, we have:

$$F''(0) = \begin{pmatrix} u_1 & u_2 \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

Derivation of second order Taylor series in 2D

So, if $\frac{\partial^2 f(x_0,y_0)}{\partial y \partial x} = \frac{\partial^2 f(x_0,y_0)}{\partial x \partial y}$, we have either

$$F''(0) = \frac{\partial^2 f(x_0, y_0)}{\partial x^2} u_1^2 + 2\frac{\partial^2 f(x_0, y_0)}{\partial y \partial x} u_1 u_2 + \frac{\partial^2 f(x_0, y_0)}{\partial y^2} u_2^2$$

or, writing $H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$,

$$F''(0) = \begin{pmatrix} u_1 & u_2 \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = u^T H u.$$

and the Taylor series for $F(\alpha)$ becomes:

$$F(\alpha) = f(m_0 + \alpha u) = f(m_0) + \alpha \nabla f(m_0)^T u + \frac{\alpha^2}{2} u^T H(m_0) u + r(\alpha^2).$$