Multivariable Calculus **Summer 2017**

Topic 1: Taylor Series Approximation

Author: Andy Lee

1.1 Introduction

In this section, we review function approximation using taylor series. Note that this is a common technique used in many settings (e.g. iterative LQR, guided cost learning, etc.). In particular, we focus on the 1st and 2nd order taylor approximations because these are the two forms most commonly encountered in machine learning and optimal control. We begin by presenting taylor series in the single variable case and then move on to the multivariable case. In addition, we provide a full derivation of the multivariable taylor series formula by first reparametrizing the problem in the single variable case and solving it accordingly.

1.2 Motivation

Intuitively, taylor series approximation is a technique that uses a polynomial function to *approximate* a target function $f(x)$.

$$f(x) \approx \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_k x^k + \cdots$$

Typically, this is done because the target function is difficult to compute. The reason we use the name taylor series comes from the fact that this polynomial approximation takes on a particular form. Specifically, a taylor series uses the derivatives of $f(x)$ as coefficients for its higher-order terms.

Recall, the kth-order taylor polynomial for a function f around base point x_0 is given as follows.

$$f(x) \approx f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \tag{1.1}$$

Notice that in this setting, we are making the assumption that $f(x - x_0$ is difficult to compute exactly; that is, f is difficult to compute. Our technique to get around this problem is to come up with an approximate answer that leverages the derivatives of f evaluated at x. We can now get precise leverage these derivatives to get an approximate value that is close for x near x_0 and by Taylor's theorem (not covered here) we can get precise guarantees on the errors surrounding this approximation. That's the basic idea.

1.3 Taylor Series (Single Variable Calculus)

Let $g : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function and consider the taylor series approximation for f around base point x_0. Note that the error term is denoted as $o(x^k)$ (pronounced "little-oh") where k is the degree of polynomial function. The general taylor series approximation formula is given as

$$f(x) = f(x_0) + \sum_{n=1}^{k} \frac{f^{(n)}(x)}{n!}(x - x_0)^n + o(x^k) \tag{1.2}$$

For convenience, the first and second order taylor series approximations are provided below, respectively.

First Order Taylor Series Approximation

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + o(x) \tag{1.3}$$

Second Order Taylor Series Approximation

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x)}{2!}(x - x_0)^2 + o(x^2) \tag{1.4}$$

1.4 Taylor Series (Multivariable Variable Calculus)

In most applications, we will be interested in taylor series approximations for scalar-valued functions taking vector-inputs. In this section, we generalize taylor series approximations to the multivariable case.

To motivate the problem, we will be interested in determining the value $f(a + y)$. In this case, we can think of a as the base point and y as a slight pertubation. Intuitively, we can have high confidence in the approximation so long as $d(a, a + y) < \delta$, where δ represents some notion of distance threshold (usually small) that we want to stay within. In other words, the approximation only works if $a + y$ is *near* a.

Note that the notation used to denote the error term is $||y||^k E_k(a, y)$. In every case, we assume the error term tends to zero faster than $||y||^k$.

First Order Taylor Series Approximation (Multivariable)

$$f(a + y) = f(a) + \nabla f(a) \cdot y + ||y||E(a, y)$$

Second Order Taylor Series Approximation (Multivariable)

$$f(a + y) = f(a) + \nabla f(a) \cdot y + \frac{1}{2} y^T \nabla^2 f(a) y + ||y||^2 E_2(a, y)$$

1.5 Derivation of Multivariable Taylor Series Approximation

In this section, we demonstrate how we can derive the mutlivariable form of the taylor series approximation by first casting it as a single variable problem and solving it accordingly. Before diving into this section, we recommend the reader review his/her multivariable calculus concepts of differentiability (gradient, hessian) and attempt the claims as exercises before looking at the proofs as it provides a good litmus test for current abilities in calculus. Note these derivations follow closely with the derivations provided by [1] and [2].

Lemma 1.1 *Let* $g(\alpha) = f(x^* + \alpha d)$, *then* $g'(\alpha) = \nabla_x f(x^* + \alpha d) \cdot d$

Proof:
Consider $g'(a)$.

$$\begin{aligned} g'(\alpha) &= \frac{d}{d\alpha} g(\alpha) \\ &= \frac{d}{d\alpha} f(x^* + \alpha d) \text{ (by definition of } g) \\ &= \frac{\partial f}{\partial x} \frac{dx}{d\alpha}(x^* + \alpha d) \text{ (by chain rule)} \end{aligned}$$

Note that our intermediate quantities are given as.

$$\begin{aligned} \frac{\partial f}{\partial x} &= \nabla_x f \text{ (by definition of gradient)} \\ &= \left[\frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_n} \right]_{1 \times n} \end{aligned}$$

$$\begin{aligned} \frac{dx}{d\alpha} &= \frac{d}{d\alpha}(x^* + \alpha d) \\ &= \frac{d}{d\alpha} x^* + \frac{d}{d\alpha}(\alpha d) \\ &= d \end{aligned}$$

Substituting our intermediate quantities back to the original equation yields

$$g'(\alpha) = \frac{\partial f}{\partial x} \frac{dx}{d\alpha}(x^* + \alpha d) \tag{1.5}$$

$$= \nabla_x f(x^* + \alpha d) \cdot d \tag{1.6}$$

■

Lemma 1.2 *Let* $g(\alpha) = f(x^* + \alpha d)$*, then* $g''(\alpha) = d^T\left(\nabla_x^2 f(x^* + \alpha d)\right)d$

Proof: Recall, we have $g'(\alpha) = \nabla_x f(x^* + \alpha d) \cdot d$ and we wish to consider $g''(\alpha)$. Note,

$$\begin{aligned}
g''(\alpha) &= \frac{d}{d\alpha} g'(\alpha) \\
&= \frac{d}{d\alpha}\left(\nabla_x f(x^* + \alpha d) \cdot d\right) \text{ (by definition of } g'(\alpha) \text{)} \\
&= \frac{d}{d\alpha} \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} d_i
\end{aligned}$$

Fix index i and consider the term $\frac{d}{d\alpha} g$, where $g = \frac{\partial f}{\partial x_i}$. Note that we have

$$\begin{aligned}
\frac{d}{d\alpha} g &= \frac{\partial g}{\partial x} \frac{dx}{d\alpha} \\
&= \left[\frac{\partial g}{\partial x_1} \cdots \frac{\partial g}{\partial x_n}\right]_{1\times n} \frac{dx}{d\alpha} \\
&= \left[\frac{\partial g}{\partial x_1} \cdots \frac{\partial g}{\partial x_n}\right]_{1\times n} \cdot d \text{ (because } x = x^* + \alpha d) \\
&= \sum_{j=1}^{n} \frac{\partial g}{\partial x_j} d_j \\
&= \sum_{j=1}^{n} \frac{\partial f}{\partial x_i \partial x_j} d_j \text{ (by definition of } g)
\end{aligned}$$

Substituting back into our original equation, we have

$$\begin{aligned}
g''(\alpha) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial f}{\partial x_i \partial x_j} d_i d_j && (1.7) \\
&= \begin{bmatrix} d_1 & d_2 & \cdots & d_n \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_n \partial x_1} & \cdots & \frac{\partial f}{\partial x_n \partial x_n} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} && (1.8) \\
&= d^T\left(\nabla_x^2 f(x^* + \alpha d)\right)d && (1.9)
\end{aligned}$$

■

Theorem 1.3 *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a continuously differentiable function. Let* $x^* \in \mathbb{R}^n$ *and let* $d \in \mathbb{R}^n$ *be some arbitrary vector. Finally, let* $\alpha \in \mathbb{R}$*, which we can think of as a scaling factor. We wish to show the following.*

$$f(x^* + \alpha d) = f(x^*) + \left(\nabla f(x^*) \cdot d\right)\alpha + \frac{1}{2}\left(d^T\left(\nabla^2 f(x^*)\right)d\right)\alpha^2 + o(\alpha^2)$$

Proof: The proof follows a straightforward substitution of the derived quantities from the two lemmas above. ■

References

[1] Tom M. Apostol. *Calculus, Vol. 2: Multi-Variable Calculus and Linear Algebra with Applications to Differential Equations and Probability.* Wiley, 2nd edition, 6 1969.

[2] Daniel Liberzon. *Calculus of variations and optimal control theory: a concise introduction.* World Publishing Corporation, 2013.