VI. Optimization

- Notation
  - 예를 들어 $L(\boldsymbol{x}) = \|\boldsymbol{b} - A\boldsymbol{x}\|^2$ 과 같은 loss function이 주어졌을 때, $L(\boldsymbol{x})$의 값을 최소로 만드는 입력값이 $\hat{\boldsymbol{x}}$일 때 $\operatorname{argmin} L(\boldsymbol{x}) = \hat{\boldsymbol{x}}$과 같이 표현함
- 단일 변수 함수의 테일러 급수 (Taylor series)
  - 변수 x 에 대한 임의의 함수 $f(x)$ 는 다항식으로 근사 가능함
    - $f(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots$
  - 다항식의 계수 $c_0, c_1, c_2, c_3, \ldots$는 어떻게 결정?
  - 위의 다항식으로 전개한 결과가 $f(x)$와 같다면,
    - $x = 0$ 을 등식에 대입했을 때, 양변이 같아야 함 ➔ $f(0) = c_0$
    - $x = 0$ 을 등식에 대입했을 때, 양변의 1차 미분이 같아야 함 ➔ $\left.\frac{df}{dx}\right|_{x=0} = c_1$
    - $x = 0$ 을 등식에 대입했을 때, 양변의 2차 미분이 같아야 함 ➔ $\left.\frac{d^2f}{dx^2}\right|_{x=0} = c_2 \cdot 2$
    - $x = 0$ 을 등식에 대입했을 때, 양변의 3차 미분이 같아야 함 ➔ $\left.\frac{d^3f}{dx^3}\right|_{x=0} = c_3 \cdot 3 \cdot 2$
    - 일반적으로 $c_n = \left.\frac{1}{n!}\frac{d^nf}{dx^n}\right|_{x=0}$
  - $f(x) = f(0) + \left.\frac{df}{dx}\right|_{x=0} x + \left.\frac{1}{2!}\frac{d^2f}{dx^2}\right|_{x=0} x^2 + \left.\frac{1}{3!}\frac{d^3f}{dx^3}\right|_{x=0} x^3 + \cdots$

VI. Optimization

- (계속) 단일 변수 함수의 테일러 급수 (Taylor series)
  - $f(x) = f(0) + \left.\frac{df}{dx}\right|_{x=0} x + \frac{1}{2!}\left.\frac{d^2f}{dx^2}\right|_{x=0} x^2 + \frac{1}{3!}\left.\frac{d^3f}{dx^3}\right|_{x=0} x^3 + \cdots$
  - 위의 결과는 $f(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots$의 형태를 가정하고 유도하였으므로, $x = 0$ 근처에서는 정확도가 높지만 0에서 먼 위치에 대해서는 정확도가 떨어짐
  - 만약 x가 0이 아닌 위치(x_0) 근처에서 정확도가 높은 다항식을 원한다면 아래와 같은 전개 가능
    - $f(x) = d_0 + d_1(x - x_0) + d_2(x - x_0)^2 + d_3(x - x_0)^3 + \cdots$
  - 위의 다항식으로 전개한 결과가 $f(x)$와 같다면,
    - $x = x_0$ 을 등식에 대입했을 때, 양변이 같아야 함 ➔ $f(x_0) = d_0$
    - $x = x_0$ 을 등식에 대입했을 때, 양변의 1차 미분이 같아야 함 ➔ $\left.\frac{df}{dx}\right|_{x=x_0} = d_1$
    - $x = x_0$을 등식에 대입했을 때, 양변의 2차 미분이 같아야 함 ➔ $\left.\frac{d^2f}{dx^2}\right|_{x=x_0} = d_2 \cdot 2$
    - 일반적으로 $d_n = \frac{1}{n!}\left.\frac{d^nf}{dx^n}\right|_{x=x_0}$
  - $f(x) = f(x_0) + \left.\frac{df}{dx}\right|_{x=x_0}(x - x_0) + \frac{1}{2!}\left.\frac{d^2f}{dx^2}\right|_{x=x_0}(x - x_0)^2 + \frac{1}{3!}\left.\frac{d^3f}{dx^3}\right|_{x=x_0}(x - x_0)^3 + \cdots$
  - Example) $f(x) = 1 + 2x + 3x^2$를 $x = 1$ 근처에서의 다항식으로 나타내면?
    - $d_0 + d_1(x - 1) + d_2(x - 1)^2 = 6 + 8(x - 1) + \frac{1}{2!}6(x - 1)^2$
    - 만약 $f(x)$가 3차 이상의 식이었는데, 2차에서 멈췄다면 근사의 정확도가 떨어짐

VI. Optimization

- 다 변수 함수(multivariable function)의 테일러 급수 (Taylor series)
  - $f(x,y) = c_0 + c_x x + c_y y + c_{xx}x^2 + c_{xy}xy + c_{yy}y^2 + \cdots$
  - 위의 다항식으로 전개한 결과가 $f(x,y)$와 같다면,
    - $x = y = 0$을 등식에 대입했을 때, 양변이 같아야 함 ➔ $f(0,0) = c_0$
    - $x = y = 0$을 등식에 대입했을 때, 양변을 x에 대해 1차 편미분한 결과가 같아야 함 ➔ $\left.\frac{\partial f}{\partial x}\right|_{x=y=0} = c_x$
    - $x = y = 0$을 등식에 대입했을 때, 양변을 x에 대해 2차 편미분한 결과가 같아야 함 ➔ $\left.\frac{\partial^2 f}{\partial x^2}\right|_{x=y=0} = c_{xx} \cdot 2$
    - $x = y = 0$을 등식에 대입했을 때, 양변을 x에 대해 1차 편미분한 후 y에 대해 1차 편미분한 결과가 같아야 함 ➔ $\left.\frac{\partial^2 f}{\partial y \partial x}\right|_{x=y=0} = c_{xy}$
    - 일반적으로 $c_{x^m y^{(n-m)}} = \frac{1}{m!(n-m)!}\left.\frac{\partial^n f}{\partial x^m \partial y^{(n-m)}}\right|_{x=y=0} = \frac{1}{n!}\binom{n}{m}\left.\frac{\partial^n f}{\partial x^m \partial y^{(n-m)}}\right|_{x=y=0} = \frac{1}{n!}\binom{n}{m}\left(\frac{\partial}{\partial x}\right)^m\left(\frac{\partial}{\partial y}\right)^{n-m} f\Big|_{x=y=0}$
  - $f(x,y) = f(0,0) + \left.\frac{\partial f}{\partial x}\right|_{x=y=0} x + \left.\frac{\partial f}{\partial y}\right|_{x=y=0} y + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial x^2}\right|_{x=y=0} x^2 + \left.\frac{\partial^2 f}{\partial y \partial x}\right|_{x=y=0} xy + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial y^2}\right|_{x=y=0} y^2 + \cdots = \sum_{n=0}^{\infty}\frac{1}{n!}\left(x\left(\frac{\partial}{\partial x}\right) + y\left(\frac{\partial}{\partial y}\right)\right)^n f\Big|_{x=y=0}$

VI. Optimization

- 다 변수 함수(multivariable function)의 테일러 급수 (Taylor series)
  - $f(x,y) = f(0,0) + \left.\frac{\partial f}{\partial x}\right|_{x=y=0} x + \left.\frac{\partial f}{\partial y}\right|_{x=y=0} y + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial x^2}\right|_{x=y=0} x^2 + \left.\frac{\partial^2 f}{\partial y \partial x}\right|_{x=y=0} xy + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial y^2}\right|_{x=y=0} y^2 + \cdots$
  - 단일 변수 함수와 마찬가지로 임의의 위치 (x_0, y_0)에서 위의 전개를 하면,
    - $f(x,y) = f(x_0,y_0) + \left.\frac{\partial f}{\partial x}\right|_{x_0,y_0} (x - x_0) + \left.\frac{\partial f}{\partial y}\right|_{x_0,y_0} (y - y_0) + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial x^2}\right|_{x_0,y_0} (x - x_0)^2 + \left.\frac{\partial^2 f}{\partial y \partial x}\right|_{x_0,y_0} (x - x_0)(y - y_0) + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial y^2}\right|_{x_0,y_0} (y - y_0)^2 + \cdots$
  - $x - x_0 \equiv \Delta x, y - y_0 \equiv \Delta y$로 표현하고, 이들 차이값을 성분으로 하는 벡터 $\Delta \boldsymbol{x} \equiv \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$를 정의하면, 1차 다항식 항들은 아래와 같이 표현 가능
    - $\left.\frac{\partial f}{\partial x}\right|_{x_0,y_0} (x - x_0) + \left.\frac{\partial f}{\partial y}\right|_{x_0,y_0} (y - y_0) = \left.\begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}\right|_{x_0,y_0} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = (\nabla f)^T \Delta \boldsymbol{x}$

VI. Optimization

- 다 변수 함수(multivariable function)의 테일러 급수 (Taylor series)
  - $f(x,y) = f(x_0,y_0) + \left.\frac{\partial f}{\partial x}\right|_{x_0,y_0}(x-x_0) + \left.\frac{\partial f}{\partial y}\right|_{x_0,y_0}(y-y_0) + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial x^2}\right|_{x_0,y_0}(x-x_0)^2 + \left.\frac{\partial^2 f}{\partial y \partial x}\right|_{x_0,y_0}(x-x_0)(y-y_0) + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial y^2}\right|_{x_0,y_0}(y-y_0)^2 + \cdots$
  - 위 전개의 2차 다항식 항들은 아래와 같이 표현 가능
    - $\frac{1}{2!}\left.\frac{\partial^2 f}{\partial x^2}\right|_{x_0,y_0}\Delta x^2 + \left.\frac{\partial^2 f}{\partial y \partial x}\right|_{x_0,y_0}\Delta x \Delta y + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial y^2}\right|_{x_0,y_0}\Delta y^2 = \frac{1}{2}\begin{bmatrix}\Delta x & \Delta y\end{bmatrix}\begin{bmatrix}\frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2}\end{bmatrix}_{x_0,y_0}\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix} = \frac{1}{2}(\Delta \boldsymbol{x})^T H (\Delta \boldsymbol{x})$
  - H는 Hessian 행렬이라고 부르고, 주어진 스칼라 함수의 2차 편미분들을 성분으로 가지는 대칭 행렬임
  - 위의 식을 일반화하여 정리해 보면,
    - $f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}) = f(\boldsymbol{x}_0) + (\nabla f)_{\boldsymbol{x}_0}{}^T \Delta \boldsymbol{x} + \frac{1}{2}(\Delta \boldsymbol{x})^T H_{\boldsymbol{x}_0}(\Delta \boldsymbol{x}) + \cdots$