부록

Quadratic Programming 알고리즘에 필요한 수식들을 도출해 내는 방법을 조금 더 자세히 살펴보도록 한다.

$$\vec{\nabla}_{\vec{w}}\mathcal{L}(\vec{w},\vec{\lambda}) + \tilde{\nabla}^2_{\vec{w}}\mathcal{L}(\vec{w},\vec{\lambda})\overrightarrow{\Delta w} + \tilde{\nabla}^2_{\vec{w},\vec{\lambda}}\mathcal{L}(\vec{w},\vec{\lambda})\overrightarrow{\Delta\lambda} = \vec{0} \tag{G.1}$$

$$\vec{\nabla}_{\vec{\lambda}}\mathcal{L}(\vec{w},\vec{\lambda}) + \tilde{\nabla}^2_{\vec{w},\vec{\lambda}}\mathcal{L}(\vec{w},\vec{\lambda})\overrightarrow{\Delta w} = \vec{0} \tag{G.2}$$

여기에서 라그랑주함수 $\mathcal{L}(\vec{w},\vec{\lambda})$는 다음과 같이 정의한다.

$$\mathcal{L}(\vec{w},\vec{\lambda}) = f(\vec{w}) - \vec{\lambda}\cdot\vec{d}(\vec{w}) \tag{G.3}$$

$$f(\vec{w}) = \frac{1}{2}\vec{w}^t\tilde{\Sigma}\vec{w} - \frac{1}{c}\sum_{i=1}^{N} Log(w_i) \tag{G.4}$$

$$\vec{d}(\vec{w}) = \tilde{A}\vec{w} - \vec{b} \tag{G.5}$$

그러면, 선형화(線形化)를 위한 첫번째 단계로써 라그랑주함수를 테일러 급수로 전개하기로 한다. 테일러 급수는 임의의 좌표 $(\vec{w}_0,\vec{\lambda}_0)$를 중심으로 하는 함수의 근접 표기이다. 여기에서$(\vec{w},\vec{\lambda}) = (\vec{w}_0,\vec{\lambda}_0) + (\overrightarrow{\Delta w},\overrightarrow{\Delta\lambda})$가 된다.

$$\mathcal{L}(\vec{w},\vec{\lambda}) \cong \mathcal{L}(\vec{w}_0,\vec{\lambda}_0) + \vec{\nabla}_{\vec{w}}\mathcal{L}(\vec{w}_0,\vec{\lambda}_0)\overrightarrow{\Delta w} + \vec{\nabla}_{\vec{\lambda}}\mathcal{L}(\vec{w}_0,\vec{\lambda}_0)\overrightarrow{\Delta\lambda} \tag{G.6}$$

$$+\frac{1}{2}\tilde{\nabla}^2_{\vec{w}}\mathcal{L}(\vec{w}_0,\vec{\lambda}_0)\overrightarrow{\Delta w}\cdot\overrightarrow{\Delta w}+\tilde{\nabla}^2_{\vec{w},\vec{\lambda}}\mathcal{L}(\vec{w}_0,\vec{\lambda}_0)\overrightarrow{\Delta w}\cdot\overrightarrow{\Delta \lambda}$$

$$+\frac{1}{2}\tilde{\nabla}^2_{\vec{\lambda}}\mathcal{L}(\vec{w}_0,\vec{\lambda}_0)\overrightarrow{\Delta \lambda}\cdot\overrightarrow{\Delta \lambda}$$

그리고, 수식(G.6)의 개개의 항에 나타나는 미분은 수식(G.3)~(G.4)를 적용해서 표현하면 다음과 같다.

$$\vec{\nabla}_{\vec{w}}\mathcal{L}(\vec{w}_0,\vec{\lambda}_0)\cong\tilde{\Sigma}\vec{w} \tag{G.7a}$$

$$\vec{\nabla}_{\vec{\lambda}}\mathcal{L}(\vec{w}_0,\vec{\lambda}_0)=-\vec{d}(\vec{w}) \tag{G.7b}$$

$$\tilde{\nabla}^2_{\vec{w}}\mathcal{L}(\vec{w}_0,\vec{\lambda}_0)=\tilde{\Sigma}+\frac{1}{c}\widetilde{\Phi} \tag{G.7c}$$

$$\tilde{\nabla}^2_{\vec{w},\vec{\lambda}}\mathcal{L}(\vec{w}_0,\vec{\lambda}_0)=-\tilde{\nabla}_{\vec{w}}\vec{d}(\vec{w})=-\tilde{A} \tag{G.7d}$$

$$\tilde{\nabla}^2_{\vec{\lambda}}\mathcal{L}(\vec{w}_0,\vec{\lambda}_0)=\vec{0} \tag{G.7e}$$

그리고, 수식(G.6)과 같은 라그랑주 함수의 극점을 구하면 최적화 방향으로 향하는 해 $(\overrightarrow{\Delta w},\overrightarrow{\Delta \lambda})$를 얻게 된다. 극점의 조건은 다음과 같다.

$$\begin{aligned}\vec{\nabla}_{\vec{w}}\mathcal{L}(\vec{w},\vec{\lambda})&=\vec{0}\\ \vec{\nabla}_{\vec{\lambda}}\mathcal{L}(\vec{w},\vec{\lambda})&=\vec{0}\end{aligned} \tag{G.9}$$

극점의 조건 중의 하나 $\vec{\nabla}_{\vec{w}}\mathcal{L}(\vec{w},\vec{\lambda}) = \vec{0}$를 수식(G.6)에 적용하면 수식(G.1)이 되고, 또다른 극점의 조건 $\vec{\nabla}_{\vec{\lambda}}\mathcal{L}(\vec{w},\vec{\lambda}) = \vec{0}$을 수식(G.6)에 적용하면 수식(G.2)가 된다. 여기에서 좌표 $(\vec{w}_0,\vec{\lambda}_0)$의 서브인덱스 0 을 생략 표기한 것을 알 수 있다.

수식(G.1)과 수식(G.2)를 확인했으니, 이제는 수식(G.7a)~수식(G.7e)를 수식(G.1)과 수식(G.2)에 대입해 보면 다음과 같게 된다.

$$\tilde{\Sigma}\vec{w} + \left(\tilde{\Sigma} + \frac{1}{c}\tilde{\Phi}\right)\overrightarrow{\Delta w} - \tilde{A}^t\overrightarrow{\Delta\lambda} = \vec{0} \tag{G.10}$$

$$-\vec{d}(\vec{w}) - \tilde{A}\,\overrightarrow{\Delta w} = \vec{0} \tag{G.11}$$

위의 수식(G.10)과 수식(G.11)를 조금 더 함축적으로 표기하면 다음과 같다.

$$\begin{pmatrix} \tilde{\Sigma} + \frac{1}{c}\tilde{\Phi} & -\tilde{A}^t \\ \tilde{A} & \vec{0} \end{pmatrix}\begin{pmatrix} \overrightarrow{\Delta w} \\ \overrightarrow{\Delta\lambda} \end{pmatrix} = -\begin{pmatrix} \tilde{\Sigma}\vec{w} \\ \vec{d}(\vec{w}) \end{pmatrix} \tag{G.12}$$

이 수식이 바로 Quadratic Programming 알고리즘의 기초가 된다.