10장 부록

보충 설명: 10.1.3절

이차 조건(Second-order Conditions)

최적화 문제에서 어떤 점이 **국소 최적해(local optimum)**인지 판단하려면, 보통 1차 미분 조건(KKT 조건)뿐만 아니라 **2차 조건**도 함께 고려해야 한다. 특히, 목적함수와 제약함수가 두 번 미분 가능할 때, **라그랑지안 함수의 곡률(curvature)**을 분석함으로써 그 점이 진짜로 최적인지를 더 정밀하게 평가할 수 있다. 아래 정리는 이러한 **이차 조건**을 명시한 것이이다.

정리 10A.1. [2차 조건(Second-Order Conditions)] [^Nocedal et al. (2006)]

함수 f, g_i, h_j가 모두 **두 번 미분 가능**하다고 하자. 어떤 실현가능한 점 $\boldsymbol{x}^*$에서 KKT 조건을 만족하는 라그랑주 승수 $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$가 존재한다고 가정하면, 다음 두 가지 조건이 성립한다.

1. 이차 필요 조건 (Second-order Necessary Conditions)

먼저, 만약 $\boldsymbol{x}^*$가 국소 최적점이고, **LICQ 조건**도 만족한다면, 다음 조건이 반드시 성립해야 한다:

$$\boldsymbol{w}^\top \nabla_{xx} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \boldsymbol{w} \geq 0 \tag{1}$$

여기서 $\boldsymbol{w}$는 특정한 조건을 만족하는 방향 벡터인데, 이 방향들은 다음을 만족해야 한다:

- 모든 등식 제약식 h_j에 대해 $\nabla h_j(\boldsymbol{x}^*)^\top \boldsymbol{w} = 0$
- 활성화된 부등식 제약식 중에서 $\lambda_i^* > 0$인 경우에는 $\nabla g_i(\boldsymbol{x}^*)^\top \boldsymbol{w} = 0$
- $\lambda_i^* = 0$인 경우에는 $\nabla g_i(\boldsymbol{x}^*)^\top \boldsymbol{w} \leq 0$

즉, 이 방향들은 현재 점에서 제약조건을 위반하지 않으면서도, 목적함수 값이 어떻게 변하는지를 볼 수 있는 '허용된' 변화 방향이다. 이 방향에 대해서 라그랑지안 함수의 **이차 도함수(즉 곡률)**가 0 이상이어야만 국소 최소점이 될 수 있다.

2. 이차 충분 조건 (Second-order Sufficient Conditions)

이번에는 반대로, 만약 위와 같은 허용된 방향 $\boldsymbol{w}$에 대해 다음 부등식이 항상 성립한다면:

$$\boldsymbol{w}^\top \nabla_{xx} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \boldsymbol{w} > 0 \tag{2}$$

$\boldsymbol{x}^*$는 그 문제의 **엄격한(strict) 국소 최적해**가 된다. 즉, 단순히 평평한 지점이 아니라, 정말로 그 근처에서 최소인 점임을 보장할 수 있다.

이차 충분조건은 **LICQ 같은 제약조건 자격 조건(CQ)**이 없어도 성립한다. 즉, 더 강한 결과를 주면서도, 필요조건보다 적용 범위가 더 넓을 수도 있다는 뜻이다.

> **예제 10A.1.** [^Nocedal et al. (2006)]

다음 예제를 살펴보자

$$\begin{array}{ll} \text{minimize} & (x_1-1)^2+x_2^2 \\ \text{subject to} & -2kx_1+x_2^2 \geq 0, \end{array} \tag{3}$$

먼저, $\nabla c_1(\boldsymbol{x}^*) = (2k, -2x_2)^\top$는 모든 $k > 0$에 대해 LICQ 조건을 만족하며, 다음을 쉽게 확인할 수 있다.

$$\nabla_x L(x,\lambda) = \begin{bmatrix} 2(x_1-1)+2\lambda k \\ 2x_2 - 2\lambda x_2 \end{bmatrix}, \qquad \nabla_{xx} L(x,\lambda) = \begin{bmatrix} 2 & 0 \\ 0 & 2-2\lambda \end{bmatrix} \tag{4}$$

(Case-I: $k>1$) $\bar{\lambda}^1 = 1/k$일 때 $\bar{\boldsymbol{x}}^1 = (0,0)^\top$만 KKT 조건을 만족한다. $\nabla_{xx} L(\bar{\boldsymbol{x}}^1, \bar{\lambda}^1)$가 양의 정부호이기 때문에, 이차 충분조건은 이 점에서 성립하며, 따라서 $\bar{\boldsymbol{x}}^1$는 엄격한 국소 최소점이다.
(Case-II: $0<k<1$) KKT 조건이 $(\bar{\boldsymbol{x}}^1, \bar{\lambda}^1) = (0,0,1/k)$, $(\bar{\boldsymbol{x}}^2, \bar{\lambda}^2) = (1-k, \sqrt{2k(1-k)}, 1)$, 그리고 $(\bar{\boldsymbol{x}}^3, \bar{\lambda}^3) = (1-k, -\sqrt{2k(1-k)}, 1)$에서 성립한다. $\bar{\lambda} > 0$이기 때문에 다음을 만족한다.

$$\mathcal{C}(\boldsymbol{x}^*, \bar{\lambda}) = \{\boldsymbol{w} \neq 0 | kw_1 = \bar{\boldsymbol{x}}_2 w_2\}. \tag{5}$$

이차 필요 조건에 의해 임의의 $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*)$에 대해,

$$\boldsymbol{w}^\top \nabla_{xx} L(\bar{\boldsymbol{x}}, \bar{\lambda}) \boldsymbol{w} = 2w_1^2 + 2(1-\lambda)w_2^2 \geq 0 \tag{6}$$

이다. $(\bar{\boldsymbol{x}}_1, \bar{\lambda}_1)$에 대해서, w_2를 충분히 크게 만들면 조건을 위반할 수 있다. 임의의 $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*)$에 대해 $w_1 \neq 0$ 이므로, $(\bar{\boldsymbol{x}}_2, \bar{\lambda}_2)$와 $(\bar{\boldsymbol{x}}_3, \bar{\lambda}_3)$에 대해 $\boldsymbol{w}^\top \nabla_{xx} L \boldsymbol{w} = 2w_1^2 > 0$ 이다.

보충 설명: 10.2.2절

라그랑주 승수와 민감도(Lagrange Multiplier and Sensitivity)

최적화 문제에서 **라그랑주 승수(Lagrange multiplier)**는 단순히 수학적인 도구를 넘어서, **제약조건이 목적함수에 얼마나 영향을 주는지**를 측정하는 중요한 의미를 가진다. 다시 말해, 각 제약조건이 존재할 때 목적함수의 최적값이 **얼마나 민감하게 변하는지(sensitivity)**를 나타낸다.

예를 들어 어떤 부등식 제약 $g_i(\boldsymbol{x}) \leq 0$에 대해 라그랑주 승수 λ_i^*가 클수록, 해당 제약조건이 목적함수 $f(\boldsymbol{x})$에 더 큰 영향을 준다는 뜻이다. 직관적으로 보면, 목적함수를 '눌러서(pushing)' 혹은 '잡아당겨서(pulling)' 최적점을 형성하고 있는 것이다.

일반적으로 최적화 문제에서는 다음과 같은 상황을 고려한다:

- 주어진 목적함수 $f(\boldsymbol{x})$를 최소화하되,
- 제약조건은 $g_i(\boldsymbol{x}) \leq 0$, $h_j(\boldsymbol{x}) = 0$ 형태로 주어진다.

이때, 최적해 $\boldsymbol{x}^*$와 함께 **라그랑주 승수** (λ^*, ν^*)도 존재하며, 이들이 **KKT 조건**을 만족한다고 가정한다. 또한 $\boldsymbol{x}^*$는 정칙점이며 활성 제약조건에 대해 **2차 충분조건**도 만족된다고 가정한다.

> **정리 10-2.6. [민감도 정리(Sensitivity Theorem)]**

이제 제약조건을 약간 변경한 다음 문제를 생각해보자:

$$\begin{array}{rll} \text{minimize} & f(\boldsymbol{x}) & \\ \text{subject to} & g_i(\boldsymbol{x}) + u_i \leq 0, & i = 1, \dots, l \\ & h_j(\boldsymbol{x}) + v_j = 0, & j = 1, \dots, m \end{array} \tag{7}$$

즉, 제약조건을 u_i, v_j만큼 살짝 밀어낸 형태다. 이 문제의 해를 $\boldsymbol{x}^*(u, v)$라고 하고, $(u, v) = (0, 0)$일 때의 해는 원래 해 $\boldsymbol{x}^*$와 동일하다고 본다. 그러면 다음이 성립한다:

$$\frac{\partial f(\boldsymbol{x}^*(u, v))}{\partial u_i}\bigg|_{(u,v)=(0,0)} = \lambda_i^*, \quad \frac{\partial f(\boldsymbol{x}^*(u, v))}{\partial v_j}\bigg|_{(u,v)=(0,0)} = \nu_j^* \tag{8}$$

즉, 제약조건의 우변이 변화할 때 목적함수의 최적값이 어떻게 변하는지를 **라그랑주 승수**가 정확히 나타낸다.

증명 개요 (Sketch of Proof)

1. 활성화되지 않은 제약조건의 경우
   만약 $g_i(\boldsymbol{x}^*) < 0$인 제약조건이라면, 현재 최적해에서 이 제약은 영향을 주지 않는다. 이 경우 라그랑주 승수 $\lambda_i^* = 0$이며, 제약조건을 u_i 만큼 바꿔도 최적해나 목적함수 값에는 변화가 없다.
2. 활성화된 제약조건의 경우
   이제 $g_i(\boldsymbol{x}^*) = 0$인 활성 제약조건 i를 생각하자. 이 제약을 다음처럼 살짝 변경한다고 하자:

   $$g_i(\boldsymbol{x}) + \epsilon_i \|\nabla g_i(\boldsymbol{x}^*)\| \leq 0 \tag{9}$$

   이때 ϵ_i가 충분히 작으면 새로운 최적해 $\boldsymbol{x}^*(\epsilon_i)$는 원래의 활성 제약조건 집합을 그대로 유지한다고 가정할 수 있다. 즉, $g_i(\boldsymbol{x}^*(\epsilon_i)) = -\epsilon_i \|\nabla g_i(\boldsymbol{x}^*)\|$. 또한 해의 변화량은 다음처럼 일차 근사된다:

   $$g_i(\boldsymbol{x}^*(\epsilon_i)) - g_i(\boldsymbol{x}^*) \approx \nabla g_i(\boldsymbol{x}^*)^\top (\boldsymbol{x}^*(\epsilon_i) - \boldsymbol{x}^*) = -\epsilon_i \|\nabla g_i(\boldsymbol{x}^*)\| \tag{10}$$

   등식 제약조건에 대해서도 마찬가지로 근사할 수 있다:

   $$h_j(\boldsymbol{x}^*(\epsilon_i)) - h_j(\boldsymbol{x}^*) \approx \nabla h_j(\boldsymbol{x}^*)^\top (\boldsymbol{x}^*(\epsilon_i) - \boldsymbol{x}^*) = 0 \tag{11}$$

3. 목적함수의 변화 근사

이제 목적함수의 변화도 일차적으로 근사해보면 다음과 같다:

$$
\begin{aligned}
f(\boldsymbol{x}^*(\epsilon_i)) - f(\boldsymbol{x}^*) &\approx \nabla f(\boldsymbol{x}^*)^\top(\boldsymbol{x}^*(\epsilon_i) - \boldsymbol{x}^*) \\
&= -\sum_{j\in\mathcal{A}_{x^*}} \lambda_j^* \nabla g_j(\boldsymbol{x}^*)^\top(\boldsymbol{x}^*(\epsilon_i) - \boldsymbol{x}^*) - \sum_{k\in\mathcal{E}} \nu_k^* \nabla h_k(\boldsymbol{x}^*)^\top(\boldsymbol{x}^*(\epsilon_i) - \boldsymbol{x}^*) \\
&\approx \epsilon_i \|\nabla g_i(\boldsymbol{x}^*)\| \lambda_i^*
\end{aligned}
$$

위의 근사를 정리하면 다음과 같이 정리된다:

$$
\frac{df(\boldsymbol{x}^*(\epsilon_i))}{d\epsilon_i} = \lambda_i^* \tag{12}
$$

따라서 λ_i^*는 g_i 제약조건이 조금 완화될 때 목적함수 값이 얼마나 감소하는지를 정확히 나타낸다. 마찬가지로 등식 제약에 대한 민감도도 ν_j^*로 주어진다.

엄밀한 증명 (Proof):

가정

- 목적함수 f, 제약조건 $g_i, h_j \in \mathcal{C}^2$ (두 번 연속 미분 가능)
- $(u, v) = (0, 0)$ 근방에서 최적해 $\boldsymbol{x}^*(u, v)$는 존재하며 연속적으로 변화함
- $\boldsymbol{x}^* := \boldsymbol{x}^*(0, 0)$는 원래 문제의 최적해이며 KKT 조건을 만족함
- 활성 제약조건 집합 $\mathcal{A}_{\boldsymbol{x}^*}$는 변하지 않음 (즉, 활성 조건이 유지된다고 가정)

활성화되지 않은 제약조건의 경우
만약 $g_i(\boldsymbol{x}^*) < 0$인 제약조건이라면, 현재 최적해에서 이 제약은 영향을 주지 않는다. 이 경우 라그랑주 승수 $\lambda_i^* = 0$이며, 제약조건을 u_i 만큼 바꿔도 최적해나 목적함수 값에는 변화가 없다.

활성화된 제약조건의 경우

변형된 문제에 대해 라그랑지안은 다음과 같다:

$$
\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}; \boldsymbol{u}, \boldsymbol{v}) = f(\boldsymbol{x}) + \sum_{i=1}^{l} \lambda_i (g_i(\boldsymbol{x}) + u_i) + \sum_{j=1}^{m} \nu_j (h_j(\boldsymbol{x}) + v_j) \tag{13}
$$

최적해 $\boldsymbol{x}^*(\boldsymbol{u}, \boldsymbol{v})$는 위 문제의 해이므로, KKT 조건에 따라 다음이 성립한다:

$$
\begin{gathered}
\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*(u, v), \boldsymbol{\lambda}^*(u, v), \boldsymbol{\nu}^*(u, v)) = 0 \quad \Rightarrow \\
\nabla f(\boldsymbol{x}^*(u, v)) + \sum_{i=1}^{l} \lambda_i^* \nabla g_i(\boldsymbol{x}^*(u, v)) + \sum_{j=1}^{m} \nu_j^* \nabla h_j(\boldsymbol{x}^*(u, v)) = 0
\end{gathered} \tag{14}
$$

목적함수의 최적값을 다음과 같이 정의하자:

$$
\varphi(u, v) := f(\boldsymbol{x}^*(u, v)) \tag{15}
$$

그럼 미분은 다음과 같이 chain rule을 적용하여 얻는다:

$$\frac{\partial \varphi}{\partial u_i} = \nabla f(\boldsymbol{x}^*)^\top \frac{\partial \boldsymbol{x}^*(u,v)}{\partial u_i} \tag{16}$$

하지만 $\nabla f(\boldsymbol{x}^*) = -\sum \lambda_i^* \nabla g_i(\boldsymbol{x}^*) - \sum \nu_j^* \nabla h_j(\boldsymbol{x}^*)$ 이므로,

$$\frac{\partial \varphi}{\partial u_i} = -\sum_{k=1}^{l} \lambda_k^* \nabla g_k(\boldsymbol{x}^*)^\top \frac{\partial \boldsymbol{x}^*}{\partial u_i} - \sum_{j=1}^{m} \nu_j^* \nabla h_j(\boldsymbol{x}^*)^\top \frac{\partial \boldsymbol{x}^*}{\partial u_i} \tag{17}$$

한편, 활성화된 제약조건 $g_i(\boldsymbol{x}) + u_i = 0$에 대해,

$$\frac{\partial}{\partial u_i}(g_k(\boldsymbol{x}^*(u,v)) + u_k) = \nabla g_k(\boldsymbol{x}^*)^\top \frac{\partial \boldsymbol{x}^*}{\partial u_i} + \delta_{ik} = 0 \tag{18}$$

(여기서 δ_{ik}는 크로네커 델타)

정리하면,

$$\frac{\partial \varphi}{\partial u_i} = \lambda_i^* \tag{19}$$

동일하게,

$$\frac{\partial \varphi}{\partial v_j} = \nu_j^* \tag{20}$$

정리하자면, 다음이 성립한다:

$$\left.\frac{\partial f(\boldsymbol{x}^*(\boldsymbol{u},\boldsymbol{v}))}{\partial u_i}\right|_{(u,v)=(0,0)} = \lambda_i^*, \quad \left.\frac{\partial f(\boldsymbol{x}^*(\boldsymbol{u},\boldsymbol{v}))}{\partial v_j}\right|_{(u,v)=(0,0)} = \nu_j^* \tag{21}$$

이는 **라그랑주 승수는 최적 목적값의 민감도(기울기)를 직접 나타낸다**는 것을 증명한 것이다. ¶

민감도 해석 (Remark on Sensitivity)

최적화 문제의 해석에서 라그랑주 승수 λ_i^* (또는 등식 제약일 경우 ν_j^*)는 단순히 수학적 보조변수가 아니라, 실제로 최적해와 목적함수가 제약조건에 얼마나 **민감하게 반응하는지**를 정량적으로 나타낸다.

- 만약 λ_i^*가 **크다**면, 이는 i번째 제약조건이 목적함수의 최적값에 **큰 영향**을 미친다는 것을 의미한다. 즉, 그 제약조건이 약간만 완화되거나 강화되어도 최적값이 크게 바뀐다.
- 반면, λ_i^*가 **작거나 0에 가깝다**면, 그 제약조건은 최적값에 별다른 영향을 주지 않는다. 특히 $\lambda_i^* = 0$이면, 해당 제약을 조금 변형시켜도 목적함수 값은 **일차적으로 아무 변화도 없다**는 뜻이다. 이는 민감도(기울기)가 0이기 때문이다.

그림자 가격 (Shadow Price)의 경제학적 해석

이제 이 민감도 개념을 경제적 맥락에서 해석해보자. 다음과 같은 설정을 생각하자:

- 변수 $\boldsymbol{x}$: 기업이 선택할 수 있는 **운영 방식** 또는 **생산 전략**을 의미한다.
- 목적함수 $f(\boldsymbol{x})$: 최소화할 비용 함수 (혹은 부호를 바꿔 최대화할 수익 함수).
- 부등식 제약조건 $g_i(\boldsymbol{x}) \leq 0$: 각종 **자원 제약**을 의미. 예: 노동력, 원자재, 장비, 창고 공간 등.
- **Slater 조건**이 만족된다고 가정하자 (즉, 실현가능한 interior 해가 존재).

이때, 각 자원 제약조건의 우변을 **소량 변화**시켰을 때(즉, 자원의 허용량을 약간 증가 또는 감소시킬 때), 목적함수(비용 또는 이윤)는 어떻게 변할까?

이를 수학적으로 분석하면 다음과 같은 식을 얻는다:

$$\lambda_i^* = \frac{\partial f(\boldsymbol{x}^*(u))}{\partial u_i}\Big|_{u=0} \tag{22}$$

즉, λ_i^*는 **자원 i를 소량 증감시켰을 때 목적함수 값(비용 혹은 이윤)의 민감도**를 나타낸다. 경제학적으로 이 수치는 **그림자 가격(Shadow Price)** 또는 **자연 가격(Natural Price)**이라고 부른다.

예시: 자원 거래와 수익 전략

이제 이 그림자 가격의 의미를 실생활에 비춰 보자.

- 회사가 자원 i를 보유하고 있다고 하자.
- 현재 자원의 허용량이 최적 운영 방식을 결정짓고 있으며, 이에 대한 그림자 가격이 λ_i^*라고 하자.
- 만약 시장에서 자원 i를 λ_i^***보다 낮은 가격에 구매**할 수 있다면?
  → 자원을 더 많이 확보하여 생산량을 늘리면, **구매비용보다 더 많은 수익**을 낼 수 있다.
- 반대로, 자원 i를 λ_i^***보다 높은 가격에 판매**할 수 있다면?
  → 자원의 일부를 외부에 팔고 생산을 줄이더라도, **손실보다 더 큰 수익**을 얻을 수 있다.

즉, λ_i^*는 **회사가 자원을 보유하거나 거래할 때의 기준 가격선**이 된다. 이 값을 중심으로 자원의 매입과 매각에 대한 의사결정을 내릴 수 있으며, 이는 자원 배분 최적화의 기본 원리다.

각주