선형계획법

- 문제

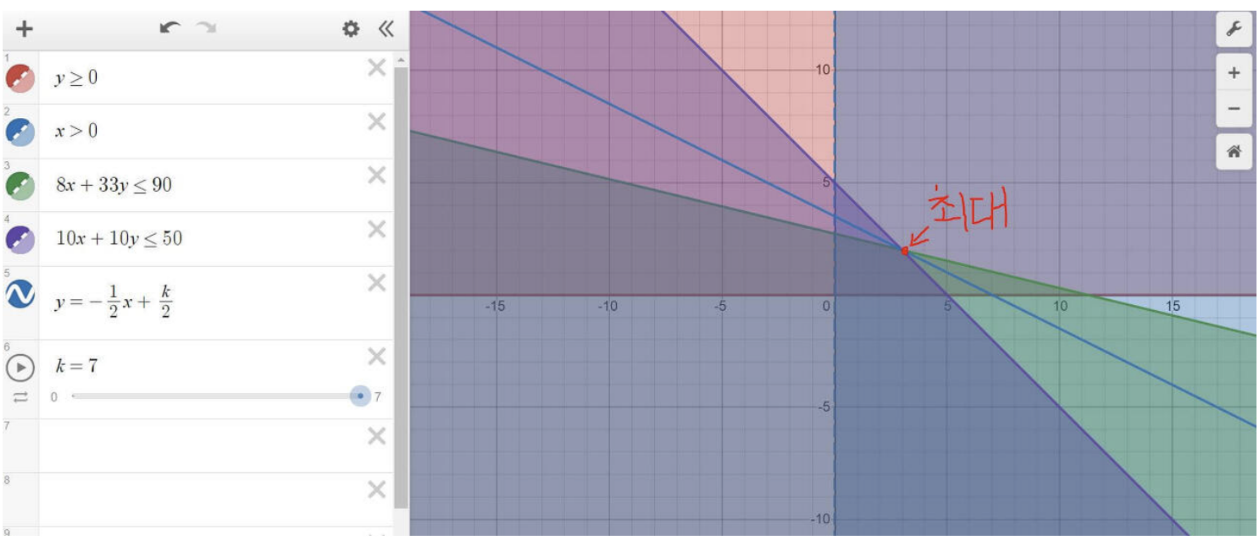
> 어느 회사에서 A제품 생산시 개당 8시간 필요하고, 10kg의 원료가 필요하다.
> B제품 생산 시 개당 33시간이 필요하고, 10kg의 원료가 필요하다.
> 제품을 만드는데 주어진 시간은 90시간 이하이고, 총 원료는 50kg 이하로 사용하고자 한다.
> A제품 한 개 판매 시 이익금은 1만원, B제품 한 개 판매시 이익금은 2만원이다.
> A제품과 B제품을 각각 몇 개 만들 때 이익이 최대가 되고, 최대이익금은 얼마인가?

- 제약조건

> (1) x>=0, y>=0
> (2) 8x+3y <= 90
> (3) 10x+10y <= 50

- 목적함수

- > O(x,y) = x+2y



일반화

$$\begin{array}{ll} \text{Find a vector} & \mathbf{x} \\ \text{that maximizes} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} \leq \mathbf{b} \\ \text{and} & \mathbf{x} \geq \mathbf{0}. \end{array}$$

Solution: simplex form

아래와 같은 문제가 있다고 가정해보자. f가 objective fuction.

$$\begin{aligned} 5x+2y &\le 900 \\ 8x+10y &\le 2800 \\ -3x-2y+f &= 0 \\ x,y &\ge 0 \end{aligned}$$

문제를 만족하는 u, v(slack variables)가 존재한다는 가정으로 수식에 도입하여 부등식을 등식으로 변경.

$$\begin{aligned} 5x+2y+u+0+0 &= 900 \\ 8x+10y+0+v+0 &= 2800 \\ -3x-2y+0+0+f &= 0 \\ x,y,u,v &\ge 0 \end{aligned}$$

Matrix form 으로 바꾸어서 생각하면,

$$\begin{aligned} R &= \begin{bmatrix} 5 & 2 & 1 & 0 & 0 & 900 \\ 8 & 10 & 0 & 1 & 0 & 2800 \\ -3 & -2 & 0 & 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix} \\ \text{set) } \mathbf{x} &= [x,y,u,v]^T \end{aligned}$$

Gauss-Jordan Elimination을 이용하여,

$$R_1-(2/5)R1,\ R3+(4/5)R1 \implies \begin{bmatrix} 1 & 0 & 25/85 & -1/17 & 0 & 100 \\ 0 & 1 & -4/17 & 5/34 & 0 & 200 \\ 0 & 0 & 2/15 & 2/17 & 1 & 700 \end{bmatrix}$$

식을 다시 정리하면,

$$\begin{aligned} x+0y+\frac{25}{85}u-\frac{1}{17}v+0 &= 900 \\ 0x+y-\frac{4}{17}u+\frac{5}{34}v+0 &= 2800 \\ 0x+0y+\frac{7}{17}u+\frac{2}{17}v+f &= 700 \end{aligned}$$

따라서 위의 수식은 $u, v \ge 0$이므로 $u=v=0$에서 700을 최소로 가짐.

이차계획법

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2}\mathbf{x}^{\mathrm{T}}Q\mathbf{x}+\mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} \le \mathbf{b} \end{aligned}$$

위의 문제를 푸는 것으로 정의됨

아래와 같은 솔루션을 가지고 있음

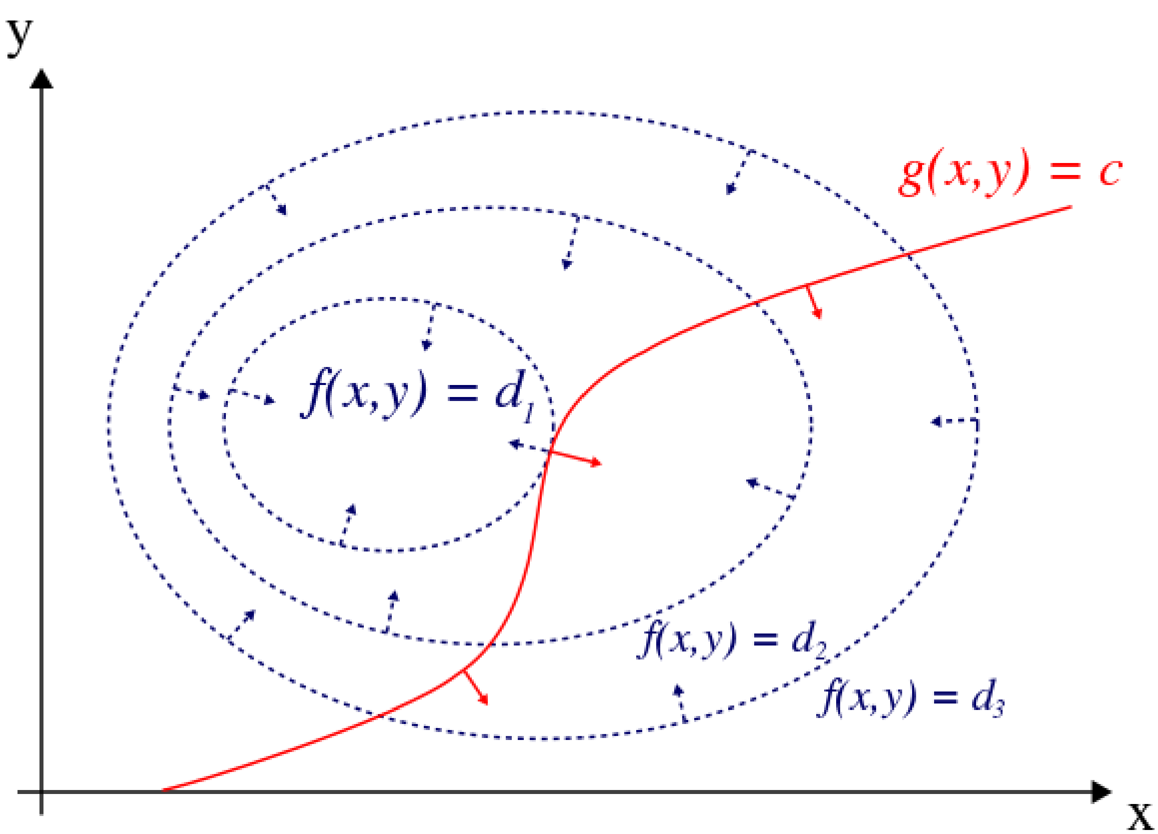
- interior point,
- active set
- augmented Lagrangian
- conjugate gradient
- gradient projection
- extensions of the simplex algorithm.

Lagrange multiplier method(라그랑주 승수법)

Objective

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}) \\ \text{subject to:} \quad & g(\mathbf{x}) = 0 \end{aligned}$$

Idea



objective function과 intercept할 수 있게 하는 c 가 존재하면 위의 그림처럼 $g(\mathbf{x}, \mathbf{y}) = c$ 를 표현 할 수 있음. 여기서 중요한 것은, intercept에서 gradient가 동일하다는 점임. 수식으로 표현하면,

$$\nabla_{x,y} f = \lambda \nabla_{x,y} g$$

특정 λ와 $\nabla_{x,y} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right), \quad \nabla_{x,y} g = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right)$ 으로 정의됨

그러므로 문제는 위의 수식을 만족하는 x, y, λ를 찾는 문제로 바뀜(a.k.a. duality)

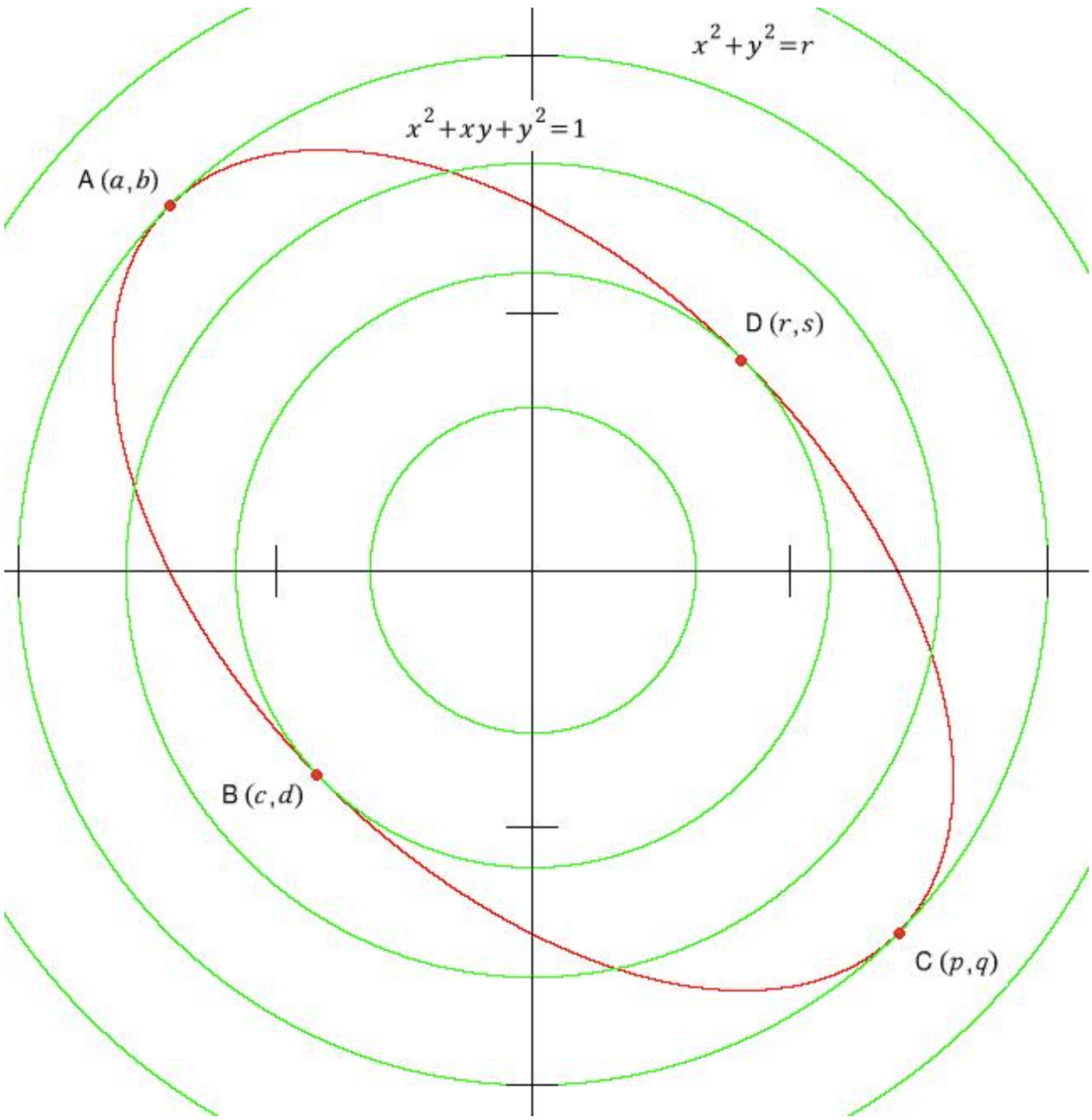
e.g.

> Solve maximum value of $x^2 + y^2$ when $x^2 + xy + y^2 = 1$

- Possible solutions

    1. substitute x to y
    2. factorize
    3. use gradient

- grdadient



Algorithm

아래와 같이 수식을 정의

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda g(x, y),$$

최대 혹은 최소가 되는 지점에서는 $\nabla_{x,y,\lambda}\mathcal{L}(x, y, \lambda) = 0$ 를 만족해야 함. 이 문제를 다시 생각하면 아래와 같이 해석 할 수 있음.

$$\nabla_{x,y,\lambda}\mathcal{L}(x, y, \lambda) = 0 \iff \begin{cases} \nabla_{x,y} f(x, y) = \lambda \nabla_{x,y} g(x, y) \\ g(x, y) = 0 \end{cases}$$

constraint qualification

- $\nabla g \neq 0$여야만 함
- $\nabla g = 0$에서는 라그랑주를 통해서 해를 구할 수 없음.

Lagrange multiplier with an inequality

- Objective

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq 0 \end{aligned}$$

풀고 싶음. 하지만 앞서 봤던 선형계획법과 동일하지만, 더 이상 선형이 아니고 선형화하기도 어려운 함수(고차원)라고 가정. 즉, 라그랑주 등 수학적 방법을 동원하여 finite solution을 구해야 함. 하지만 라그랑주는 inequlity에서는 동작하지 않음.

Duality in the linear system(back to the linear problem)

$$\begin{aligned} \text{minimize} \quad & px + qy \\ \text{subject to} \quad & x + y \geq 2 \\ & x, y \geq 0. \end{aligned}$$

수식을 조작하여 아래와 같은 과정을 거치면,

$$\begin{aligned} & a(x + y) \geq 2a \\ + \quad & bx \geq 0 \\ - \quad & cy \geq 0 \\ = \quad & (a + b)x + (a + c)y \geq 2a \end{aligned}$$

$$\begin{aligned} \text{Lower bound } B = 2a \\ \text{for } a, b, c \geq 0 \end{aligned}$$

여기서 $a + b = p$와 $a + c = q$라는 조건을 걸면, 수식은 다시 이렇게 정리된다.

$$px + qy \geq 2a$$

좌변은 우리가 구하고자 했던 objective function이다. 결과적으로 inequality를 이용하여 objective function의 lower bound를 구한 셈이 되버렸고, objective(좌변)의 최소화를 찾기 위해서 우리는 이 lower boundary를 최대화하여 $px + qy$와 최대한 가깝게(접하게)해주어야 한다. 즉, 수식으로 정리하면 다음과 같은 문제로 바뀌게 된다.

$$\begin{aligned} \text{maximize} \quad & 2a \\ \text{subject to} \quad & a + b = p \\ & a + c = q \\ & a, b, c \geq 0. \end{aligned}$$

첫번째 식을 primal LP(Linear progamming)이라고 부르고 두번째 식을 (dual linear program)이라고 부른다.

LP 라그랑지안적 접근

앞서 봤던 문제(LP)을 일반화하여 표현하면($x \in R^n$)

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & G\mathbf{x} \leq h \end{aligned}$$

여기서 라그랑주 승수 벡터 u, v를 도입하여 라그랑주 함수를 만들면, 아래와 같이 표현 할 수 있음. 단 $v \geq 0$. (밑에 밑에 수식 관계를 만들어 주기 위한 제약 조건)

$$\mathcal{L}(x, u, v) = c^T x + u^T(Ax - b) + v^T(Gx - h)$$

여기서 두번째 항은 0이고 세번째 항은 항상 0보다 크기 때문에

$$\mathcal{L}(x, u, v) = c^T x + u^T(Ax - b) + v^T(Gx - h) \leq c^T x$$

위의 부등식 관계에 의해서 우리가 찾고 싶은 최적의 값이 f^*라고 하고 primal function의 x의 range를 C라고 표현하면 아래와 같은 부등식을 만족함

$$f^* \geq \min_{x \in C} \mathcal{L}(x, u, v) \geq \min_x \mathcal{L}(x, u, v)$$

제약조건(inequality)가 없는 상황에서 최적값을 더 작게 만들 수 있다는 의미로 해석 할 수 있음. 여기서 가장 오른쪽 항(Lower boundary)을 $g(u, v)$라고 표현하고 **라그랑지 듀얼 함수**라고 부름. $\mathcal{L}$은 우리가 알고 싶은 미지수로 편미분 했을때 결과가 0이 되는 지점에서 최소값을 가짐. 따라서,

$$\begin{aligned} \nabla_x \mathcal{L} = c^T + u^T A + v^T G = 0 \\ \therefore c \ = -u^T A - v^T G \end{aligned}$$

이 수식을 다시 대입해서 최종적으로 수식을 정리하면,

$$\begin{aligned} \mathcal{L}(x, u, v) &= c^T x + u^T(Ax - b) + v^T(Gx - h) \\ &= -u^T b - v^T h \\ &= g(u, v) \end{aligned}$$

따라서 dual problem은 아래와 같이 다시 정의 됌

$$\begin{aligned} \text{minimize} \quad & -u^T b - v^T h \\ \text{subject to} \quad & -u^T A - v^T G = c \\ & v \geq 0 \end{aligned}$$

이러한 관계 말고도 모든 함수에 대해서도 primal & dual problem으로 나누어서 생각 할 수 있음.

일반화

- Objective(primal problem)

$$\begin{aligned} \text{minimize} \quad & f(x) & \\ \text{subject to} \quad & l_j(x) = 0 & i = 1, \ldots, m \\ & h_i(x) \leq 0 & j = 1, \ldots, r \end{aligned}$$

- 라그랑지 적용

$$\begin{aligned} \mathcal{L}(x, u, v) &= f(x) + \sum_i u_i h_i(x) + \sum v_j l_j(x) \\ g(u, v) &= \min_x \mathcal{L}(x, u, v) \end{aligned}$$

- Dual problem

$$\begin{aligned} \text{maximize} \quad & g(u, v) & \\ \text{subject to} \quad & u \geq 0 & \text{where } u \in R^m \end{aligned}$$

Regularization

- Regularization는 overfitting을 방지하기 위해서 모델의 복잡성을 포기하게끔 학습을 진행하는 것.
- 다시 말하면 **학습 단계**에서 bias를 낮추는 것
- 따라서 몇개의 파라미터는 0이 될 수도 있음(dimension drop)

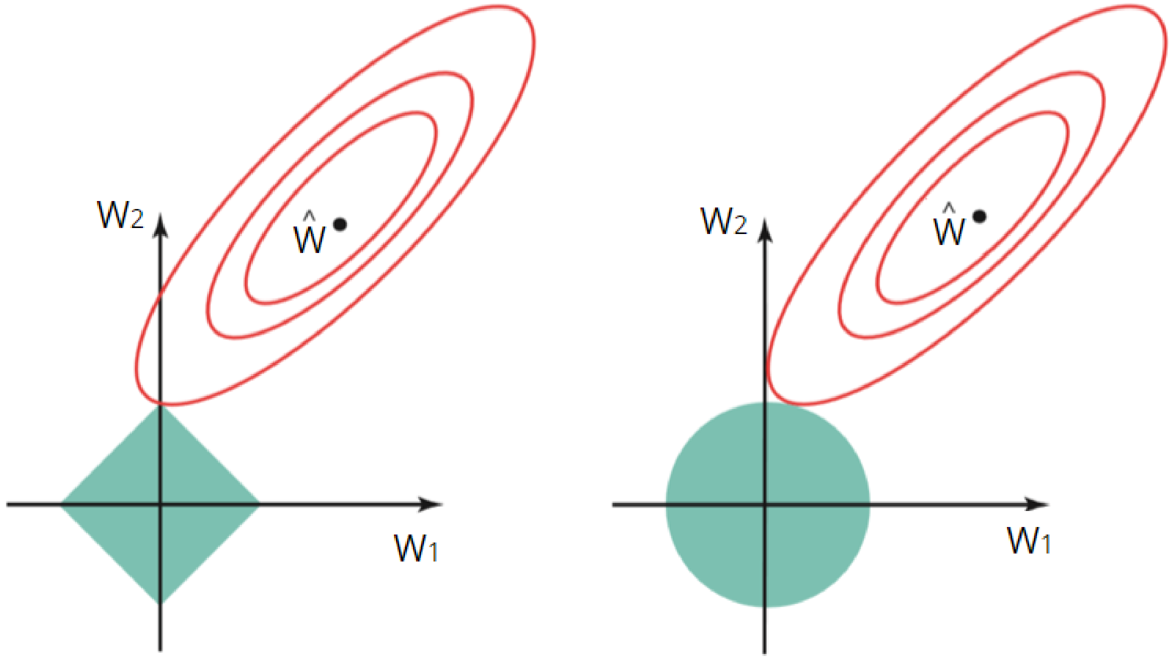
Objective function

$$f = \text{Loss}(\mathbf{x}, \mathbf{w}) + \lambda \, \text{reg}(\mathbf{w})$$

$$\Longleftrightarrow$$

$$\begin{aligned} \text{minimize} \quad & \text{Loss}(\mathbf{x}, \mathbf{w}) \\ \text{subject to} \quad & \text{reg}(\mathbf{w}) \geq c \end{aligned}$$

- 위의 수식은 duality를 갖는 관계로 설명 될 수 있으며, regualarization을 다른 각도에서 해석 할 수 있게 해줌.



- 여기서 λ와 c는 서로 역수의 관계로 λ는 inequality를 만족시키기 위한 slack variable의 허용치로 해석 될 수 있음.(서희님 추가 자료 참고)

References

https://greatjoy.tistory.com/29
https://en.wikipedia.org/wiki/Linear_programming#Simplex_algorithm_of_Dantzig
https://en.wikipedia.org/wiki/Quadratic_programming
https://en.wikipedia.org/wiki/Lagrange_multiplier
https://en.wikipedia.org/wiki/Karush%E2%80%93Kuhn%E2%80%93Tucker_conditions