③ Constrained optimization problem

$$\begin{cases} \min_{\theta} & f_0(\theta) \\ \text{s.t.} & f_i(\theta) \le 0 \quad (\lambda) \\ & h_i(\theta) = 0 \quad (\nu) \end{cases}$$

* Lagrangian: $L(\theta, \lambda, \nu) = f_0(\theta) + \lambda_i f_i(\theta) + \nu_i h_i(\theta)$

$\{\lambda_i, \nu_i\} \equiv \begin{Bmatrix} \text{Lagrange multipliers} \\ \text{Dual variables} \end{Bmatrix}$

$\theta \equiv \begin{Bmatrix} \text{Design variable} \\ \text{primal variables} \end{Bmatrix}$

* Dual function: $g(\lambda, \nu) = \inf_{\theta} L(\theta, \lambda, \nu)$

$$= \inf_{\theta} \left(f_0(\theta) + \lambda_i f_i(\theta) + \nu_i h_i(\theta) \right)$$

* Dual is a lower bound of the optimal value $p^* = f(\theta^*)$

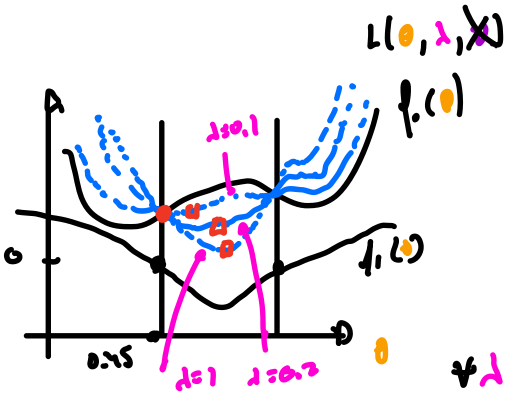
Take $\tilde{\theta}$ feasible ($f_i(\tilde{\theta}) \le 0$; $h_i(\tilde{\theta}) = 0$)

$$L(\tilde{\theta}, \lambda, \nu) = f_0(\tilde{\theta}) + \lambda_i \underbrace{f_i(\tilde{\theta})}_{\le 0} + \nu_i \underbrace{h_i(\tilde{\theta})}_{=0} \overset{\text{if } \lambda \ge 0}{\le} f_0(\tilde{\theta}) \quad \forall \tilde{\theta}$$

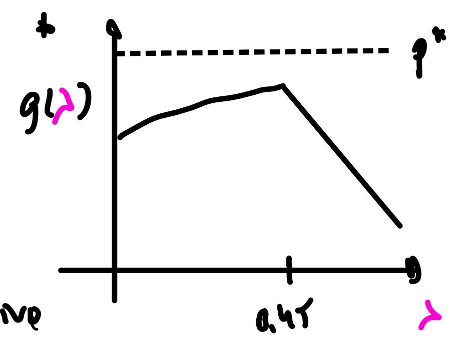
$$g(\lambda, \nu) = \inf_{\theta} L(\theta, \lambda, \nu) \le L(\tilde{\theta}, \lambda, \nu) \overset{\text{if } \lambda \ge 0}{\le} f_0(\tilde{\theta}) \quad \forall \tilde{\theta}$$

Thus for θ^*, which is also feasible $g(\lambda, \nu) \le f_0(\theta^*) = p^*$ if $\lambda \ge 0$

*



*



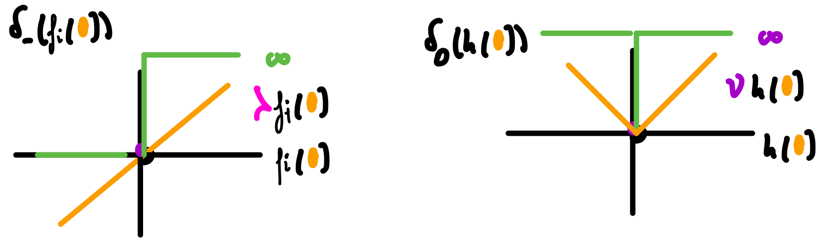
$\forall \lambda$ positive

In regions where $f_1(\theta) \le 0$; $L(\tilde{\theta}, \lambda, \nu) \le f_0(\theta^*)$

* Linear interpolation approximation

$$\min_{\theta} \; f_0(\theta) + \sum \delta_-(f_i(\theta)) + \sum \delta_0(h_i(\theta))$$

where
$$\begin{cases} \delta_-(u) = \begin{cases} 0 & \text{if } u \le 0 \\ \infty & \text{otherwise} \end{cases} \\ \delta_0(u) = \begin{cases} 0 & \text{if } u = 0 \\ \infty & \text{otherwise} \end{cases} \end{cases}$$



$$L(\theta, \lambda, \nu) = f_0(\theta) + \lambda_i f_i(\theta) + \nu_i h_i(\theta) \le f_0(\theta) + \sum \delta(f_i(\theta)) + \sum \delta_0(h_i(\theta))$$

We see that the Lagrangian is a soft approximation of the original problem.

* The Lagrange dual problem
  * What is the best lower bound:

$$\begin{cases} \max_{\lambda, \nu} & g(\lambda, \nu) \\ \text{s.t.} & \lambda \ge 0 \end{cases} \quad \text{(convex optimization)}$$

* Example: least square solution of linear equations

$$\begin{cases} \min_{\theta} & \theta^T \theta \\ & X\theta = y \end{cases} \Rightarrow \begin{cases} g(\nu) = \inf_{\theta} L(\nu, \theta) = \inf_{\theta} \theta^T\theta + \nu^T(X\theta - y) \\ \frac{\partial L}{\partial \theta}(\bar{\theta}) = 2\bar{\theta}\,\theta + X^T\nu\,\bar{\theta} = 0 \Rightarrow \theta = -\frac{1}{2}X^T\nu \\ g(\nu) = -\frac{1}{4}\nu^T X X^T \nu - y^T\nu \end{cases}$$

$$\left\{ \max_{\nu} \; -\frac{1}{4}\nu^T X X^T \nu - y^T \nu \right.$$

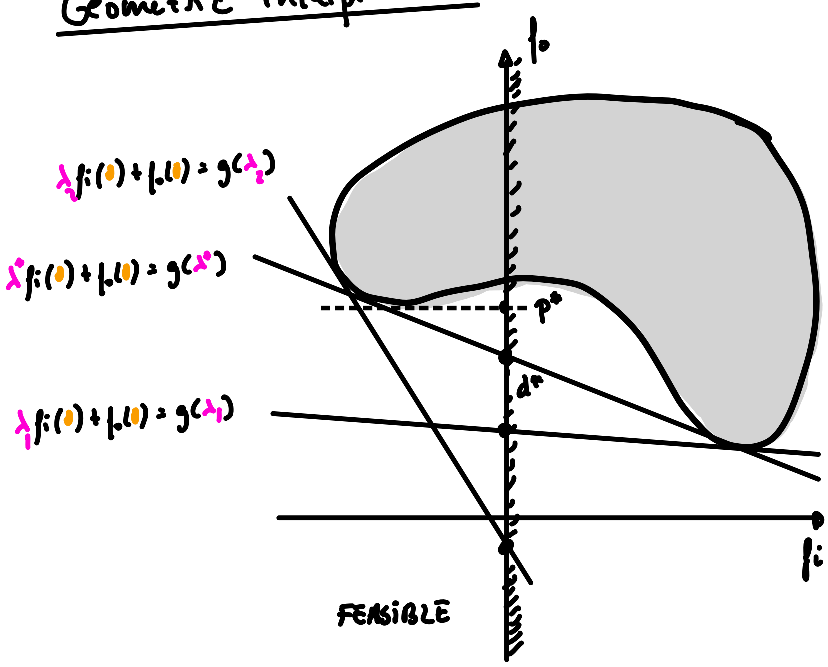
* Example: Linear programming (standard form)

$$\begin{cases}\min & c^T\theta \\ \text{s.t.} & A\theta = b \\ & \underbrace{\theta \geq 0}_{-\theta \leq 0}\end{cases}$$

$$\begin{cases} g(\nu) = \inf_{\theta} L(\nu, \theta) = \inf_{\theta} c^T\theta - \lambda^T\theta + \nu^T(A\theta - b) \\ = \inf_{\theta} -\nu^T b + (c - \lambda + A^T\nu)^T\theta = \\ = -\nu^T b + \inf_{\theta} (c - \lambda + A^T\nu)^T\theta = \begin{cases} -b^T\nu & c - \lambda + A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \end{cases}$$

Equivalently: $\begin{cases} \max_{\lambda, \nu} & -b^T\nu \\ & c - \lambda + A^T\nu = 0 \end{cases} \Rightarrow \begin{cases} \min_{\nu} & b^T\nu \\ & A^T\nu + c \geq 0 \end{cases}$ Dual of linear programming

Geometric interpretation



Saddle-point interpretation

Note that: $\delta_-(f_1(\theta)) = \begin{cases} 0 & \text{if } f_1(\theta) \leq 0 \\ \infty & \text{otherwise} \end{cases} = \left\{ \sup_{\lambda \geq 0} \lambda f_1(\theta) \right\}$

$$p^* = \inf_\theta f_0(\theta) + \delta_-(f_1(\theta)) = \inf_\theta f_0(\theta) + \sup_{\lambda \geq 0} \lambda f_1(\theta) = \inf_\theta \sup_{\lambda \geq 0} f_0(\theta) + \lambda f_1(\theta)$$

$$d^* = \sup_{\lambda \geq 0} \inf_\theta f_0(\theta) + \lambda f_1(\theta)$$

Then: $\sup_{\lambda \geq 0} \inf_\theta f_0(\theta) + \lambda f_1(\theta) = d^* \leq p^* = \inf_\theta \sup_{\lambda \geq 0} f_0(\theta) + \lambda f_1(\theta)$

In addition: $g(\lambda) \leq g(\lambda^*) \leq f_0(\theta^*) \leq f_0(\tilde{\theta})$

$$\underbrace{\min_\theta \mathcal{L}(\theta, \lambda)}_{g(\lambda)} \leq \underbrace{\max_\lambda \min_\theta \mathcal{L}(\theta, \lambda)}_{g(\lambda^*)} = d^* \leq p^* = \underbrace{\mathcal{L}(\theta^*, \lambda)}_{f_0(\theta^*)} \leq \underbrace{\mathcal{L}(\tilde{\theta}, \lambda)}_{f_0(\tilde{\theta})}$$

* Optimality conditions

* Dual feasible provides lower bound $g(\lambda, \nu) \leq p^*$
* During optimization we can check $g(\lambda^*, \nu^*) \leq f_0(\tilde{\theta}^*)$ ← feasible
* Complementary slackness. Given optimal θ^* and (λ^*, ν^*)

$$f_0(\theta^*) = g(\lambda^*, \nu^*) = \inf_\theta f_0(\theta) + \lambda_i^* f_i(\theta) + \nu^* h(\theta)$$

$$\underset{\text{inf}}{\leq} f_0(\theta^*) + \lambda_i^* f_i(\theta^*) + \nu^* h(\theta^*) \underset{\text{feasible}}{\leq} f_0(\theta^*)$$

Then all inequalities become equalities.

1) $\theta^* = \arg\inf_\theta f_0(\theta) + \lambda_i^* f_i(\theta) + \nu^* h(\theta) = \inf_\theta \mathcal{L}(\theta, \lambda^*, \nu^*)$

2) $\lambda_i^* f_i(\theta^*) = 0 \quad \begin{cases} \lambda_i^* > 0 \Rightarrow f_i(\theta^*) = 0 & \text{(active constraint)} \\ f_i(\theta^*) < 0 \Rightarrow \lambda_i^* = 0 & \text{(inactive constraint)} \end{cases}$

* **KKT conditions**

From 1) we have:

$$\left\{ \nabla f_0(\theta^*) + \lambda_i^* \nabla f_i(\theta^*) + \nu_i^* \nabla h_i(\theta^*) = 0 \right.$$

and from 2 and other conditions:

$$\left. \begin{aligned} f_i(\theta^*) &\leq 0 \\ h_i(\theta^*) &= 0 \\ \lambda_i^* &\geq 0 \\ \lambda_i^* f_i(\theta^*) &= 0 \end{aligned} \right\} \text{KKT conditions}$$

* Example: Quadratic function with linear constraints

$$\left\{ \begin{array}{ll} \text{minimize} & \frac{1}{2} \theta^T K \theta - \theta^T f \\ \text{s.t.} & A\theta = b \end{array} \right\} \qquad \underline{\text{KKT}}: \begin{bmatrix} K & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \theta^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} f \\ b \end{bmatrix}$$

* **Perturbation and sensitivity analysis**

$$\left\{ \begin{array}{ll} \min & f_0(\theta) \\ \text{s.t.} & f_i(\theta) \leq u_i \\ & h_i(\theta) = v_i \end{array} \right.$$

Let's define the optimal value of this problem

$$p^*(u, v) = \inf \{ f_0(\theta) \mid f_i(\theta) \leq u_i,\ h_i(\theta) = v_i \}$$

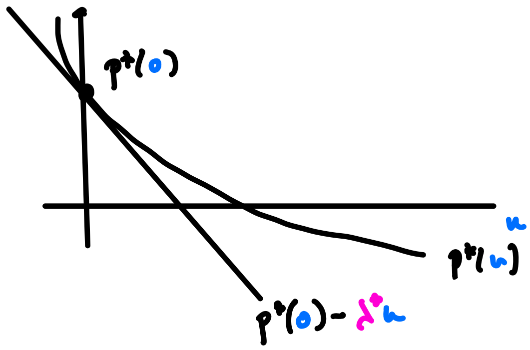
$$p^*(0,0) = p^*$$

For any feasible $\tilde{\theta}$,

$$\left\{ \begin{array}{l} p^*(0,0) = g(\lambda^*, \nu^*) \underset{\text{inf}}{\leq} f_0(\tilde{\theta}) + \lambda_i^* f_i(\tilde{\theta}) + \nu_i^* h(\tilde{\theta}) \underset{\lambda \geq 0}{\leq} f_0(\tilde{\theta}) + \lambda_i^* u_i + \nu_i^* v_i \\ \text{Thus: } \forall \tilde{\theta} \quad f_0(\tilde{\theta}) \geq p^*(0,0) - \lambda_i^* u_i - \nu_i^* v_i \end{array} \right.$$

In particular $\inf f_0(\tilde{\theta}) = p^*(u, v) \geq p^*(0,0) - \lambda_i^* u_i - \nu_i^* v_i$

* Sensitivity interpretations
  * if λ^* is large, then the optimal value $p^*(u, v)$ is highly increased when the constraint is tightened ($u_i < 0$)
    ↳ Also the contrary
  * if ν_i^* is large and positive, then the optimal value $p^*(u, v)$ is highly increased when we take $v_i > 0$
    ↳ Also the contrary



* Local sensitivity analysis

$$\begin{cases} \lambda_i^* = -\dfrac{\partial p^*(0,0)}{\partial u_i} \\[2ex] \nu_i^* = -\dfrac{\partial p^*(0,0)}{\partial v_i} \end{cases}$$

Proof: $p^*(t e_i, 0) \geq p^*(0,0) - \lambda_i^* t$

$$\frac{p^*(t e_i, 0) - p^*(0,0)}{t} \geq \lambda_i^*$$

taking
$$\left.\begin{cases} t > 0 & \lim\limits_{t \to 0} \dfrac{p^*(t e_i, 0) - p^*(0,0)}{|t|} = \dfrac{\partial p^*(0,0)}{\partial u_i} \geq -\lambda_i^* \\[2ex] t < 0 & \lim\limits_{t \to 0} \dfrac{p^*(t e_i, 0) - p^*(0,0)}{|t|} = \dfrac{\partial p^*(0,0)}{\partial u_i} \leq -\lambda_i^* \end{cases}\right\} \Rightarrow \frac{\partial p^*(0,0)}{\partial u_i} = -\lambda_i^*$$

* The value of the lagrange multiplier measures how tigthned a constraint is, or similarly how much the cost will change if the constrain is perturbed