Lec. 16 (11/15/2022)

Example: SDP duality

Primal problem:

$$p^* = \min_{\underline{x} \in \mathbb{R}^n} \underbrace{\langle \underline{c}, \underline{x} \rangle}_{f_0(\underline{x})}$$

$$-F(\underline{x}) \preceq 0 \iff \begin{cases} \text{s.t.} \quad F(\underline{x}) := F_0 + x_1 F_1 + x_2 F_2 + \ldots + x_n F_n \succeq 0 \\ \text{where } F_0, F_1, \ldots, F_n \in \mathbb{S}^m \end{cases}$$

SDP standard form (Lec. 12, p. 1)

Step 1: Lagrangian:

$$L(\underline{x}, \Lambda) = f_0(\underline{x}) + \underbrace{\langle \Lambda, -F(\underline{x}) \rangle}_{\text{Frobenius inner product}}$$

$\Lambda \in \mathbb{S}^m_+$

$$= x_1\left(c_1 - \mathrm{tr}(F_1 \Lambda)\right) + x_2\left(c_2 - \mathrm{tr}(F_2 \Lambda)\right) + \dots + x_n\left(c_n - \mathrm{tr}(F_n \Lambda)\right) - \mathrm{tr}(F_0 \Lambda)$$

Step 2: Lagrange dual function:

$$g(\Lambda) = \inf_{\underline{x} \in \mathbb{R}^n} L(\underline{x}, \Lambda) = \begin{cases} -\mathrm{tr}(F_0 \Lambda) & \text{if } \mathrm{tr}(F_i \Lambda) = c_i \ \forall i = 1, \dots, n \\ -\infty & \text{otherwise} \end{cases}$$

Step 3: Lagrange dual problem:

$$d^* = \boxed{\begin{array}{ll} \max\limits_{\Lambda \in S^m_+} & -\operatorname{tr}(F_0 \Lambda) \\ \text{s.t.} & \operatorname{tr}(F_i \Lambda) = c_i \quad \forall\, i = 1, \dots, n \end{array}}$$

$$\Updownarrow$$

Another SDP

$$\boxed{\begin{array}{ll} \min\limits_{\Lambda \in S^m_+} & \operatorname{tr}(F_0 \Lambda) \\ \text{s.t.} & \operatorname{tr}(F_i \Lambda) = c_i \quad \forall\, i = 1, \dots, n \end{array}}$$

Remark: Strong SDP duality ($p^* = d^*$) holds if $\exists\, \underline{x} \in \mathbb{R}^n$ s.t. the primal is strictly feasible, i.e.,

$$\exists\, \underline{x} \in \mathbb{R}^n \text{ s.t. } F_0 + x_1 F_1 + \dots + x_n F_n \succ 0 .$$

(strictly)

KKT conditions

Karush – $\underbrace{\text{Kuhn – Tucker}}_{(1951)}$ Conditions
(1939)

1st order Necessary Conditions for Optimality in
$\boxed{\text{Constrained}}$ Optimization problems :

Suppose that $\underbrace{f_0}_{\substack{\text{objective}\\ \text{function}}}, \underbrace{f_1, \ldots, f_m}_{\text{LHS of ineq. constraints}}, \underbrace{h_1, h_2, \ldots, h_p}_{\substack{\text{LHS of}\\ \text{equality}\\ \text{constraints}}}$

are C^1 (continuously differentiable) functions.

KKT conditions are a collection of equalities and inequalities relating

$$\underbrace{\underline{x}^*}_{\text{primal optimizer}} \quad \text{and} \quad \underbrace{(\underline{\lambda}^*, \underline{\nu}^*)}_{\text{dual optimizer}}$$

Conditions :

① (Stationarity of the Lagrangian) $\nabla_{\underline{x}} L \Big|_{(\underline{x}^*, \underline{\lambda}^*, \underline{\nu}^*)} = \underline{0}$

$$\Leftrightarrow \nabla_{\underline{x}} f_0(\underline{x}) \Big|_{\underline{x}=\underline{x}^*} + \sum_{i=1}^{m} \lambda_i \nabla_{\underline{x}} f_i(\underline{x}) \Big|_{(\underline{x}^*, \underline{\lambda}^*, \underline{\nu}^*)} + \sum_{j=1}^{p} \nu_j \nabla_{\underline{x}} h_j(\underline{x}) \Big|_{(\underline{x}^*, \underline{\lambda}^*, \underline{\nu}^*)} = \underline{0}$$

② (Complimentary slackness) : $\lambda_i^* f_i(\underline{x}^*) = 0$
[From last lecture] $\quad \forall\, i = 1, \ldots, m$

③ (Primal feasibility) : $f_i(\underline{x}^*) \leq 0 \quad \forall\, i = 1, \ldots, m$

$h_j(\underline{x}^*) = 0 \quad \forall\, j = 1, \ldots, p$

④ (Dual feasibility) : $\underline{\lambda}^* \geq \underline{0}$

(componentwise)

Statement #1: (No convexity assumption)

If
① $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ are C^1, AND
② strong duality holds ($d^* = p^*$)

then the tuple $(\underline{x}^*, \underline{\lambda}^*, \underline{\nu}^*)$ must satisfy the KKT condition.

---

next pg.

Statement # 2 : (Convexity needed)

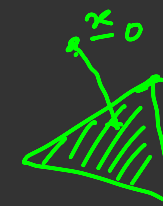
If
1. $f_0, f_1, \dots, f_m, h_1, h_2, \dots, h_p$ are C^1,

and

2. the problem is convex,

then any tuple $(\tilde{\underline{x}}, \tilde{\underline{\lambda}}, \tilde{\underline{\nu}})$ satisfying the KKT conditions, must be primal and dual optimizers with $\boxed{d^* = p^*}$.

Example: Generalized Kullback-Leibler projection of a nonnegative vector onto the standard simplex



Want to compute:

$$\boxed{\underline{x}^{opt} = \text{proj}^{D_{KL}}_{\Delta^{n-1}}(\underline{x}_0)}$$

where $\underline{x}_0 \in \mathbb{R}^n_+ \setminus \{\underline{0}\}$

and

$$\underbrace{D_{KL}(\underline{x} \| \underline{x}_0)}_{\text{generalized Kullback-Leibler divergence}} := \sum_{i=1}^{n} \left(x_i \log \frac{x_i}{x_{0i}} - x_i + x_{0i} \right)$$

Our problem:

$$\underline{x}^{opt} = \operatorname{proj}^{D_{KL}}_{\Delta^{n-1}} (\underline{x}_0)$$

$$= \underset{\underline{x} \in \Delta^{n-1}}{\operatorname{argmin}} \; D_{KL}(\underline{x} \,\|\, \underline{x}_0)$$

$$\underbrace{= \underset{\underline{x} \in \Delta^{n-1}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(x_i \log \frac{x_i}{x_{0i}} - x_i + x_{0i} \right) \right\}}_{\text{convex optimization problem}}$$

We note that $f_0(\underline{x})$ is a convex C^1 function

Now apply KKT condition:

$$L(\underline{x}, \underline{\lambda}, \nu) = \sum_{i=1}^{n} \left(x_i \log \frac{x_i}{x_{0i}} - x_i + x_{0i} \right)$$

$$+ \langle \underline{\lambda}, -\underline{x} \rangle + \nu \left(\mathbb{1}^T \underline{x} - 1 \right),$$

$$\underline{\lambda} \in \mathbb{R}^n_{\geq 0}, \ \nu \in \mathbb{R}$$

Stationarity of Lagrangian: $\nabla_{\underline{x}} L = \underline{0}$

$$\frac{\partial}{\partial x_i} L \Big|_{(\underline{x}^*, \underline{\lambda}^*, \nu^*)} = 0 \quad \forall \ i = 1, \ldots, n$$

$$\Rightarrow \quad x_i^* = \exp(-\nu^*)\, x_{0i} \exp(\lambda_i^*), \quad i = 1, \dots, n$$

primal feasibility: $x_i^* \geq 0$

$$\sum_{i=1}^{n} x_i^* = 1$$

$$\exp(-\nu^*) = \frac{1}{\sum_{i=1}^{n} x_{0i} \exp(\lambda_i^*)}$$

$$\therefore \quad x_i^* = \frac{x_{0i} \exp(\lambda_i^*)}{\sum_{i=1}^{n} x_{0i} \exp(\lambda_i^*)} \quad \forall\, i = 1, \dots, n$$

Dual feasibility: $\lambda_i^* \geq 0 \quad \forall\, i = 1, \dots, n$

Complimentary slackness: $\lambda_i^* x_i^* = 0 \quad \forall\, i = 1, \dots, n$

From the formula: $$x_i^* = \frac{x_{0i} \exp(\lambda_i^*)}{\sum_{i=1}^{n} x_{0i} \exp(\lambda_i^*)},$$

we see that if $x_{0i} = 0$ then $x_i^* = 0$

for some i

else $(x_{0i} > 0)$ then $x_i^* > 0$

But complimentary slackness says: $x_i^* > 0 \Leftrightarrow \lambda_i^* = 0$

and $x_i^* = 0 \Leftrightarrow \lambda_i^* > 0$

$\therefore$ If $x_{0i} \neq 0$ ($\because > 0$), then

$$x_i^* = \frac{x_{0i}\, \cancel{\exp(0)}^{\to 1}}{\sum_{i=1}^{n} x_{0i}\, \cancel{\exp(0)}^{\to 1}}$$

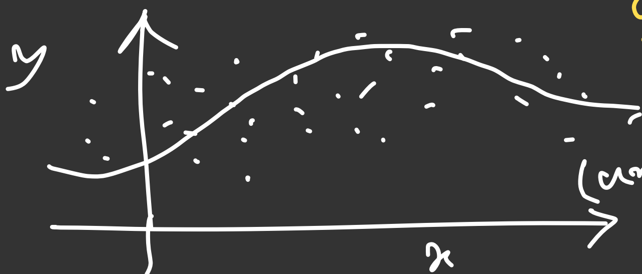
$$= \frac{x_{0i}}{\sum_{i=1}^{n} x_{0i}}$$

Combining $x_{0i} = 0$ & $x_{0i} \neq 0$ (hence > 0) cases:

we conclude: $\boxed{x_i^* = x_{0i} \Big/ \sum_{i=1}^{n} x_{0i}}$ ∎

Application of convex optimization to approximation, estimation, machine learning:

Linear regression:



ordinary least squares (unconstrained QP)

$$\underset{\underline{\theta} \in \mathbb{R}^p}{\arg\min} \; \| A\underline{\theta} - \underline{y} \|_2^2$$

where tall $A \in \mathbb{R}^{m \times n}$, $m > n$

(more equations, less unknowns)

$$\boxed{y = \theta_1 f_1(\underline{x}) + \ldots + \theta_n f_n(\underline{x})}$$

MATLAB A\y

numpy.linalg.lstq

$$\underline{\theta}^{opt} = (A^T A)^{-1} A^T \underline{y}$$

$$= A^{+} \underline{y}$$

Constrained least squares:

No analytical solution

$$\left.\begin{aligned} &\underset{\underline{\theta}\in\mathbb{R}^p}{\arg\min} \ \|A\underline{\theta} - \underline{y}\|_2^2 \\ &\text{s.t.} \ \begin{cases} \underline{0} \preceq \underline{\theta} \preceq \underline{1} \\ \text{or more generally:} \\ \quad G\underline{\theta} \preceq \underline{h} \end{cases} \end{aligned}\right\} \text{constrained } \underline{\text{QP}}$$

Linear Regression with other norms:

e.g.,

$$\left.\begin{aligned} &\underset{\underline{\theta}\in\mathbb{R}^p}{\arg\min} \ \|A\underline{\theta} - \underline{y}\|_\infty \\ &\underset{\underline{\theta}\in\mathbb{R}^p}{\arg\min} \ \|A\underline{\theta} - \underline{y}\|_1 \end{aligned}\right\} \text{LP (see Lec. 12, p. 13,15)}$$