ELE 538B: Mathematics of High-Dimensional Data	Fall 2018

Homework 3 Solutions

Please do not distribute.

1. Gaussian graphical models (*20 points*)

(a) Consider a p-dimensional Gaussian vector $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. For any $1 \leq u, v \leq p$, show that

$$x_u \perp\!\!\!\perp x_v \mid \boldsymbol{x}_{\mathcal{V}\setminus\{u,v\}}$$

(namely, x_u and x_v are conditionally independent given all other variables) if and only if $\Theta_{u,v} = 0$. Here, $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$.

Solution:
Let

$$\boldsymbol{a} = \begin{bmatrix} x_u \\ x_v \end{bmatrix}, \quad \boldsymbol{b} = \boldsymbol{x}_{\mathcal{V}\setminus\{u,v\}}.$$

Then, $\boldsymbol{a}, \boldsymbol{b}$ are jointly Gaussian as

$$\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_{\boldsymbol{a}} & \Sigma_{\boldsymbol{ab}} \\ \Sigma_{\boldsymbol{ab}}^\top & \Sigma_{\boldsymbol{b}} \end{bmatrix} \right),$$

where covariance matrices are

$$\Sigma_{\boldsymbol{a}} = \mathbb{E}\left[\boldsymbol{a}\boldsymbol{a}^\top\right], \quad \Sigma_{\boldsymbol{ab}} = \mathbb{E}\left[\boldsymbol{a}\boldsymbol{b}^\top\right], \quad \Sigma_{\boldsymbol{b}} = \mathbb{E}\left[\boldsymbol{b}\boldsymbol{b}^\top\right].$$

The conditional density of $\boldsymbol{a}$ given $\boldsymbol{b}$ is

$$p_{\boldsymbol{a}|\boldsymbol{b}}(\boldsymbol{r}|\boldsymbol{s}) = \frac{1}{2\pi}(\det \Lambda)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{r} - \boldsymbol{w})^\top \Lambda^{-1} (\boldsymbol{r} - \boldsymbol{w}) \right)$$

where

$$\begin{aligned} \boldsymbol{w} &= \Sigma_{\boldsymbol{ab}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{s}, \\ \Lambda &= \Sigma_{\boldsymbol{a}} - \Sigma_{\boldsymbol{ab}} \Sigma_{\boldsymbol{b}}^{-1} \Sigma_{\boldsymbol{ab}}^\top \end{aligned}$$

In the similar way, we can have the conditional density of x_u and x_v given $\boldsymbol{b}$ respectively as

$$\begin{aligned} p_{x_u|\boldsymbol{b}}(t|\boldsymbol{s}) &= \frac{1}{\sqrt{2\pi\lambda_u}} \exp\left(-\frac{1}{2\lambda_u}(t - w_u)^2 \right), \\ p_{x_v|\boldsymbol{b}}(t|\boldsymbol{s}) &= \frac{1}{\sqrt{2\pi\lambda_v}} \exp\left(-\frac{1}{2\lambda_v}(t - w_v)^2 \right), \end{aligned}$$

where

$$\begin{aligned} \lambda_u &= \Sigma_{x_u} - \Sigma_{x_u\boldsymbol{b}} \Sigma_{\boldsymbol{b}}^{-1} \Sigma_{x_u\boldsymbol{b}}^\top, \\ \lambda_v &= \Sigma_{x_v} - \Sigma_{x_v\boldsymbol{b}} \Sigma_{\boldsymbol{b}}^{-1} \Sigma_{x_v\boldsymbol{b}}^\top, \\ w_u &= \Sigma_{x_u\boldsymbol{b}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{s}, \\ w_v &= \Sigma_{x_v\boldsymbol{b}} \Sigma_{\boldsymbol{b}}^{-1} \boldsymbol{s}. \end{aligned}$$

We have

$$\Lambda_{11} = \lambda_u, \ \Lambda_{22} = \lambda_v, \ \boldsymbol{w}_1 = w_u, \ \boldsymbol{w}_2 = w_v.$$

Assume $\Theta_{u,v} = 0$. *Using Schur complement,*

$$\Theta_{u,v} = \left[\Sigma^{-1}\right]_{u,v} = \left[\left(\Sigma_{\boldsymbol{a}} - \Sigma_{\boldsymbol{ab}}\Sigma_{\boldsymbol{b}}^{-1}\Sigma_{\boldsymbol{ab}}^\top\right)^{-1}\right]_{12} = \left[\Lambda^{-1}\right]_{12} = 0. \tag{1}$$

Therefore, since Λ *is symmetric,*

$$\Lambda = \begin{bmatrix} \lambda_u & 0 \\ 0 & \lambda_v \end{bmatrix}, \tag{2}$$

and

$$p_{\boldsymbol{a}|\boldsymbol{b}}(\boldsymbol{r}|\boldsymbol{s}) = \frac{1}{2\pi}\frac{1}{\sqrt{\lambda_u\lambda_v}}\exp\left(-\frac{1}{2\lambda_u}(r_u - w_u)^2 - \frac{1}{2\lambda_v}(r_v - w_v)^2\right) = p_{x_u|\boldsymbol{b}}(r_u|\boldsymbol{s})p_{x_v|\boldsymbol{b}}(r_v|\boldsymbol{s}),$$

which implies that x_u *and* x_v *are conditionally independent given* $\boldsymbol{b} = \boldsymbol{x}_{\mathcal{V}\setminus\{u,v\}}$.

Now, prove that $\Theta_{u,v} = 0$ *if* x_u *and* x_v *are conditionally independent given* $\boldsymbol{b}$. *For any* $\boldsymbol{r} = (r_u, r_v)^\top$, *we have*

$$(\det\Lambda)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{w})^\top\Lambda^{-1}(\boldsymbol{r}-\boldsymbol{w})\right) = \frac{1}{\sqrt{\lambda_u\lambda_v}}\exp\left(-\frac{1}{2\lambda_u}(r_u - w_u)^2 - \frac{1}{2\lambda_v}(r_v - w_v)^2\right).$$

This implies that

$$\begin{aligned}
&\det\Lambda = \lambda_u\lambda_v, \\
&(\boldsymbol{r}-\boldsymbol{w})^\top\Lambda^{-1}(\boldsymbol{r}-\boldsymbol{w}) = (\boldsymbol{r}-\boldsymbol{w})^\top \begin{bmatrix} \frac{1}{\lambda_u} & 0 \\ 0 & \frac{1}{\lambda_v} \end{bmatrix} (\boldsymbol{r}-\boldsymbol{w}),
\end{aligned}$$

which shows that (2) *and* (1) *are true.*

(b) In graphical lasso, the objective function includes a term $\log\det\boldsymbol{\Theta}$. Show that $g(\boldsymbol{\Theta}) := \log\det(\boldsymbol{\Theta})$ $(\boldsymbol{\Theta} \succ \boldsymbol{0})$ is a concave function.

Hint: A function $g(\boldsymbol{\Theta})$ is concave if $h(t) := g(\boldsymbol{\Theta} + t\boldsymbol{V})$ is concave for all t and $\boldsymbol{V}$ obeying $\boldsymbol{\Theta} + t\boldsymbol{V} \succ \boldsymbol{0}$.

Solution: *We have*

$$\begin{aligned}
h(t) &= \log\det(\Theta + t\boldsymbol{V}) \\
&= \log\det\left(\Theta^{\frac{1}{2}}\left(\boldsymbol{I} + t\Theta^{-\frac{1}{2}}\boldsymbol{V}\Theta^{-\frac{1}{2}}\right)\Theta^{\frac{1}{2}}\right) \\
&= \log\det\left(\boldsymbol{I} + t\Theta^{-\frac{1}{2}}\boldsymbol{V}\Theta^{-\frac{1}{2}}\right) + \log\det\Theta \\
&= \sum_{i=1}^{n}\log(1 + t\lambda_i) + \log\det\Theta,
\end{aligned}$$

where $\lambda_1, \cdots, \lambda_n$ *are the eigenvalues of* $\Theta^{-\frac{1}{2}}\boldsymbol{V}\Theta^{-\frac{1}{2}}$. $\Theta^{\frac{1}{2}}$ *and* $\Theta^{-\frac{1}{2}}$ *are well defined since* $\Theta \succ \boldsymbol{0}$. *It is shown that* $h''(t) \le 0$, *because*

$$h'(t) = \sum_{i=1}^{n}\frac{\lambda_i}{1 + t\lambda_i}, \quad h''(t) = -\sum_{i=1}^{n}\frac{\lambda_i^2}{(1 + t\lambda_i)^2}.$$

Thus, $h(t)$ is concave in t.

2. Restricted isometry properties (*30 points*)

Recall that the restricted isometry constant $\delta_s \geq 0$ of $\boldsymbol{A}$ is the smallest constant such that

$$(1-\delta_s)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leq (1+\delta_s)\|\boldsymbol{x}\|_2^2 \tag{3}$$

holds for all s-sparse vector $\boldsymbol{x} \in \mathbb{R}^p$.

(a) Show that

$$|\langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2\rangle| \leq \delta_{s_1+s_2}\|\boldsymbol{x}_1\|_2\|\boldsymbol{x}_2\|_2$$

for all pairs of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ that are supported on disjoint subsets $S_1, S_2 \subset \{1, \cdots, n\}$ with $|S_1| \leq s_1$ and $|S_2| \leq s_2$.

Solution: *WLOG, assume $\|\boldsymbol{x}_1\| = \|\boldsymbol{x}_2\| = 1$. Since $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ have disjoint support, we get*

$$\begin{aligned}
|\langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2\rangle| &= \frac{1}{4}\left|\|\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\boldsymbol{x}_2\|_2^2 - \|\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{A}\boldsymbol{x}_2\|_2^2\right| \\
&= \frac{1}{4}\left|\left\|\boldsymbol{A}\begin{bmatrix}\boldsymbol{x}_1 \\ \boldsymbol{x}_2\end{bmatrix}\right\|^2 - \left\|\boldsymbol{A}\begin{bmatrix}\boldsymbol{x}_1 \\ -\boldsymbol{x}_2\end{bmatrix}\right\|^2\right| \\
&\leq \frac{1}{4}|2(1+\delta_{s_1+s_2}) - 2(1-\delta_{s_1+s_2})| \\
&\leq \delta_{s_1+s_2}.
\end{aligned}$$

(b) For any $\boldsymbol{u}$ and $\boldsymbol{v}$, show that

$$|\langle \boldsymbol{u},\ (\boldsymbol{I} - \boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{v}\rangle| \leq \delta_s\|\boldsymbol{u}\|_2 \cdot \|\boldsymbol{v}\|_2,$$

where s is the cardinality of support $(\boldsymbol{u}) \cup$ support $(\boldsymbol{v})$.

Solution: *Let $\mathcal{S} =$ support $(\boldsymbol{u}) \cup$ support $(\boldsymbol{v})$.*

$$\begin{aligned}
|\langle \boldsymbol{u}, (\boldsymbol{I} - \boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{v}\rangle| &= |\langle \boldsymbol{u}, \boldsymbol{v}\rangle - \langle \boldsymbol{A}\boldsymbol{u}, \boldsymbol{A}\boldsymbol{v}\rangle| \\
&= |\langle \boldsymbol{u}_{\mathcal{S}}, \boldsymbol{v}_{\mathcal{S}}\rangle - \langle \boldsymbol{A}_{\mathcal{S}}\boldsymbol{u}_{\mathcal{S}}, \boldsymbol{A}_{\mathcal{S}}\boldsymbol{v}_{\mathcal{S}}\rangle| \\
&= |\langle \boldsymbol{u}_{\mathcal{S}}, (\boldsymbol{I} - \boldsymbol{A}_{\mathcal{S}}^\top\boldsymbol{A}_{\mathcal{S}})\boldsymbol{v}_{\mathcal{S}}\rangle| \\
&\leq \|\boldsymbol{u}_{\mathcal{S}}\|_2\|\boldsymbol{I} - \boldsymbol{A}_{\mathcal{S}}^\top\boldsymbol{A}_{\mathcal{S}}\|_{\mathsf{op}}\|\boldsymbol{v}_{\mathcal{S}}\|_2,
\end{aligned}$$

where $\|\cdot\|_{\mathsf{op}}$ denotes the operator norm of a matrix as

$$\|\boldsymbol{A}\|_{\mathsf{op}} = \max_{\|\boldsymbol{x}\|_2=1}\|\boldsymbol{A}\boldsymbol{x}\|_2.$$

By the definition of the restricted isometry constant,

$$|\langle(\boldsymbol{A}_{\mathcal{S}}^\top\boldsymbol{A}_{\mathcal{S}} - \boldsymbol{I})\boldsymbol{x}, \boldsymbol{x}\rangle| = |\langle \boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}, \boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}\rangle - \langle \boldsymbol{x}, \boldsymbol{x}\rangle| = |\|\boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}\|^2 - \|\boldsymbol{x}\|_2^2| \leq \delta_s\|\boldsymbol{x}\|_2^2.$$

Therefore,

$$\|\boldsymbol{I} - \boldsymbol{A}_{\mathcal{S}}^\top\boldsymbol{A}_{\mathcal{S}}\|_{\mathsf{op}} \leq \delta_s$$

and

$$|\langle \boldsymbol{u}, (\boldsymbol{I} - \boldsymbol{A}^\top \boldsymbol{A})\boldsymbol{v}\rangle| \leq \delta_s \|\boldsymbol{u}_{\mathcal{S}}\|_2 \|\boldsymbol{v}_{\mathcal{S}}\|_2 = \delta_s \|\boldsymbol{u}\|_2 \|\boldsymbol{v}\|_2.$$

(c) Suppose that each column of $\boldsymbol{A}$ has unit norm. Show that $\delta_2 = \mu(\boldsymbol{A})$, where $\mu(\boldsymbol{A})$ is the mutual coherence of $\boldsymbol{A}$.

Solution: *Given that*

$$|\langle (\boldsymbol{A}_{\mathcal{S}}^\top \boldsymbol{A}_{\mathcal{S}} - \boldsymbol{I})\boldsymbol{x}, \boldsymbol{x}\rangle| = |\langle \boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}, \boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}\rangle - \langle \boldsymbol{x}, \boldsymbol{x}\rangle| = |\|\boldsymbol{A}_{\mathcal{S}}\boldsymbol{x}\|^2 - \|\boldsymbol{x}\|^2| \leq \delta_s \|\boldsymbol{x}\|^2,$$

δ_s *is the same as*

$$\delta_s = \max_{|\mathcal{S}| \leq s} \|\boldsymbol{A}_{\mathcal{S}}^\top \boldsymbol{A}_{\mathcal{S}} - \boldsymbol{I}\|_{\mathsf{op}}.$$

When $s = 2$,

$$\delta_2 = \max_{i \neq j} \left\| \begin{bmatrix} \boldsymbol{a}_i & \boldsymbol{a}_j \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{a}_i & \boldsymbol{a}_j \end{bmatrix} - \boldsymbol{I} \right\|_{\mathsf{op}}.$$

The eigenvalues of the following matrix

$$\begin{bmatrix} \boldsymbol{a}_i & \boldsymbol{a}_j \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{a}_i & \boldsymbol{a}_j \end{bmatrix} - \boldsymbol{I} = \begin{bmatrix} 1 & \langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle \\ \langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle & 1 \end{bmatrix} - \boldsymbol{I} = \begin{bmatrix} 0 & \langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle \\ \langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle & 0 \end{bmatrix}$$

are $\pm\langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle$, *and accordingly*

$$\delta_2 = \max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle| = \mu(\boldsymbol{A}).$$

3. Statistical dimension (*10 points*) Recall that for any convex cone $\mathcal{K}$, its statistical dimension and Gaussian width are defined respectively as

$$\text{stat-dim}(\mathcal{K}) := \mathbb{E}\big[\|\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\|_2^2\big]$$

and

$$w(\mathcal{K}) := \mathbb{E}\left[\sup_{\boldsymbol{z} \in \mathcal{K}, \|\boldsymbol{z}\|_2 = 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle\right],$$

where $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\mathcal{P}_{\mathcal{K}}$ denotes the projection to $\mathcal{K}$ as

$$\mathcal{P}_{\mathcal{K}}(\boldsymbol{g}) = \arg\min_{\boldsymbol{z} \in \mathcal{K}} \|\boldsymbol{g} - \boldsymbol{z}\|_2.$$

(a) Prove that $w^2(\mathcal{K}) \leq \text{stat-dim}(\mathcal{K})$.

Solution:

$$\begin{aligned} w^2(\mathcal{K}) &= \left(\mathbb{E}\left[\sup_{\boldsymbol{z} \in \mathcal{K}, \|\boldsymbol{z}\|_2 = 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle\right]\right)^2 && (4) \\ &\leq \left(\mathbb{E}\left[\sup_{\boldsymbol{z} \in \mathcal{K}, \|\boldsymbol{z}\|_2 \leq 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle\right]\right)^2 && (5) \\ &\leq \mathbb{E}\left[\left(\sup_{\boldsymbol{z} \in \mathcal{K}, \|\boldsymbol{z}\|_2 \leq 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle\right)^2\right], && (6) \end{aligned}$$

where (5) *holds because* $\{\boldsymbol{z} \ : \ \boldsymbol{z} \in \mathcal{K}, \|\boldsymbol{z}\|_2 = 1\} \subset \{\boldsymbol{z} \ : \ \boldsymbol{z} \in \mathcal{K}, \|\boldsymbol{z}\|_2 \leq 1\}$, *and* (6) *holds by Jensen's inequality. Further, the statistical dimension can be represented as*

$$\textit{stat-dim}(\mathcal{K}) = \mathbb{E}\left[\left(\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2\leq 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle\right)^2\right],$$

since

$$\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2\leq 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle = \sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2\leq 1} \langle \boldsymbol{z}, \mathcal{P}_{\mathcal{K}}(\boldsymbol{g}) + \mathcal{P}_{\mathcal{K}^\circ}(\boldsymbol{g})\rangle \leq \sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2\leq 1} \langle \boldsymbol{z}, \mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\rangle = \left\langle \frac{\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})}{\|\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\|_2}, \mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\right\rangle = \|\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\|_2,$$

where

$$\mathcal{K}^\circ = \{\boldsymbol{u} \ : \ \langle \boldsymbol{u}, \boldsymbol{x}\rangle \leq 0 \quad \forall \boldsymbol{x} \in \mathcal{K}\}, \quad \boldsymbol{g} = \mathcal{P}_{\mathcal{K}}(\boldsymbol{g}) + \mathcal{P}_{\mathcal{K}^\circ}(\boldsymbol{g}), \quad \langle \mathcal{P}_{\mathcal{K}}(\boldsymbol{g}), \mathcal{P}_{\mathcal{K}^\circ}(\boldsymbol{g})\rangle = 0,$$

and

$$\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|\leq 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle \geq \left\langle \frac{\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})}{\|\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\|_2}, \boldsymbol{g}\right\rangle = \|\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\|_2.$$

Therefore, $w^2(\mathcal{K}) \leq$ *stat-dim*$(\mathcal{K})$.

(b) (Optional (10 bonus points)) Prove the reverse inequality stat-dim$(\mathcal{K}) \leq w^2(\mathcal{K}) + 1$.
hint: Let $f(\cdot)$ be a function that is Lipschitz with respect to the Euclidean norm:

$$|f(\boldsymbol{u}) - f(\boldsymbol{v})| \leq M\|\boldsymbol{u} - \boldsymbol{v}\|_2 \qquad \forall \boldsymbol{u}, \boldsymbol{v}.$$

Then, $\mathrm{Var}(f(\boldsymbol{g})) \leq M^2$.

Solution: *It was shown before that*

$$\textit{stat-dim}(\mathcal{K}) = \mathbb{E}\left[\left(\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2\leq 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle\right)^2\right],$$

and it is enough to show that

$$\mathbb{E}\left[\left(\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2\leq 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle\right)^2\right] \leq w^2(\mathcal{K}) + 1.$$

For any $\boldsymbol{g} \notin \mathcal{K}^\circ$, *we have*

$$\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2\leq 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle = \sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2 = 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle.$$

Also, if $\boldsymbol{g} \in \mathcal{K}^\circ$, *then*

$$\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2\leq 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle = \langle \boldsymbol{0}, \boldsymbol{g}\rangle = 0,$$

because $\langle \boldsymbol{z}, \boldsymbol{g}\rangle \leq 0$ *for all* $\boldsymbol{z} \neq 0, \boldsymbol{z} \in \mathcal{K}$. *Therefore,*

$$\left(\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2\leq 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle\right)^2 = \left(\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2 = 1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle\right)^2 \mathbf{1}\left\{\boldsymbol{g} \notin \mathcal{K}^\circ\right\},$$

$$\mathbb{E}\left[\left(\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2\leq 1}\langle\boldsymbol{z},\boldsymbol{g}\rangle\right)^2\right]\leq\mathbb{E}\left[f^2(\boldsymbol{g})\right],$$

where $f(\cdot)$ defined as

$$f(\boldsymbol{g})=\sup_{\boldsymbol{z}\in\mathcal{K},\|\boldsymbol{z}\|_2=1}\langle\boldsymbol{z},\boldsymbol{g}\rangle.$$

This function $f(\cdot)$ is 1-Lipschitz function because

$$\begin{aligned}&|f(\boldsymbol{u})|\leq\|\boldsymbol{u}\|_2 \qquad \forall\boldsymbol{u},\\ &|f(\boldsymbol{u})-f(\boldsymbol{v})|\leq|\|\boldsymbol{u}\|_2-\|\boldsymbol{v}\|_2|\leq\|\boldsymbol{u}-\boldsymbol{v}\|_2 \qquad \forall\boldsymbol{u},\boldsymbol{v}.\end{aligned}$$

Using the hint,

$$\mathbb{E}\left[f^2(\boldsymbol{g})\right]-w^2(\mathcal{K})=\mathit{Var}(f(\boldsymbol{g}))\leq 1.$$

Thus, we have proven stat-dim$(\mathcal{K})\leq w^2(\mathcal{K})+1$.