Recall our previous development:

$$(\text{DR-S}) \iff \sup_{\mathbb{P}\succeq 0} \inf_{\substack{r\in\mathbb{R} \\ Q\succeq 0 \\ \begin{bmatrix} P & p \\ p^T & s\end{bmatrix}\succeq 0}} \mathcal{L}(\mathbb{P}, r, Q, P, p, s)$$

where the Lagrangian is given by

$$\mathcal{L}(\mathbb{P}, r, Q, P, p, s) = \int_C h(x,\xi)\, d\mathbb{P}(\xi) + r\left(1 - \int_C d\mathbb{P}(\xi)\right)$$
$$- \left\langle Q, \int_C (\xi\xi^T - \xi\mu_0^T - \mu_0\xi^T)\, d\mathbb{P}(\xi) - \gamma_2\Gamma_0 + \mu_0\mu_0^T \right\rangle$$
$$+ \langle P, \Gamma_0\rangle + 2p^T\left(\int_C \xi\, d\mathbb{P}(\xi) - \mu_0\right) + s\gamma_1$$

The dual takes the form

$$\inf_{\substack{r\in\mathbb{R} \\ Q\succeq 0 \\ \begin{bmatrix} P & p \\ p^T & s\end{bmatrix}\succeq 0}} \sup_{\mathbb{P}\succeq 0} \mathcal{L}(\mathbb{P}, r, Q, P, p, s)$$

$$= \inf \quad r + \gamma_2\langle \Gamma_0, Q\rangle + \mu_0^T Q\mu_0 - 2\mu_0^T q + \sqrt{\gamma_1}\, \|\Gamma_0^{1/2}(q - Q\mu_0)\|_2$$

$$\text{s.t.} \quad h(x,\xi) - r - \xi^T Q\xi + 2\xi^T q \le 0 \quad \forall \xi \in C$$

$$r\in\mathbb{R}, \quad Q\succeq 0, \quad q\in\mathbb{R}^n.$$

* Why strong duality holds?

The original problem is

$$\sup \int_C h(x,\xi)\, d\mathbb{P}(\xi)$$

$$\text{s.t.} \quad \int_C d\mathbb{P}(\xi) = 1,$$

$$\int_C (\xi\xi^T - \xi\mu_0^T - \mu_0\xi^T)\, d\mathbb{P}(\xi) \preceq \gamma_2\Gamma_0 - \mu_0\mu_0^T,$$

$$\begin{bmatrix} \Gamma_0 & \int_C \xi\, d\mathbb{P}(\xi) - \mu_0 \\ \int_C \xi^T d\mathbb{P}(\xi) - \mu_0^T & \gamma_1 \end{bmatrix} \succeq 0,$$

$$\mathbb{P}\succeq 0,$$

which can be written in the form

$$
\begin{aligned}
\sup \quad & \langle h(x,\cdot), \mathbb{P} \rangle \\
\text{s.t.} \quad & \mathcal{A}(\mathbb{P}) - \underbrace{\begin{bmatrix} 1 \\ -\gamma_2 \Gamma_0 + \mu_0 \mu_0^T \\ \begin{bmatrix} -\Gamma_0 & \mu_0 \\ \mu_0^T & -\gamma_1 \end{bmatrix} \end{bmatrix}}_{\triangleq\, b} \in \underbrace{\{0\} \times \mathcal{S}^n_+ \times \mathcal{S}^{n+1}_+}_{\triangleq\, K}, \\
& \mathbb{P} \in \mathcal{C},
\end{aligned}
$$

where $\mathcal{A}$ is the linear map

$$
\mathbb{P} \mapsto \left(\int_C d\mathbb{P}(\xi),\ \int_C (-\xi\xi^T + \xi\mu_0^T + \mu_0 \xi^T)\, d\mathbb{P}(\xi),\ \begin{bmatrix} 0 & \int_C \xi\, d\mathbb{P}(\xi) \\ \int_C \xi^T d\mathbb{P}(\xi) & 0 \end{bmatrix} \right)
$$

and $\mathcal{C} = \text{cone}(\mathcal{M})$ with $\mathcal{M} = \{ \mathbb{P} : \mathbb{P} \geq 0,\ \mathbb{P}(\mathbb{R}^n) = 1 \}$

Now, since $\mu_0 \in C$, note that

$$
\mathcal{A}(\mathcal{C}) - K \underset{\substack{\uparrow \\ \text{Take} \\ \mathbb{P}(\xi) = \theta \delta_{\mu_0}(\xi)}}{\supseteq} \left\{ \left(\theta,\ \theta \mu_0 \mu_0^T - \mathcal{S}^n_+,\ \theta \begin{bmatrix} 0 & \mu_0 \\ \mu_0^T & 0 \end{bmatrix} - \mathcal{S}^{n+1}_+ \right) : \theta > 0 \right\}
$$

and hence $b \in \text{int}[\mathcal{A}(\mathcal{C}) - K]$. Thus, by Proposition 3.4 of Shapiro: On Duality Theory of Conic Linear Problems, we conclude that strong duality holds.

* In summary, we can reformulate our original problem as

$$
\text{(DR-M)} \qquad \inf_{x \in \mathcal{U}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[h(x,\xi)]
$$

$$
\begin{aligned}
= \inf \quad & r + t \\
\text{s.t.} \quad & h(x,\xi) - r - \xi^T Q \xi + 2\xi^T q \leq 0 \quad \forall \xi \in C \qquad \text{— (!!)} \\
& t \geq \gamma_2 \langle \Gamma_0, Q \rangle + \langle \mu_0 \mu_0^T, Q \rangle - 2\mu_0^T q + \sqrt{\gamma_1} \| \Gamma_0^{1/2} (q - Q\mu_0) \|_2 \\
& r \in \mathbb{R},\ Q \succeq 0,\ q \in \mathbb{R}^n,\ x \in \mathcal{U}.
\end{aligned}
$$

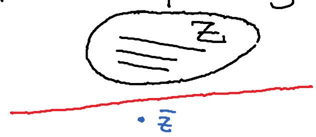
Is this a tractable formulation? Note that (!!) still has infinitely many constraints!

Consider the following abstract problem:

$$(P) \qquad \inf_{z \in Z} \langle c, z \rangle$$

where Z is a full-dimensional compact convex set. Suppose that (P) has an optimal solution.

Theorem: Suppose that we have an oracle that can perform the following in polynomial time for any given $\bar{z}$:



(1) confirm $\bar{z} \in Z$, or

(2) confirm $\bar{z} \notin Z$ by returning a hyperplane that separates $\bar{z}$ from Z.

Then, (P) can be solved in polynomial time by the ellipsoid method.

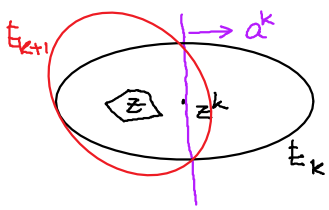
Idea: (1) Start with an ellipsoid E_0 containing Z; $k \leftarrow 0$

(2) while $z^k \notin Z$, where z^k is center of E_k

get from oracle a^k s.t $\{ z : (a^k)^T z \le (a^k)^T z^k \} \supseteq Z$

set E_{k+1} be the minimum-volume ellipsoid containing $E_k \cap \{ z : (a^k)^T z \le (a^k)^T z^k \}$

$k \leftarrow k+1$



Key: The volumes of $\{E_k\}$ decrease geometrically.

Reference:

Grötschel, Lovász, Schrijver: Geometric Algorithms and Combinatorial Optimization, Springer, 1993.

Returning to our problem, we need to show how the oracle can be implemented. Given $\bar{x} \in \mathcal{U}$, $\bar{Q} \succeq 0$, $\bar{q} \in \mathbb{R}^n$, $\bar{r} \in \mathbb{R}$, observe that whether

$$h(\bar{x}, \xi) - \bar{r} - \xi^T \bar{Q} \xi + 2 \xi^T \bar{q} \le 0 \quad \forall \xi \in C$$

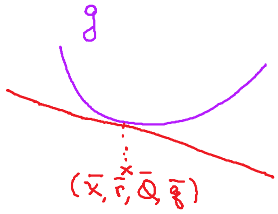
is feasible can be determined by solving the following:

$$(\Delta) \qquad v^* = \max \; h(\bar{x}, \xi) - \bar{r} - \xi^T \bar{Q} \xi + 2 \xi^T \bar{q} \quad \text{s.t.} \quad \xi \in C$$

Assume that $\xi \mapsto h(x,\xi)$ is concave for any x, (Δ) is a convex optimization problem. Thus, if C is compact convex and equipped with a polynomial-time oracle, and a supergradient of $\xi \mapsto h(x,\xi)$ can be found in polynomial time, then (Δ) is polynomial-time solvable.

Now, if $v^* > 0$, then there exists $\bar{\xi} \in C$ such that

$$\underbrace{h(\bar{x},\bar{\xi}) - \bar{r} - \bar{\xi}^T \bar{Q} \bar{\xi} + 2\bar{\xi}^T \bar{q}}_{g(\bar{x},\bar{r},\bar{Q},\bar{q})} > 0 .$$



Assuming $x \mapsto h(x,\xi)$ is convex for any ξ, g is convex in (x,r,Q,q). Hence,

$$g(x,r,Q,q) \geq g(\bar{x},\bar{r},\bar{Q},\bar{q}) + \left\langle (s,-1,-\bar{\xi}\bar{\xi}^T, 2\bar{\xi}), (x-\bar{x}, r-\bar{r}, Q-\bar{Q}, q-\bar{q}) \right\rangle$$

Since for any (x,r,Q,q) that is feasible for (!!), we have $g(x,r,Q,q) \leq 0$. It follows that

$$0 \geq g(\bar{x},\bar{r},\bar{Q},\bar{q}) + \left\langle (s,-1,-\bar{\xi}\bar{\xi}^T, 2\bar{\xi}), (x-\bar{x}, r-\bar{r}, Q-\bar{Q}, q-\bar{q}) \right\rangle$$

is a hyperplane separating $(\bar{x},\bar{r},\bar{Q},\bar{q})$ and the convex set

$$\left\{ (x,r,Q,q) : h(x,\xi) - r - \xi^T Q \xi + 2\xi^T q \leq 0 \quad \forall \xi \in C \right\}$$

$$= \bigcap_{\xi \in C} \left\{ (x,r,Q,q) : h(x,\xi) - r - \xi^T Q \xi + 2\xi^T q \leq 0 \right\}.$$

Finally, let us revisit the uncertainty set used:

$$\mathcal{P} \triangleq \mathcal{P}(C, \mu_0, \Gamma_0, \gamma_1, \gamma_2) = \left\{ \mathbb{P} \geq 0 : \begin{array}{l} \mathbb{P}(C) = 1, \\ (\mathbb{E}_{\mathbb{P}}[\xi] - \mu_0)^T \Gamma_0^{-1} (\mathbb{E}_{\mathbb{P}}[\xi] - \mu_0) \leq \gamma_1, \\ \mathbb{E}_{\mathbb{P}}[(\xi - \mu_0)(\xi - \mu_0)^T] \preceq \gamma_2 \Gamma_0 \end{array} \right\}.$$

* How to we choose γ_1, γ_2 so that the true distribution $\mathbb{P}^*$ belongs to $\mathcal{P}$?

<u>Idea</u>: Make use of concentration inequalities.

Setup: Let $\xi_1, \cdots, \xi_N$ be iid according to $\mathbb{P}^*$. Set

Setup: Let $\xi_1, \cdots, \xi_N$ be iid according to $\mathbb{P}^*$. Set

$$\mu_0 = \frac{1}{N}\sum_{i=1}^{N}\xi_i \quad , \quad \Gamma_0 = \frac{1}{N}\sum_{i=1}^{N}(\xi_i-\mu_0)(\xi_i-\mu_0)^T$$

Let μ and Γ be the mean and covariance of $\mathbb{P}^*$, respectively.

Assume that $\mathbb{P}^*$ has bounded support:

$$\mathbb{P}^*\left((\xi-\mu)^T\Gamma^{-1}(\xi-\mu) \le R^2\right) = 1 \quad \text{for some } R\in(0,+\infty)$$

Then, we compute, for any given $t>0$,

$$\mathbb{P}^*\left[(\mu_0-\underbrace{\mu}_{\mathbb{E}_{\mathbb{P}^*}[\xi]})^T\Gamma^{-1}(\mu_0-\underbrace{\mu}_{\mathbb{E}_{\mathbb{P}^*}[\xi]}) \le t\right] = \mathbb{P}^*\left[\left\| \Gamma^{-1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i-\mu\right)\right\|_2^2 \le t\right]$$

$$= \mathbb{P}^*\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\Gamma^{-1/2}(\xi_i-\mu)\right\|_2^2 \le t\right] = \mathbb{P}^*\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\zeta_i\right\|_2^2 \le t\right].$$

Note that $\zeta = \Gamma^{-1/2}(\xi-\mu)$ has mean 0 and covariance I.

Choosing $t = \frac{R^2}{N}\left(2+\sqrt{2\ln(1/\delta)}\right)^2$ and invoking a suitable concentration inequality (McDiarmid), we get

$$\mathbb{P}^*\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\zeta_i\right\|_2^2 \le \frac{R^2}{N}\left(2+\sqrt{2\ln(1/\delta)}\right)^2\right] \ge 1-\delta. \qquad \text{—— } (+)$$

Theorem (McDiarmid)

Let $\{\xi_i\}_{i=1}^N$ be independent random vectors and g be a real-valued function. Suppose that

$$|g(\xi_1,\cdots,\xi_j,\cdots,\xi_N) - g(\xi_1,\cdots,\xi_j',\cdots,\xi_N)| \le c_j$$

for all j. Then, for any $t\ge 0$,

$$\Pr\left[g(\xi_1,\cdots,\xi_N) - \mathbb{E}[g(\xi_1,\cdots,\xi_N)] \le -t\right] \le \exp\left(-\frac{2t^2}{\sum_{j=1}^{N}c_j^2}\right).$$

Exercise: Deduce (+) from McDiarmid's theorem

* The above shows that

$$\mathbb{P}^*\left[(\mu_0-\mu)^T\Gamma^{-1}(\mu_0-\mu) \le \frac{R^2}{N}\left(2+\sqrt{2\ln(1/\delta)}\right)^2\right] \ge 1-\delta.$$

This almost suggests that by taking $\gamma_1 = \frac{R^2}{N}\left(2+\sqrt{2\ln(1/\delta)}\right)^2$ and choosing $\delta \in (0,1)$ appropriately, we have $\mathbb{P}^* \in \mathcal{P}$. However, the definition of $\mathcal{P}$ uses

$$(\mathbb{E}_{\mathbb{P}}[\xi] - \mu_0)^T \Gamma_0^{-1} (\mathbb{E}_{\mathbb{P}}[\xi] - \mu_0) \le \gamma_1$$

Hence, we still need to bound the difference between Γ^{-1} and Γ_0^{-1}

<u>Theorem</u>: Let $\{\xi_i\}_{i=1}^N$ be as before. Then, with probability $\geq 1-\delta$,

$$\frac{1}{1+\alpha} \Gamma_0 \preceq \Gamma \preceq \frac{1}{1-\alpha-\beta} \Gamma_0 ,$$

where

$$\alpha = \frac{R^2}{\sqrt{N}} \left(\sqrt{1 - \frac{n}{R^4}} + \sqrt{\ln 4/\delta} \right)$$

$$\beta = \frac{R^2}{M} \left(2 + \sqrt{2 \ln 2/\delta} \right)^2 ,$$

n = dimension of ξ (i.e., $\xi \in \mathbb{R}^n$),

$$N > R^4 \left(\sqrt{1 - \frac{n}{R^4}} + \sqrt{\ln 2/\delta} \right)^2 .$$

Putting the above pieces together, we obtain a choice for γ_1.

* Using a similar analysis, we can get a choice for γ_2.

The material in this section is taken from

Delage, Ye: Distributionally Robust Optimization under Moment Uncertainty with Application to Data-Driven Problems. Oper. Res. 58(3): 595-612, 2008.

Extension of the above probabilistic results to distributions with possibly unbounded support can be found in

So: Moment Inequalities for Sums of Random Matrices and Their Applications in Optimization. Math. Prog. 130: 125-151, 2011.