DFS Robust Optimization Problem Formulation

The goal of daily fantasy sports is to draft a proper lineup $(x \in \{0,1\}^N)$ for which each player has a cost $(c \in \mathbb{R}_+^N)$, subject to a budget constraint, that maximizes the number of points you will receive. Given that you have projections for what each player will score $(p \in \mathbb{R}^N)$, this problem can be formulated as a simple mixed integer linear programming problem (MILP):

$$
\begin{aligned}
&\max_x p^T x \\
s.t \quad &\sum_{i \in \{QB\}} x_i = 1 \\
&\sum_{i \in \{WR\}} x_i \geq 3 \\
&\sum_{i \in \{RB\}} x_i \geq 2 \\
&\sum_{i \in \{TE\}} x_i \geq 1 \\
&\sum_{i \in \{DST\}} x_i = 1 \\
&\sum_{i} x_i = 9 \\
&c^T x \leq Budget \\
&x \in \{0,1\}^n, p \in \mathbb{R}^n, c \in \mathbb{R}_+^n
\end{aligned}
$$

The lineup, budget, and variable constraints remain untouched throughout the problem formulation so they will be dropped until the very end.

The goal then becomes to maximize the worst-case scenario of *actual* points scored. The objective function is altered to include the random error vector:

$$\max_x (p + e)^T x$$

Now "worst-case" scenario needs to be captured. Although it may seem direct to include an inner minimization over the error vector, shifting the objective to the constraints and making the constraint a robust constraint makes the problem simpler.

$$\max_{x,\theta} \theta$$

$$s.t \qquad \theta \leq \ (p+e)^T x \qquad \forall e \in \mathcal{U}$$

This is a standard method for shifting the objective function to the constraints. A new "dummy" variable is added, then the dummy variable is upper bounded by the old objective function. Thus, if the goal is to maximize the dummy variable, the dummy variable will always be the maximum of the objective function. The problem is unchanged. Now the constraint is made "robust" to the uncertainty in error by including the stipulation that the constraint must hold *for all* errors in the uncertainty set. Because the constraint is an upper bound, if the dummy variable meets the constraint in the worst-case error scenario, it will meet the constraint for any error scenario. This is how the "worst-case" scenario is maximized. The robust constraint can then be rewritten below:

$$\theta \leq \ (p+e)^T x \quad \forall e \in \mathcal{U}$$

$$\theta \leq \min_{e \in \mathcal{U}} (p+e)^T x$$

$$\theta \leq p^T x + \min_{e \in \mathcal{U}} e^T x$$

The focus is now shifted to solving the minimization problem in the constraint. The definition of the uncertainty set in the second line below is derived in the "data whitening" proof. This minimization problem, for which will now be called the primal problem, is a conic programming problem.

$$\min_{e} x^T e$$

$$s.t \qquad (e,\rho) \in \mathcal{K} = \left\{ (e,\rho) \mid \left\| \Sigma_e^{-1/2} e \right\|_p \leq \rho \right\}$$

The conic constraint is removed and Lagrangian variables are added. A new constraint called the dual-cone is also added.

$$\mathcal{L}(e,s,t) = x^T e - \mathrm{s}^{\mathrm{T}} \Sigma_e^{-1/2} e - t\rho$$

$$s.t \qquad (s,t) \in \mathcal{K}^* = \{ (s,t) \mid \langle (s,t),(e,\rho) \rangle \geq 0 \ \forall (e,\rho) \in \mathcal{K} \}$$

The Lagrangian is first minimized over the errors then maximized over the dual variables. If strong duality holds (which it does here), the solution to the final maximization problem, which will now be called the dual problem, will be the same as the primal problem. Because the dual problem is a lower bound to the primal problem and the case of interest is when they are equal, only the case where a finite solution is obtained is considered. The dual problem is found by the following steps with proof of the solution to the dual cone.

$$g(s,t) = \min_{e} \mathcal{L}(e,s,t) = \min_{e} \left[\left(x^T - \mathrm{s}^{\mathrm{T}} \Sigma_e^{-1/2} \right) e - t\rho \right] = \begin{cases} -t\rho & x = \Sigma_e^{-1/2} s \\ -\infty & otw \end{cases}$$

Because only the finite case is considered, the problem becomes:

$$\max_{s,t} g(s,t) = \max_{s,t} -t\rho$$
$$s.t. \quad x = \Sigma_e^{-1/2} s$$
$$(s,t) \in \mathcal{K}^* = \left\{(s,t) \mid \|s\|_q \le t\right\} \, where \;\; \frac{1}{p} + \frac{1}{q} = 1$$

Proof of the dual cone

$$\mathcal{K}^* = \{(s,t) \mid \langle (s,t), (e,\rho) \rangle \ge 0 \quad \forall (e,\rho) \in \mathcal{K}\}$$
$$= \{(s,t) \mid \langle (s,t), (-e,\rho) \rangle \ge 0 \quad \forall (-e,\rho) \in \mathcal{K}\}$$
$$= \{(s,t) \mid -s^T e + t\rho \ge 0 \quad \forall (-e,\rho) \in \mathcal{K}\}$$
$$= \left\{(s,t) \mid s^T e \le t\rho \quad \forall \|-e\|_p \le \rho\right\}$$
$$= \left\{(s,t) \mid \frac{1}{\rho} \max_{\|e\|_p \le \rho} s^T e \le t\right\}$$
$$= \left\{(s,t) \mid \|s\|_q \le t\right\}$$

To maximize the objective function, the smallest value for t must be selected. t is lower bounded by $\|s\|_q$ so t can be replaced by $\|s\|_q$. Then s is constrained by equality so it can be replaced with $\Sigma_e^{1/2} x$. The dual problem simplifies to:

$$\max_{s,t} -t\rho = \max_{s,t} -\rho\|s\|_q$$
$$s.t \quad s = \Sigma_e^{1/2} x$$
$$= \max_{s,t} -\rho\left\|\Sigma_e^{1/2} x\right\|_q = -\rho\left\|\Sigma_e^{1/2} x\right\|_q$$

The solution to the dual problem can now replace the primal problem in the initial robust constraint. The dummy variable can be removed, and the right-hand side of the constraint can be shifted back to the objective function. The objective function looks like MILP but now we have the extra term subtracted. This can be thought of as the "safety factor" from the effect of the errors. The robust problem has been formulated.

$$\theta \le p^T x - \rho\left\|\Sigma_e^{1/2} x\right\|_q$$
$$\max_x p^T x - \rho\left\|\Sigma_e^{1/2} x\right\|_q$$