Robust Optimization with Applications

Prof. Daniel P. Palomar

ELEC5470/IEDA6100A - Convex Optimization
The Hong Kong University of Science and Technology (HKUST)
Fall 2020-21

Outline

Outline

Convex optimization

- A convex optimization problem is written as

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0 \qquad i = 1, \ldots, m \\ & \mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b} = 0 \end{array}$$

  where $f_0, f_1, \ldots, f_m$ are convex and equality constraints are affine.
- Convex problems enjoy a rich theory (KKT conditions, zero duality gap, etc.) as well as a large number of efficient numerical algorithms guaranteed to deliver an optimal solution $\mathbf{x}^{\star}$.
- Many off-the-shelf solvers exist in all the programming languages (e.g., R, Python, Matlab, Julia, C, etc.), tailored to specific classes of problems, namely, LP, QP, QCQP, SOCP, SDP, GP, etc.

Useless in practice!

- However, the obtained optimal solution $\mathbf{x}^{\star}$ typically performs very poorly in practice.
- In many cases, it can be totally useless!
- Why is that?
- Recall that a problem formulation contains not only the optimization variables $\mathbf{x}$ but also the parameters $\boldsymbol{\theta}$.
- Such parameters define the problem instance and are typically estimated in practice, i.e., they are not exact: $\hat{\boldsymbol{\theta}} \neq \boldsymbol{\theta}$ but hopefully close $\hat{\boldsymbol{\theta}} \simeq \boldsymbol{\theta}$.
- The question is whether a small error in the parameters is going to be detrimental or can be ignored. That depends on each particular type of problem.
- In the case of portfolio optimization, small errors in the parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ happen to have a huge effect in the solution $\mathbf{x}^{\star}$. To the point that most practitioners avoid the use of portfolio optimization!

Parameters: θ

- To make explicit the fact that the functions depend on parameters $\boldsymbol{\theta}$, we can explicitly write $f_i(\mathbf{x}; \boldsymbol{\theta})$ and $h_i(\mathbf{x}; \boldsymbol{\theta})$.

- For example, consider an LP:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \mathbf{c}^T\mathbf{x} + \mathbf{d} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array}$$

- The parameters are $\boldsymbol{\theta} = (\mathbf{A}, \mathbf{b}, \mathbf{c}, \mathbf{d})$.
- The objective function is $f_0(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{c}^T\mathbf{x} + \mathbf{d}$
- The constraint function is $\mathbf{h}(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{A}\mathbf{x} - \mathbf{b}$

- In practice, we only have an estimation $\hat{\boldsymbol{\theta}}$. So the problem can only be formulated and solved using $\hat{\boldsymbol{\theta}}$ obtaining the solution $\mathbf{x}^\star(\hat{\boldsymbol{\theta}})$, which is different from the desired one $\mathbf{x}^\star(\boldsymbol{\theta})$.

Robust optimization

- The naive approach is to pretend that $\hat{\boldsymbol{\theta}}$ is close enough to $\boldsymbol{\theta}$ and solve the approximated problem, obtaining $\mathbf{x}^{\star}(\hat{\boldsymbol{\theta}})$.
- For some type of problems, it may be that $\mathbf{x}^{\star}(\hat{\boldsymbol{\theta}}) \approx \mathbf{x}^{\star}(\boldsymbol{\theta})$ and that's it.
- For many other problems, however, that's not the case. So we cannot really rely on the naive solution $\mathbf{x}^{\star}(\hat{\boldsymbol{\theta}})$.
- The solution is to consider instead a robust formulation that takes into account the fact that we know we only have an estimation of the parameters.
- There are several ways to make the problem robust to parameters errors, mainly:
  - stochastic robust optimization (involving expectations)
  - worst-case robust optimization
  - chance programming or chance robust optimization.

Taxonomy of robust optimization

- **Stochastic optimization (SO)**: this includes expectations as well as chance constraints (requires probabilistic modeling of the parameter):

  J. R. Birge and F. V. Louveaux. *Introduction to Stochastic Programming*. Springer, 2011.

  A. P. Ruszczynski and A. Shapiro. *Stochastic Programming*. Elsevier, 2003.

  A. Prekopa. *Stochastic Programming*. Kluwer Academic Publishers, 1995.

- **Robust optimization (RO)**: this includes the worst-case approach (requires definition of hard uncertainty set for the parameter):

  A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

  A. Ben-Tal and A. Nemirovski, "Selected topics in robust convex optimization", *Mathematical Programming*, 112 (1), 2008.

  D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization", *SIAM Review*, 53 (3), 2011.

  M. S. Lobo. *Robust and convex optimization with applications in finance*. PhD thesis, Stanford University, 2000.

Stochastic optimization: Expectations

- In stochastic robust optimization, one models the estimation $\hat{\boldsymbol{\theta}}$ as a random variable that fluctuates around its true value $\boldsymbol{\theta}$.
- Then, instead of considering the approximated function $f(\mathbf{x};\hat{\boldsymbol{\theta}})$, it uses its expected value $E_{\boldsymbol{\theta}}[f(\mathbf{x};\boldsymbol{\theta})]$, where $E_{\boldsymbol{\theta}}[\cdot]$ denotes expectation over the random variable $\boldsymbol{\theta}$.
- The random variable is typically modeled around the estimated value as $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + \boldsymbol{\delta}$ with $\boldsymbol{\delta}$ following a zero-mean distribution such as Gaussian.
- For example, if the function is quadratic, say, $f(\mathbf{x};\hat{\boldsymbol{\theta}}) = (\hat{\mathbf{c}}^T\mathbf{x})^2$, and we model the parameter as $\mathbf{c} = \hat{\mathbf{c}} + \boldsymbol{\delta}$ with $\boldsymbol{\delta}$ zero-mean and covariance matrix $\mathbf{Q}$, then the expected value is

$$\begin{aligned} E_{\boldsymbol{\theta}}[f(\mathbf{x};\boldsymbol{\theta})] &= E_{\boldsymbol{\delta}}[((\hat{\mathbf{c}} + \boldsymbol{\delta})^T\mathbf{x})^2] \\ &= E_{\boldsymbol{\delta}}[\mathbf{x}^T\hat{\mathbf{c}}\hat{\mathbf{c}}^T\mathbf{x} + \mathbf{x}^T\boldsymbol{\delta}\boldsymbol{\delta}^T\mathbf{x}] \\ &= (\hat{\mathbf{c}}^T\mathbf{x})^2 + \mathbf{x}^T\mathbf{Q}\mathbf{x} \end{aligned}$$

  where the additional term $\mathbf{x}^T\mathbf{Q}\mathbf{x}$ serves as a regularizer.

Worst-case robust optimization

- In worst-case robust optimization, the parameter is not characterized statistically. Instead, it is assumed that the true parameter lies in an uncertainty region centered around the estimated value: $\boldsymbol{\theta} \in \mathcal{U}$.

- The uncertainty region can be chosen depending on the problem. Typical choices include:
  - sphere region:
  $$\mathcal{U} = \{\boldsymbol{\theta} | \parallel \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \parallel_2) \leq \delta\}$$
  - box region:
  $$\mathcal{U} = \{\boldsymbol{\theta} | \parallel \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \parallel_\infty) \leq \delta\}$$
  - elliptical region:
  $$\mathcal{U} = \{\boldsymbol{\theta} | (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{S}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})) \leq \delta^2\}$$

    where $\mathbf{S} \succ \mathbf{0}$ defines the shape of the ellipsoid.

Worst-case robust optimization: Example

Take the previous quadratic function

$$f(\mathbf{x};\hat{\boldsymbol{\theta}}) = (\hat{\mathbf{c}}^T\mathbf{x})^2$$

and consider a sphere uncertainty region

$$\mathcal{U} = \{\mathbf{c} | \parallel \mathbf{c} - \hat{\mathbf{c}} \parallel_2) \le \delta\}.$$

- If the function is the objective to be **minimized** or it is a constraint of the form $f(\mathbf{x};\hat{\boldsymbol{\theta}}) \le 0$, then the worst-case value of that function is

$$\begin{aligned}\max_{\mathbf{c}\in\mathcal{U}} \left|\mathbf{c}^T\mathbf{x}\right| &= \max_{\|\mathbf{e}\|\le\delta} \left|(\hat{\mathbf{c}}+\mathbf{e})^T\mathbf{x}\right| \\ &\le \max_{\|\mathbf{e}\|\le\delta} \left|\hat{\mathbf{c}}^T\mathbf{x}\right| + \left|\mathbf{e}^T\mathbf{x}\right| \\ &\le \left|\hat{\mathbf{c}}^T\mathbf{x}\right| + \delta\,\|\mathbf{x}\|\end{aligned}$$

  with upper bound achieved by $\mathbf{e} = \frac{\mathbf{x}}{\|\mathbf{x}\|}\delta$.

Worst-case robust optimization: Example

Take the previous quadratic function

$$f(\mathbf{x};\hat{\theta}) = (\hat{\mathbf{c}}^T\mathbf{x})^2$$

and consider a sphere uncertainty region

$$\mathcal{U} = \{\mathbf{c} | \parallel \mathbf{c} - \hat{\mathbf{c}} \parallel_2) \leq \delta\}.$$

- If the function is the objective to be **maximized** or it is a constraint of the form $f(\mathbf{x};\hat{\theta}) \geq 0$, then the worst-case value of that function is

$$\begin{aligned}\min_{\mathbf{c}\in\mathcal{U}} \left|\mathbf{c}^T\mathbf{x}\right| &= \min_{\|\mathbf{e}\|\leq\delta} \left|(\hat{\mathbf{c}}+\mathbf{e})^T\mathbf{x}\right| \\ &\geq \min_{\|\mathbf{e}\|\leq\delta} \left|\hat{\mathbf{c}}^T\mathbf{x}\right| - \left|\mathbf{e}^T\mathbf{x}\right| \\ &\geq \left|\hat{\mathbf{c}}^T\mathbf{x}\right| - \delta\,\|\mathbf{x}\|\end{aligned}$$

  with lower bound achieved by $\mathbf{e} = -\frac{\mathbf{x}}{\|\mathbf{x}\|}\delta$.

Stochastic optimization: Chance constraints

- The problem with expectations is that only the average behavior is concerned and nothing is under control about the realizations worse than the average. For example, on average some constraint will be satisfied but it will be violated for many realizations.
- The problem with worst-case programming is that it is too conservative as one deals with the worst possible case.
- Chance programming tries to find a compromise. In particular, it also models the estimation errors statistically but instead of focusing on the average it guarantees a performance for, say, 95% of the cases.
- The naive constraint $f(\mathbf{x};\hat{\boldsymbol{\theta}}) \leq 0$ is replaced with $\Pr_{\boldsymbol{\theta}}\left[f(\mathbf{x};\boldsymbol{\theta}) \leq 0\right] \geq 1-\epsilon = 0.95$ with $\epsilon = 0.05$.
- Chance or probabilistic constraints are generally very hard to deal with and one typically has to resort to approximations.

Outline

Robust Beamforming in Wireless Communications

👉 *Use whiteboard.*

Outline

Markowitz mean-variance portfolio (1952)

- The idea of the **Markowitz mean-variance portfolio (MVP)** (Markowitz 1952)[1] is to find a trade-off between the expected return $\mathbf{w}^T\boldsymbol{\mu}$ and the risk of the portfolio measured by the variance $\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\boldsymbol{\mu} - \lambda\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1 \end{array}$$

  where $\mathbf{w}^T\mathbf{1} = 1$ is the capital budget constraint and λ is a parameter that controls how risk-averse the investor is.

- This is a convex quadratic problem (QP) with only one linear constraint which admits a closed-form solution:

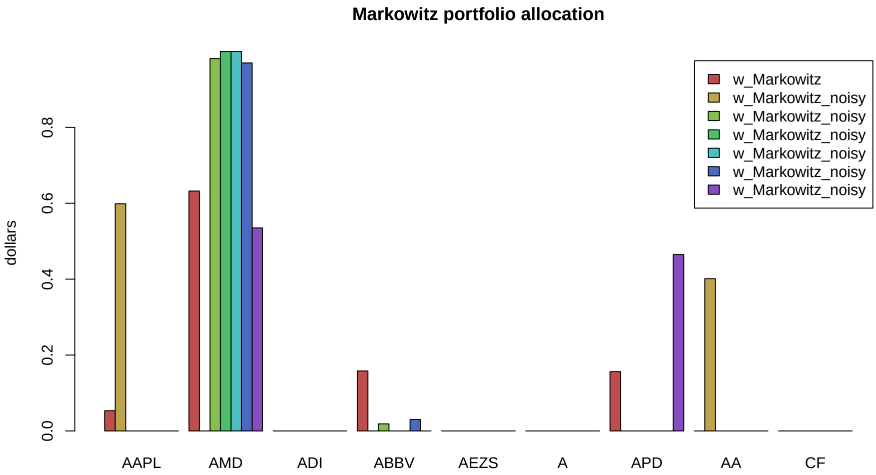
$$\mathbf{w}_{\text{MVP}} = \frac{1}{2\lambda}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} + \nu\mathbf{1}\right),$$

  where ν is the optimal dual variable $\nu = \frac{2\lambda - \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}$.

[1] H. Markowitz, "Portfolio selection," *J. Financ.*, vol. 7, no. 1, pp. 77–91, 1952.

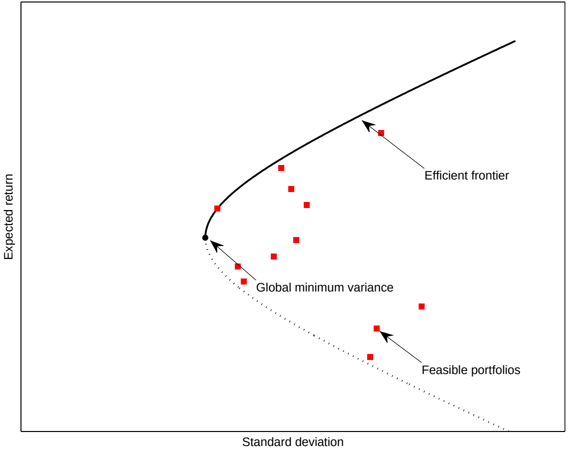
Sensitivity of Markowitz's portfolio

Markowitz's portfolio is extremely sensitivity of the estimatior errors of the parameters:



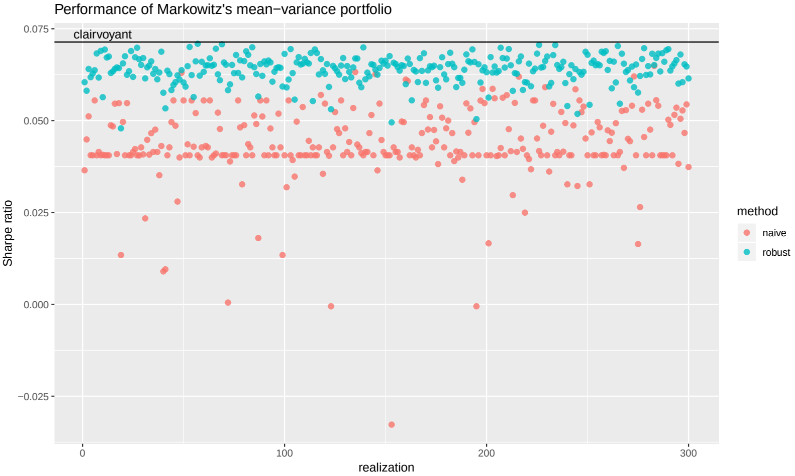
Mean-variance tradeoff

Efficient frontier: mean-variance trade-off curve (Pareto curve) but it is not achieved in practice due to parameter estimation errors:



Outline

Markowitz's portfolio: Naive vs robust



In terms of Sharpe ratio, the robust is clearly superior to the naive (note that the mean-variance portfolio is not the same as the maximum Sharpe ration portfolio).

Outline

Global maximum return portfolio (GMRP)

- The portfolio that maximizes the return (while ignoring the variance) is the linear program (LP)

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\boldsymbol{\mu} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1. \end{array}$$

- In practice, however, $\boldsymbol{\mu}$ is unknown and has to be estimated $\hat{\boldsymbol{\mu}}$, e.g., with the sample mean

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t,$$

where $\mathbf{x}_t$ is the return at day t.

- Unfortunately, it is well known that the estimation of $\boldsymbol{\mu}$ is extremely noisy in practice (Chopra and Ziemba 1993)[2].

[2] V. Chopra and W. Ziemba, "The effect of errors in means, variances and covariances on optimal portfolio choice," *Journal of Portfolio Management*, 1993.

Worst-case robust GMRP

- Instead of assuming that $\boldsymbol{\mu}$ is known perfectly, we now assume it belongs to some convex uncertainty set, denoted by $\mathcal{U}_{\boldsymbol{\mu}}$.
- The worst-case robust formulation is

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \underset{\boldsymbol{\mu} \in \mathcal{U}_{\boldsymbol{\mu}}}{\min} \mathbf{w}^T \boldsymbol{\mu} \\ \text{subject to} & \mathbf{1}^T \mathbf{w} = 1. \end{array}$$

- We assume the expected returns are only known within an ellipsoid:

$$\mathcal{U}_{\boldsymbol{\mu}} = \left\{ \boldsymbol{\mu} = \hat{\boldsymbol{\mu}} + \kappa \mathbf{S}^{1/2} \mathbf{u} \mid \|\mathbf{u}\|_2 \leq 1 \right\}$$

where one can use the estimated covariance matrix to shape the uncertainty ellipsoid, i.e., $\mathbf{S} = \hat{\boldsymbol{\Sigma}}$.

Worst-case robust GMRP

- We can solve easily the inner minimization:

$$\begin{array}{ll} \underset{\boldsymbol{\mu},\mathbf{u}}{\text{minimize}} & \mathbf{w}^T\boldsymbol{\mu} \\ \text{subject to} & \boldsymbol{\mu} = \hat{\boldsymbol{\mu}} + \kappa \mathbf{S}^{1/2}\mathbf{u}, \\ & \|\mathbf{u}\|_2 \leq 1. \end{array}$$

- It's easy to find the minimum value using Cauchy-Schwartz's inequality:

$$\begin{aligned} \mathbf{w}^T\boldsymbol{\mu} &= \mathbf{w}^T\left(\hat{\boldsymbol{\mu}} + \kappa \mathbf{S}^{1/2}\mathbf{u}\right) \\ &= \mathbf{w}^T\hat{\boldsymbol{\mu}} + \kappa \mathbf{w}^T\mathbf{S}^{1/2}\mathbf{u} \\ &\geq \mathbf{w}^T\hat{\boldsymbol{\mu}} - \kappa \left\|\mathbf{S}^{1/2}\mathbf{w}\right\|_2 \end{aligned}$$

  with equality achieved with $\mathbf{u} = -\frac{\mathbf{S}^{1/2}\mathbf{w}}{\left\|\mathbf{S}^{1/2}\mathbf{w}\right\|_2}$.

Worst-case robust GMRP

- Finally, the robust formulation becomes the SOCP

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\hat{\boldsymbol{\mu}} - \kappa \left\| \mathbf{S}^{1/2}\mathbf{w} \right\|_2 \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1. \end{array}$$

- Recall the vanilla problem formulation was the LP

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\hat{\boldsymbol{\mu}} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1. \end{array}$$

- So, we have gone from an LP to an SOCP.
- In general, when a problem is robustified, the complexity of the problem increases. For example:
  - LP becomes SOCP
  - QP also becomes SOCP
  - SOCP becomes SDP.

Outline

Global Minimum Variance Portfolio (GMVP)

- The **global minimum variance portfolio (GMVP)** ignores the expected return and focuses on the risk only:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} & \mathbf{1}^T \mathbf{w} = 1. \end{array}$$

- It is a simple convex QP with solution

$$\mathbf{w}_{\text{GMVP}} = \frac{1}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}} \boldsymbol{\Sigma}^{-1} \mathbf{1}.$$

- It is widely used in academic papers for simplicity of evaluation and comparison of different estimators of the covariance matrix $\boldsymbol{\Sigma}$ (while ignoring the estimation of $\boldsymbol{\mu}$).
- In practice, $\boldsymbol{\Sigma}$ is unknown and has to be estimated $\hat{\boldsymbol{\Sigma}}$, e.g., with the sample covariance matrix.
- Then, the naive portfolio becomes

$$\mathbf{w}_{\text{GMVP}} = \frac{1}{\mathbf{1}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}.$$

Worst-case robust GMVP

- Instead of assuming that $\boldsymbol{\Sigma}$ is known perfectly, we now assume it belongs to some convex uncertainty set, denoted by $\mathcal{U}_{\boldsymbol{\Sigma}}$.
- The worst-case robust formulation is

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \max_{\boldsymbol{\Sigma} \in \mathcal{U}_{\boldsymbol{\Sigma}}} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} & \mathbf{1}^T \mathbf{w} = 1. \end{array}$$

- In particular, we will assume that the estimation comes from the sample covariance matrix $\hat{\boldsymbol{\Sigma}} = \frac{1}{T} \mathbf{X}^T \mathbf{X}$ where $\mathbf{X}$ is a $T \times N$ matrix containing the return data (assumed demeaned already).
- However, we will assume that the data matrix is noisy $\hat{\mathbf{X}}$ and the actual matrix can be written as $\mathbf{X} = \hat{\mathbf{X}} + \boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ is some error matrix bounded in its norm.
- Thus, we will then model the data matrix as

$$\mathcal{U}_{\mathbf{X}} = \left\{ \mathbf{X} \mid \left\| \mathbf{X} - \hat{\mathbf{X}} \right\|_F \leq \delta_{\mathbf{X}} \right\}.$$

Worst-case robust GMVP

- The worst-case robust formulation becomes:

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{minimize}} & \max_{\mathbf{X}\in\mathcal{U}_{\mathbf{X}}} \mathbf{w}^T\frac{1}{T}\mathbf{X}^T\mathbf{X}\mathbf{w}\\ \text{subject to} & \mathbf{1}^T\mathbf{w}=1.\end{array}$$

- Let's focus on the inner maximization:

$$\max_{\mathbf{X}\in\mathcal{U}_{\mathbf{X}}} \mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} = \max_{\|\boldsymbol{\Delta}\|_F\leq\delta_{\mathbf{X}}} \left\|\left(\hat{\mathbf{X}}+\boldsymbol{\Delta}\right)\mathbf{w}\right\|_2^2$$

- We first invoke the triangle inequality to get an upper bound:

$$\left\|\left(\hat{\mathbf{X}}+\boldsymbol{\Delta}\right)\mathbf{w}\right\|_2 \leq \left\|\hat{\mathbf{X}}\mathbf{w}\right\|_2 + \|\boldsymbol{\Delta}\mathbf{w}\|_2$$

  with equality achieved when the two vectors $\hat{\mathbf{X}}\mathbf{w}$ and $\boldsymbol{\Delta}\mathbf{w}$ are aligned.
- Next, we invoke the norm inequality

$$\|\boldsymbol{\Delta}\mathbf{w}\|_2 \leq \|\boldsymbol{\Delta}\|_F\|\mathbf{w}\|_2 \leq \delta_{\mathbf{X}}\|\mathbf{w}\|_2$$

  with equality achieved when $\boldsymbol{\Delta}$ is rank-one with right singular vector aligned with $\mathbf{w}$ with $\|\boldsymbol{\Delta}\|_F=\delta_{\mathbf{X}}$. (This follows from $\mathbf{w}^T\mathbf{M}\mathbf{w}\leq\lambda_{\max}(\mathbf{M})\|\mathbf{w}\|^2\leq\text{Tr}(\mathbf{M})\|\mathbf{w}\|^2$ for $\mathbf{M}\succeq\mathbf{0}$.)

Worst-case robust GMVP

- Finally, we can see that both upper bounds can be actually achieved if the error is properly chosen as

$$\boldsymbol{\Delta} = \delta_{\mathbf{X}} \frac{\hat{\mathbf{X}}\mathbf{w}\mathbf{w}^T}{\|\mathbf{w}\|_2 \left\|\hat{\mathbf{X}}\mathbf{w}\right\|_2}.$$

- Thus,

$$\max_{\mathbf{X} \in \mathcal{U}_{\mathbf{X}}} \mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} = \left(\left\|\hat{\mathbf{X}}\mathbf{w}\right\|_2 + \delta_{\mathbf{X}} \|\mathbf{w}\|_2\right)^2.$$

- The robust problem formulation finally becomes:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \left\|\hat{\mathbf{X}}\mathbf{w}\right\|_2 + \delta_{\mathbf{X}} \|\mathbf{w}\|_2 \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1 \end{array}$$

which is a (convex) SOCP.

Worst-case robust GMVP

- Recall the vanilla problem formulation was the QP

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \left\|\hat{\mathbf{X}}\mathbf{w}\right\|_2^2 \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1 \end{array}$$

- Now, the robust problem formulation is the SOCP (from QP to SOCP)

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \left\|\hat{\mathbf{X}}\mathbf{w}\right\|_2 + \delta_{\mathbf{X}}\left\|\mathbf{w}\right\|_2 \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1 \end{array}$$

  which contains the regularization term $\delta_{\mathbf{X}}\left\|\mathbf{w}\right\|_2$.
- One common heuristic, called Tikhonov regularization, is to consider instead

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \left\|\hat{\mathbf{X}}\mathbf{w}\right\|_2^2 + \delta_{\mathbf{X}}\left\|\mathbf{w}\right\|_2^2 = \mathbf{w}^T\left(\hat{\mathbf{X}}^T\hat{\mathbf{X}} + \delta_{\mathbf{X}}\mathbf{I}\right)\mathbf{w} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1 \end{array}$$

  which is equivalent to the vanilla formulation but using the regularized sample covariance matrix $\hat{\boldsymbol{\Sigma}}^{\text{tik}} = \frac{1}{T}(\hat{\mathbf{X}}^T\hat{\mathbf{X}} + \delta_{\mathbf{X}}\mathbf{I}) = \hat{\boldsymbol{\Sigma}} + \frac{\delta_{\mathbf{X}}}{T}\mathbf{I}$.

Outline

Markowitz's portfolio formulation

- Recall Markowitz's formulation:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\boldsymbol{\mu} - \lambda\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1, \quad \mathbf{w} \in \mathcal{W}, \end{array}$$

  where $\mathcal{W}$ denotes some other constraints on $\mathbf{w}$.

- Instead of assuming $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known perfectly, now we assume they belong to some convex uncertainty sets, denoted as $\mathcal{U}_{\boldsymbol{\mu}}$ and $\mathcal{U}_{\boldsymbol{\Sigma}}$, respectively.
- The worst-case formulation will consider the worst-case points within those uncertainty sets.

Worst-case Markowitz's portfolio

- A conservative and practical investment approach is to optimize the worst-case objective over the uncertainty sets (Cornuejols and Tütüncü 2006)[3], (Fabozzi 2007)[4]:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \min_{\boldsymbol{\mu} \in \mathcal{U}_{\boldsymbol{\mu}}} \mathbf{w}^T \boldsymbol{\mu} - \lambda \max_{\boldsymbol{\Sigma} \in \mathcal{U}_{\boldsymbol{\Sigma}}} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} & \mathbf{w}^T \mathbf{1} = 1, \quad \mathbf{w} \in \mathcal{W}. \end{array}$$

- The two key issues are:
  1. How to choose the uncertainty sets $\mathcal{U}_{\boldsymbol{\mu}}$ and $\mathcal{U}_{\boldsymbol{\Sigma}}$ so that they are meaningful in practice.
  2. To make sure the optimization problem above is still easy to solve.

[3] G. Cornuejols and R. Tütüncü, *Optimization Methods in Finance*. Cambridge University Press, 2006.
[4] F. J. Fabozzi, *Robust Portfolio Optimization and Management*. Wiley, 2007.

Worst-case mean: Box set

- Box uncertainty set:
$$\mathcal{U}_{\boldsymbol{\mu}}^{b} = \left\{\boldsymbol{\mu} \mid -\boldsymbol{\delta} \le \boldsymbol{\mu} - \hat{\boldsymbol{\mu}} \le \boldsymbol{\delta}\right\},$$
where the predefined parameters $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\delta}$ denote the location and size of the box uncertainty set, respectively.
- Easily, the worst-case mean is
$$\min_{\boldsymbol{\mu} \in \mathcal{U}_{\boldsymbol{\mu}}^{b}} \mathbf{w}^T \boldsymbol{\mu} = \mathbf{w}^T \hat{\boldsymbol{\mu}} + \min_{-\boldsymbol{\delta} \le \boldsymbol{\gamma} \le \boldsymbol{\delta}} \mathbf{w}^T \boldsymbol{\gamma} = \mathbf{w}^T \hat{\boldsymbol{\mu}} - |\mathbf{w}|^T \boldsymbol{\delta},$$
where $|\mathbf{w}|$ denotes elementwise absolute value of $\mathbf{w}$.
- Note that it is a concave function of $\mathbf{w}$ (as it should be since it is the minimum of linear functions).

Worst-case mean: Elliptical set

- Elliptical uncertainty set:

$$\mathcal{U}_{\mu}^{e} = \left\{ \boldsymbol{\mu} \mid (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^{T} \mathbf{S}_{\mu}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \delta_{\mu}^{2} \right\},$$

  where the predefined parameters $\hat{\boldsymbol{\mu}}$, $\delta_{\mu} > 0$, and $\mathbf{S}_{\mu} \succ \mathbf{0}$ denote the location, size, and the shape of the uncertainty set, respectively.

- The worst-case mean is

$$\begin{aligned} \min_{\boldsymbol{\mu} \in \mathcal{U}_{\mu}^{e}} \mathbf{w}^{T} \boldsymbol{\mu} &= \min_{\left\| \mathbf{S}_{\mu}^{-1/2} \gamma \right\|_{2} \leq \delta_{\mu}} \mathbf{w}^{T} (\hat{\boldsymbol{\mu}} + \gamma) = \mathbf{w}^{T} \hat{\boldsymbol{\mu}} + \min_{\left\| \mathbf{S}_{\mu}^{-1/2} \gamma \right\|_{2} \leq \delta_{\mu}} \mathbf{w}^{T} \gamma \\ &= \mathbf{w}^{T} \hat{\boldsymbol{\mu}} + \min_{\|\tilde{\gamma}\|_{2} \leq \delta_{\mu}} \mathbf{w}^{T} \mathbf{S}_{\mu}^{1/2} \tilde{\gamma} = \mathbf{w}^{T} \hat{\boldsymbol{\mu}} - \delta_{\mu} \left\| \mathbf{S}_{\mu}^{1/2} \mathbf{w} \right\|_{2}. \end{aligned}$$

- Note that it is a concave function of $\mathbf{w}$ (as it should be since it is the minimum of linear functions).

Worst-case variance: Box set

- Box uncertainty set:
$$\mathcal{U}_{\boldsymbol{\Sigma}}^{b} = \left\{ \boldsymbol{\Sigma} \mid \underline{\boldsymbol{\Sigma}} \leq \boldsymbol{\Sigma} \leq \overline{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma} \succeq \mathbf{0} \right\}.$$

- The worst-case value $\max_{\boldsymbol{\Sigma} \in \mathcal{U}_{\boldsymbol{\Sigma}}^{b}} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$ is given by the (convex) semidefinite problem (SDP)
$$\begin{array}{ll} \underset{\boldsymbol{\Sigma}}{\text{maximize}} & \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} & \underline{\boldsymbol{\Sigma}} \leq \boldsymbol{\Sigma} \leq \overline{\boldsymbol{\Sigma}}, \\ & \boldsymbol{\Sigma} \succeq \mathbf{0}. \end{array}$$

Worst-case variance: Box set

- The equivalent dual problem is (Lobo and Boyd 2000)[5]

$$\begin{array}{ll} \underset{\overline{\boldsymbol{\Lambda}}, \underline{\boldsymbol{\Lambda}}}{\text{minimize}} & \text{Tr}(\overline{\boldsymbol{\Lambda}}\,\overline{\boldsymbol{\Sigma}}) - \text{Tr}(\underline{\boldsymbol{\Lambda}}\,\underline{\boldsymbol{\Sigma}}) \\ \text{subject to} & \begin{bmatrix} \overline{\boldsymbol{\Lambda}} - \underline{\boldsymbol{\Lambda}} & \mathbf{w} \\ \mathbf{w}^T & 1 \end{bmatrix} \succeq \mathbf{0}, \\ & \overline{\boldsymbol{\Lambda}} \geq \mathbf{0}, \quad \underline{\boldsymbol{\Lambda}} \geq \mathbf{0}, \end{array}$$

  which is a convex SDP.
- The constraints are jointly convex in the inner dual variable variables $\overline{\boldsymbol{\Lambda}}$ and $\underline{\boldsymbol{\Lambda}}$ and the outer variable $\mathbf{w}$.

[5] M. S. Lobo and S. Boyd, "The worst-case risk of a portfolio," Tech. Rep., 2000.

Worst-case variance: Elliptical set

- Elliptical uncertainty set:

$$\mathcal{U}_{\boldsymbol{\Sigma}}^{e}=\left\{\boldsymbol{\Sigma}\mid\left(\mathrm{vec}(\boldsymbol{\Sigma})-\mathrm{vec}(\hat{\boldsymbol{\Sigma}})\right)^{T}\mathbf{S}_{\boldsymbol{\Sigma}}^{-1}\left(\mathrm{vec}(\boldsymbol{\Sigma})-\mathrm{vec}(\hat{\boldsymbol{\Sigma}})\right)\leq\delta_{\boldsymbol{\Sigma}}^{2},\boldsymbol{\Sigma}\succeq\mathbf{0}\right\}$$

where $\hat{\boldsymbol{\Sigma}}$ denotes the location, $\delta_{\boldsymbol{\Sigma}}$ denotes the size, and $\mathbf{S}_{\boldsymbol{\Sigma}}$ determines the shape.

- The worst-case value $\max_{\boldsymbol{\Sigma}\in\mathcal{U}_{\boldsymbol{\Sigma}}^{e}}\mathbf{w}^{T}\boldsymbol{\Sigma}\mathbf{w}$ is given by the (convex) SDP

$$\begin{array}{ll}\underset{\boldsymbol{\Sigma}}{\text{maximize}} & \mathbf{w}^{T}\boldsymbol{\Sigma}\mathbf{w}\\ \text{subject to} & \left(\mathrm{vec}(\boldsymbol{\Sigma})-\mathrm{vec}(\hat{\boldsymbol{\Sigma}})\right)^{T}\mathbf{S}_{\boldsymbol{\Sigma}}^{-1}\left(\mathrm{vec}(\boldsymbol{\Sigma})-\mathrm{vec}(\hat{\boldsymbol{\Sigma}})\right)\leq\delta_{\boldsymbol{\Sigma}}^{2},\\ & \boldsymbol{\Sigma}\succeq\mathbf{0}.\end{array}$$

Worst-case variance: Elliptical set

- Since the problem is convex, strong duality holds (zero duality gap).
- Thus, the maximum objective value equals the minimum objective value of its dual problem:

$$\begin{array}{ll} \underset{\mathbf{Z}}{\text{minimize}} & \text{Tr}\left(\hat{\boldsymbol{\Sigma}}\left(\mathbf{w}\mathbf{w}^T + \mathbf{Z}\right)\right) + \delta_{\boldsymbol{\Sigma}}\left\|\mathbf{S}_{\boldsymbol{\Sigma}}^{1/2}\left(\text{vec}(\mathbf{w}\mathbf{w}^T) + \text{vec}(\mathbf{Z})\right)\right\|_2 \\ \text{subject to} & \mathbf{Z} \succeq \mathbf{0}. \end{array}$$

- We can now plug in this inner problem in the original problem:

$$\begin{array}{ll} \underset{\mathbf{w},\mathbf{Z}}{\text{maximize}} & \underset{\boldsymbol{\mu}\in\mathcal{U}_{\boldsymbol{\mu}}}{\min}\, \mathbf{w}^T\boldsymbol{\mu} - \lambda\left(\text{Tr}\left(\hat{\boldsymbol{\Sigma}}\left(\mathbf{w}\mathbf{w}^T + \mathbf{Z}\right)\right)\right. \\ & \qquad \left. +\delta_{\boldsymbol{\Sigma}}\left\|\mathbf{S}_{\boldsymbol{\Sigma}}^{1/2}\left(\text{vec}(\mathbf{w}\mathbf{w}^T) + \text{vec}(\mathbf{Z})\right)\right\|_2\right) \\ \text{subject to} & \mathbf{w}^T\mathbf{1} = 1, \quad \mathbf{w} \in \mathcal{W} \\ & \mathbf{Z} \succeq \mathbf{0}. \end{array}$$

- However, now this problem contains a complicated term with the composition of $\text{vec}(\mathbf{w}\mathbf{w}^T)$ and the norm $\|\cdot\|_2$.

Worst-case variance: Elliptical set

- We can further include a new variable as $\mathbf{X} = \mathbf{w}\mathbf{w}^T$ so that the objective function becomes nicer.
- But this constraint is not convex!
- Luckily we can instead use $\mathbf{X} \succeq \mathbf{w}\mathbf{w}^T$ (because we can easily show that at the optimal point it will be achieved with equality), which can be further expressed as $\begin{bmatrix} \mathbf{X} & \mathbf{w} \\ \mathbf{w}^T & 1 \end{bmatrix} \succeq \mathbf{0}$.
- The final problem is

$$
\begin{array}{ll}
\underset{\mathbf{w},\mathbf{X},\mathbf{Z}}{\text{maximize}} & \underset{\boldsymbol{\mu}\in\mathcal{U}_{\boldsymbol{\mu}}}{\min} \mathbf{w}^T\boldsymbol{\mu} - \lambda\left(\mathrm{Tr}\left(\hat{\boldsymbol{\Sigma}}\left(\mathbf{X}+\mathbf{Z}\right)\right)\right. \\
& \qquad\qquad \left. +\delta_{\boldsymbol{\Sigma}}\left\|\mathbf{S}_{\boldsymbol{\Sigma}}^{1/2}\left(\mathrm{vec}(\mathbf{X})+\mathrm{vec}(\mathbf{Z})\right)\right\|_2\right) \\
\text{subject to} & \mathbf{w}^T\mathbf{1} = 1, \quad \mathbf{w}\in\mathcal{W} \\
& \begin{bmatrix} \mathbf{X} & \mathbf{w} \\ \mathbf{w}^T & 1 \end{bmatrix} \succeq \mathbf{0} \\
& \mathbf{Z} \succeq \mathbf{0}.
\end{array}
$$

Equivalent formulation

- Consider the two uncertainty sets:

$$\mathcal{U}^b_{\boldsymbol{\mu}} = \{\boldsymbol{\mu} \mid -\boldsymbol{\delta} \le \boldsymbol{\mu} - \hat{\boldsymbol{\mu}} \le \boldsymbol{\delta}\},$$
$$\mathcal{U}^b_{\boldsymbol{\Sigma}} = \left\{\boldsymbol{\Sigma} \mid \underline{\boldsymbol{\Sigma}} \le \boldsymbol{\Sigma} \le \overline{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma} \succeq \mathbf{0}\right\}.$$

- The worst-case robust formulation

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \min_{\boldsymbol{\mu}\in\mathcal{U}_{\boldsymbol{\mu}}} \mathbf{w}^T\boldsymbol{\mu} - \lambda \max_{\boldsymbol{\Sigma}\in\mathcal{U}_{\boldsymbol{\Sigma}}} \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{w}^T\mathbf{1} = 1, \quad \mathbf{w} \in \mathcal{W}. \end{array}$$

becomes

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\hat{\boldsymbol{\mu}} - |\mathbf{w}|^T\boldsymbol{\delta} - \lambda \begin{array}[t]{ll} \underset{\overline{\boldsymbol{\Lambda}},\underline{\boldsymbol{\Lambda}}}{\min} & \left\{\mathrm{Tr}(\overline{\boldsymbol{\Lambda}}\,\overline{\boldsymbol{\Sigma}}) - \mathrm{Tr}(\underline{\boldsymbol{\Lambda}}\,\underline{\boldsymbol{\Sigma}})\right\} \\ \text{subject to} & \begin{bmatrix} \overline{\boldsymbol{\Lambda}} - \underline{\boldsymbol{\Lambda}} & \mathbf{w} \\ \mathbf{w}^T & 1 \end{bmatrix} \succeq \mathbf{0}, \\ & \overline{\boldsymbol{\Lambda}} \ge \mathbf{0}, \quad \underline{\boldsymbol{\Lambda}} \ge \mathbf{0}, \end{array} \\ \text{subject to} & \mathbf{w}^T\mathbf{1} = 1, \quad \mathbf{w} \in \mathcal{W}. \end{array}$$

Equivalent formulation

- Finally, the worst-case robust formulation can be written as the (convex) SDP:

$$\begin{array}{ll}\underset{\mathbf{w},\overline{\mathbf{\Lambda}},\underline{\mathbf{\Lambda}}}{\text{maximize}} & \mathbf{w}^T\hat{\mu} - |\mathbf{w}|^T\delta - \lambda\left(\text{Tr}(\overline{\mathbf{\Lambda}}\,\overline{\mathbf{\Sigma}}) - \text{Tr}(\underline{\mathbf{\Lambda}}\,\underline{\mathbf{\Sigma}})\right)\\ \text{subject to} & \mathbf{w}^T\mathbf{1} = 1, \quad \mathbf{w} \in \mathcal{W},\\ & \begin{bmatrix}\overline{\mathbf{\Lambda}} - \underline{\mathbf{\Lambda}} & \mathbf{w}\\ \mathbf{w}^T & 1\end{bmatrix} \succeq \mathbf{0},\\ & \overline{\mathbf{\Lambda}} \geq \mathbf{0}, \quad \underline{\mathbf{\Lambda}} \geq \mathbf{0}.\end{array}$$

- This problem does not have a closed-form solution, but it is an SDP that can be easily solved with an off-the-shelf SDP solver.

More cases of robust formulations

F. J. Fabozzi. *Robust Portfolio Optimization and Management*. Wiley, 2007.

D. Goldfarb and G. Iyengar, "Robust portfolio selection problems", *Mathematics of Operations Research*, vol. 28, no. 1, pp. 1–38, 2003.

R. H. Tutuncu and M. Koenig, "Robust asset allocation", *Annals Operations Research*, vol. 132, no. 1, pp. 157–187, 2004.

L. El Ghaoui, M. Oks, and F. Oustry, "Worst-case value-at-risk and robust portfolio optimization: A conic programming approach", *Operations Research*, pp. 543–556, 2003.

Z. Lu, "Robust portfolio selection based on a joint ellipsoidal uncertainty set", *Optimization Methods & Software*, vol. 26, no. 1, pp. 89–104, 2011.

Outline

Summary

Naive optimization: optimization problems formulated assuming that the parameters are perfectly known when they are not.
👉 the naive solution $\mathbf{x}^{\star}(\hat{\boldsymbol{\theta}})$ may totally differ from the desired one $\mathbf{x}^{\star}(\boldsymbol{\theta})$ (or not, depends on the type of problem)

Optimization under uncertainty of parameters:

- **stochastic optimization (SO)**: models the parameters statistically and uses expectations and probabilities
  - requires modeling the probability distribution function
  - expectations only satisfy constraints on average, not for every instance
  - chance constraints are very hard to manipulate
- **robust optimization (RO)**: assumes the true parameter is inside an uncertainty region centered around the estimation
  - the shape of the uncertainty region has to be chosen appropriately for the problem at hand
  - the size of the uncertainty region has to be carefully chosen or the solution may be too conservative to the point of being meaningless
  - usually easy to manipulate.

Thanks

For more information visit:

https://www.danielppalomar.com

References I

Chopra, V., & Ziemba, W. (1993). The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*.

Cornuejols, G., & Tütüncü, R. (2006). *Optimization methods in finance*. Cambridge University Press.

Fabozzi, F. J. (2007). *Robust portfolio optimization and management*. Wiley.

Lobo, M. S., & Boyd, S. (2000). *The worst-case risk of a portfolio*.

Markowitz, H. (1952). Portfolio selection. *J. Financ.*, *7*(1), 77–91.