EE/ACM 150 - Applications of Convex Optimization in Signal Processing and Communications
Lecture 10

Andre Tkacenko

Signal Processing Research Group
Jet Propulsion Laboratory

May 3, 2012

Outline

Vector Optimization Problems

As opposed to problems with generalized inequality constraints, where the inequality constraint functions were vector-valued, in vector optimization problems, the *objective* is vector-valued and the optimization is carried out with respect to a proper cone.

General Vector Optimization Problem

$$\begin{array}{ll} \text{minimize with respect to } \mathcal{K} & \mathbf{f}_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0 \,,\ i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0 \,,\ i = 1, \ldots, p \end{array} .$$

The vector objective $\mathbf{f}_0 : \mathbb{R}^n \to \mathbb{R}^q$ is minimized with respect to the proper cone $\mathcal{K} \subseteq \mathbb{R}^q$.

Convex Vector Optimization Problem

$$\begin{array}{ll} \text{minimize with respect to } \mathcal{K} & \mathbf{f}_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0 \,,\ i = 1, \ldots, m \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} .$$

The objective $\mathbf{f}_0 : \mathbb{R}^n \to \mathbb{R}^q$ is $\mathcal{K}$-convex, $f_1, \ldots, f_m$ are convex, $\mathbf{A} \in \mathbb{R}^{p \times n}$, and $\mathbf{b} \in \mathbb{R}^p$.

Optimal / Pareto Optimal Points for Vector Problems

Let $\mathcal{O}$ denote the set of achievable *objective* values given by

$$\mathcal{O} = \{\mathbf{f}_0(\mathbf{x}) : \exists\, \mathbf{x} \in \mathcal{D}\,,\ f_i(\mathbf{x}) \leq 0\,,\ i = 1, \ldots, m\,,\ h_i(\mathbf{x}) = 0\,,\ i = 1, \ldots, p\} \subseteq \mathbb{R}^q\,.$$

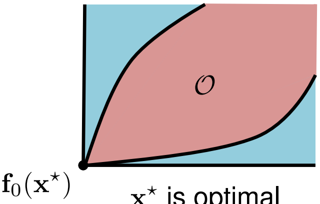
In other words, $\mathcal{O}$ consists of the values of $\mathbf{f}_0(\mathbf{x})$ such that $\mathbf{x}$ is feasible.

- A point $\mathbf{x}^\star$ is said to be *optimal* if $\mathbf{f}_0(\mathbf{x}^\star)$ is the minimum element of $\mathcal{O}$, i.e., if and only if it is feasible and satisfies

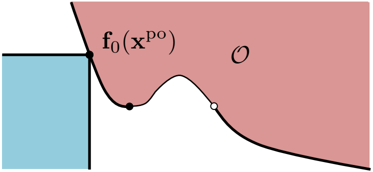
$$\mathcal{O} \subseteq \mathbf{f}_0(\mathbf{x}^\star) + \mathcal{K}\,.$$

- A point $\mathbf{x}^{\mathrm{po}}$ is said to be *Pareto optimal* if $\mathbf{f}_0(\mathbf{x}^{\mathrm{po}})$ is a minimal element of $\mathcal{O}$, i.e., if and only if it is feasible and satisfies

$$\left(\mathbf{f}_0(\mathbf{x}^{\mathrm{po}}) - \mathcal{K}\right) \cap \mathcal{O} = \{\mathbf{f}_0(\mathbf{x}^{\mathrm{po}})\}\ .$$



$\mathbf{x}^\star$ is optimal



$\mathbf{x}^{\mathrm{po}}$ is Pareto optimal

Multicriterion Optimization

A vector optimization problem with $\mathcal{K} = \mathbb{R}_+^q$ is said to be a *multicriterion* or *multi-objective* optimization problem. In this case, we have

$$\mathbf{f}_0(\mathbf{x}) = \begin{bmatrix} F_1(\mathbf{x}) & \cdots & F_q(\mathbf{x}) \end{bmatrix}^T ,$$

where $F_1, \ldots, F_q$ represent the q different objectives that we wish to minimize. In other words, we want every F_i for $i = 1, \ldots, q$ to be small.

- A feasible $\mathbf{x}^\star$ is optimal if

$$\mathbf{y} \text{ feasible} \Longrightarrow \mathbf{f}_0(\mathbf{x}^\star) \preceq \mathbf{f}_0(\mathbf{y}) \Longleftrightarrow F_i(\mathbf{x}^\star) \leq F_i(\mathbf{y}) \ , \ i = 1, \ldots, q \, .$$

- A feasible $\mathbf{x}^{\mathrm{po}}$ is Pareto optimal if

$$\mathbf{y} \text{ feasible, } \mathbf{f}_0(\mathbf{y}) \preceq \mathbf{f}_0(\mathbf{x}^{\mathrm{po}}) \Longrightarrow \mathbf{f}_0(\mathbf{x}^{\mathrm{po}}) = \mathbf{f}_0(\mathbf{y}) \ .$$

  If there are multiple Pareto optimal values, then there is a *trade-off* between the objectives. Namely, if $\mathbf{x}$ and $\mathbf{y}$ are Pareto optimal points with

$$\begin{aligned} F_i(\mathbf{x}) &< F_i(\mathbf{y}) \ , \quad i \in \mathcal{A} \, , \\ F_i(\mathbf{x}) &= F_i(\mathbf{y}) \ , \quad i \in \mathcal{B} \, , \\ F_i(\mathbf{x}) &> F_i(\mathbf{y}) \ , \quad i \in \mathcal{C} \, , \end{aligned}$$

  where $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} = \{1, \ldots, q\}$, then either $\mathcal{A}$ and $\mathcal{C}$ are both empty or both nonempty.

Solving Vector Optimization Problems Via Scalarization

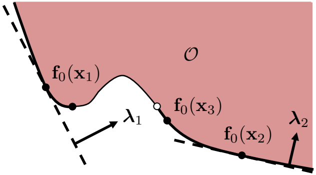
A standard method for finding optimal or Pareto optimal points is through *scalarization*, which is based on the characterization of minimum and minimal elements via dual generalized inequalities.

Scalarization

Choose any $\boldsymbol{\lambda} \succ_{\mathcal{K}^\circledast} \mathbf{0}$ and solve the *scalar* problem

$$\begin{array}{ll} \text{minimize} & \boldsymbol{\lambda}^T \mathbf{f}_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0\,,\ i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0\,,\ i = 1, \ldots, p \end{array} .$$

If $\mathbf{x}$ is optimal for the scalar problem, then it is Pareto optimal for the vector optimization problem. However, there may be Pareto optimal points which cannot be obtained through scalarization.



Scalarization of Convex Vector Optimization Problems:

- Almost all Pareto optimal points can be obtained by varying $\boldsymbol{\lambda} \succ_{\mathcal{K}^\circledast} \mathbf{0}$.
- For every Pareto optimal point $\mathbf{x}^{\text{po}}$, there is some nonzero $\boldsymbol{\lambda} \succeq_{\mathcal{K}^\circledast} \mathbf{0}$ such that $\mathbf{x}^{\text{po}}$ is a solution of the scalarized problem. However, not every solution of the scalarized problem with $\boldsymbol{\lambda} \succeq_{\mathcal{K}^\circledast} \mathbf{0}$ and $\boldsymbol{\lambda} \neq \mathbf{0}$ is a Pareto optimal point for the vector problem.

Scalarization for Multicriterion Problems

To find Pareto optimal points via scalarization for multicriterion problems, we minimize

$$\boldsymbol{\lambda}^T \mathbf{f}_0(\mathbf{x}) = \sum_{i=1}^{q} \lambda_i F_i(\mathbf{x}) \ ,$$

where $\lambda_i > 0$ for $i = 1, \ldots, q$. In other words, we minimize a positive weighted sum of the objectives.

- By varying $\boldsymbol{\lambda}$ over $\boldsymbol{\lambda} \succ \mathbf{0}$, this will trace out or explore the *optimal trade-off surface* (for $q > 2$) or the *optimal trade-off curve* (for $q = 2$).
- The quantity λ_k / λ_ℓ is the *relative weight* or relative importance of the k-th objective compared to the ℓ-th objective. Alternatively, λ_k / λ_ℓ is an *exchange rate* between the two objectives: a decrease in F_k by α is the same as an increase in F_ℓ by $(\lambda_k / \lambda_\ell) \, \alpha$.
- Starting from a weight vector $\boldsymbol{\lambda}$ yielding a Pareto optimal point $\mathbf{x}^{\mathrm{po}}$, to obtain a new Pareto optimal point which trades off a better k-th objective value for possibly worse values for the other objectives, we form a new weight vector $\widetilde{\boldsymbol{\lambda}}$ with

$$\widetilde{\lambda}_k > \lambda_k \ , \ \widetilde{\lambda}_\ell = \lambda_\ell \ , \ \ell \neq k \ , \ \ell = 1, \ldots, q \, .$$

  This yields a new Pareto optimal point $\widetilde{\mathbf{x}}^{\mathrm{po}}$ with $F_k(\widetilde{\mathbf{x}}^{\mathrm{po}}) \leq F_k(\mathbf{x}^{\mathrm{po}})$. Typically, we have $F_k(\widetilde{\mathbf{x}}^{\mathrm{po}}) < F_k(\mathbf{x}^{\mathrm{po}})$.

Regularized Least-Squares

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, we want to choose $\mathbf{x} \in \mathbb{R}^n$ to minimize the two quadratic objectives:

- $F_1(\mathbf{x}) = ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 = \mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} - 2\mathbf{b}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{b}$, which is a measure of the misfit between $\mathbf{A}\mathbf{x}$ and $\mathbf{b}$,
- $F_2(\mathbf{x}) = ||\mathbf{x}||_2^2 = \mathbf{x}^T \mathbf{x}$, which is a measure of the length of $\mathbf{x}$.

This yields the bi-criterion problem:

$$\text{minimize (with respect to } \mathbb{R}_+^2) \quad \mathbf{f}_0(\mathbf{x}) = \left(F_1(\mathbf{x}), F_2(\mathbf{x})\right) = \left(||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 , ||\mathbf{x}||_2^2\right) \ .$$

Scalarization yields the scalar weighted sum objective

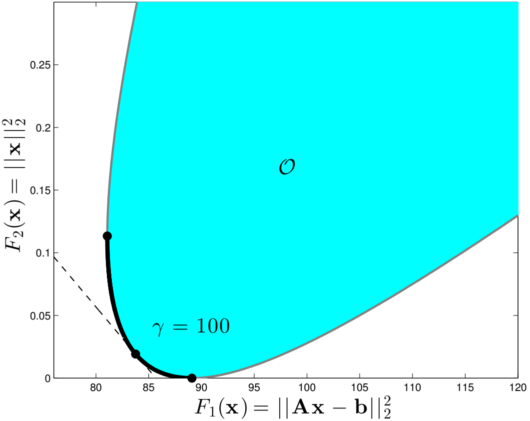
$$\begin{aligned} \boldsymbol{\lambda}^T \mathbf{f}_0(\mathbf{x}) &= \lambda_1 ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \lambda_2 ||\mathbf{x}||_2^2 \ , \\ & \mathbf{x}^T \left(\lambda_1 \mathbf{A}^T \mathbf{A} + \lambda_2 \mathbf{I}\right) \mathbf{x} - 2\lambda_1 \mathbf{b}^T \mathbf{A}\mathbf{x} + \lambda_1 \mathbf{b}^T \mathbf{b} \, . \end{aligned}$$

The first expression for the scalarized objective shows that this is a form of *Tikhonov regularization*. Defining $\gamma \triangleq \lambda_2/\lambda_1$ as the relative weight assigned to F_2 compared to F_1, the optimizing value of $\mathbf{x}$ is given by

$$\mathbf{x}^\star(\gamma) = \left(\mathbf{A}^T \mathbf{A} + \gamma \mathbf{I}\right)^{-1} \mathbf{A}^T \mathbf{b} \, .$$

Regularized Least-Squares (Continued)

For example problem data $\mathbf{A} \in \mathbb{R}^{100\times 10}$ and $\mathbf{b} \in \mathbb{R}^{100}$, we have the following set of achievable values $\left(||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2, ||\mathbf{x}||_2^2\right)$ and optimal trade-off curve.



The shaded set is the set of achievable values, while the thick line is the optimal trade-off curve formed by Pareto optimal points.

Risk-Return Trade-Off in Portfolio Optimization

Bi-criterion expression of Markowitz portfolio optimization problem

In economics, the classical *Markowitz portfolio optimization problem* consists of minimizing the *return variance* or *risk*, subject to a minimum acceptable mean return. Assuming that a portfolio budget constraint and no-shorting constraints are in effect, this becomes the QP:

$$\begin{array}{ll} \text{minimize} & \mathbf{x}^T \mathbf{\Sigma} \mathbf{x} \\ \text{subject to} & \overline{\mathbf{p}}^T \mathbf{x} \geq r_{\min} \, , \; \mathbf{1}^T \mathbf{x} = 1 \, , \; \mathbf{x} \succeq \mathbf{0} \end{array} .$$

- $\mathbf{x} \in \mathbb{R}^n$ is the investment portfolio, where x_i is the fraction invested in asset i.
- $\mathbf{p} \in \mathbb{R}^n$ is the vector of relative asset price changes, modeled as a random vector with mean $\overline{\mathbf{p}}$ and covariance $\mathbf{\Sigma}$.
- $r \triangleq \mathbf{p}^T \mathbf{x} \in \mathbb{R}$ is the return, with a mean return of $\overline{\mathbf{p}}^T \mathbf{x}$ and return variance (risk) of $\mathbf{x}^T \mathbf{\Sigma} \mathbf{x}$.
- $r_{\min} \in \mathbb{R}$ is the minimum acceptable mean return.

By varying $r_{\min}$, this is naturally equivalent to the bi-criterion problem

$$\begin{array}{ll} \text{minimize (with respect to } \mathbb{R}^2_+) & \left(-\overline{\mathbf{p}}^T \mathbf{x}, \mathbf{x}^T \mathbf{\Sigma} \mathbf{x} \right) \\ \text{subject to} & \mathbf{1}^T \mathbf{x} = 1 \, , \; \mathbf{x} \succeq \mathbf{0} \end{array} .$$

Through scalarization, this in turn becomes the problem

$$\begin{array}{ll} \text{minimize} & -\lambda_1 \overline{\mathbf{p}}^T \mathbf{x} + \lambda_2 \mathbf{x}^T \mathbf{\Sigma} \mathbf{x} \\ \text{subject to} & \mathbf{1}^T \mathbf{x} = 1 \, , \; \mathbf{x} \succeq \mathbf{0} \end{array} \iff \begin{array}{ll} \text{minimize} & -\overline{\mathbf{p}}^T \mathbf{x} + \gamma \mathbf{x}^T \mathbf{\Sigma} \mathbf{x} \\ \text{subject to} & \mathbf{1}^T \mathbf{x} = 1 \, , \; \mathbf{x} \succeq \mathbf{0} \end{array} ,$$

where $\gamma \triangleq \lambda_2 / \lambda_1 > 0$ is the exchange rate between mean return and risk. For fixed γ, this problem is a QP.

Risk-Return Trade-Off in Portfolio Optimization

Optimal risk-return trade-off curve and corresponding optimal allocations

Consider a problem consisting of 4 assets with statistics as follows.

Asset	$\overline{p}_i$	$\Sigma_{i,i}^{1/2}$
1	12%	20%
2	10%	10%
3	7%	5%
4	3%	0%

- $\rho_{1,2} = 30\%$
- $\rho_{1,3} = -40\%$
- $\rho_{2,3} = 0\%$

Here, $\rho_{k,\ell}$ is the correlation coefficient between p_k and p_ℓ given by

$$\rho_{k,\ell} = \Sigma_{k,\ell} / \left(\Sigma_{k,k}^{1/2} \Sigma_{\ell,\ell}^{1/2} \right).$$

The optimal risk-return trade-off curve and corresponding portfolio allocations are as follows.