More Advanced Topics

Stephen Boyd Steven Diamond
Junzi Zhang Akshay Agrawal

EE & CS Departments

Stanford University

Outline

- Nonconvex Optimization Methods
  - Difference of convex and multi-convex programming
  - Quasiconvex programming
- Formulating convex problems (wisely)
  - Convex formulation from modeling
  - Convexifying nonconvex problems
- Miscellaneous topics on algorithms and solvers

Outline

Outline

- Nonconvex Optimization Methods
  - Difference of convex and multi-convex programming
  - Quasiconvex programming
- Formulating convex problems (wisely)
  - Convex formulation from modeling
  - Convexifying nonconvex problems
- Miscellaneous topics on algorithms and solvers

Methods for nonconvex optimization problems

- **convex optimization methods** are (roughly) always global, always fast
- for general nonconvex problems, we have to give up one
  - **local optimization methods** are fast, but need not find global solution (and even when they do, cannot certify it)
  - **global optimization methods** find global solution (and certify it), but are not always fast (indeed, are often slow)
- **in this lecture**: local optimization methods that are based on solving a sequence of convex problems

Outline

- Nonconvex Optimization Methods
  - Difference of convex and multi-convex programming
  - Quasiconvex programming
- Formulating convex problems (wisely)
  - Convex formulation from modeling
  - Convexifying nonconvex problems
- Miscellaneous topics on algorithms and solvers

Difference of convex programming

- express problem as

$$\begin{array}{ll} \text{minimize} & f_0(x) - g_0(x) \\ \text{subject to} & f_i(x) - g_i(x) \leq 0, \quad i = 1, \dots, m \end{array}$$

  where f_i and g_i are convex
- $f_i - g_i$ are called difference of convex functions
- problem is sometimes called difference of convex programming

Convex-concave procedure

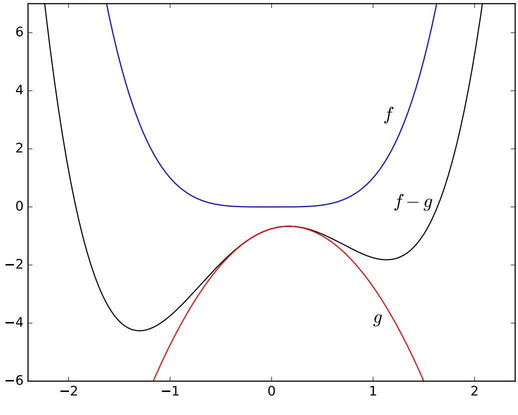
- iterative method for difference of convex programming
- obvious convexification at $x^{(k)}$: replace $f(x) - g(x)$ with

$$\hat{f}(x) = f(x) - g(x^{(k)}) - \nabla g(x^{(k)})^T (x - x^{(k)})$$

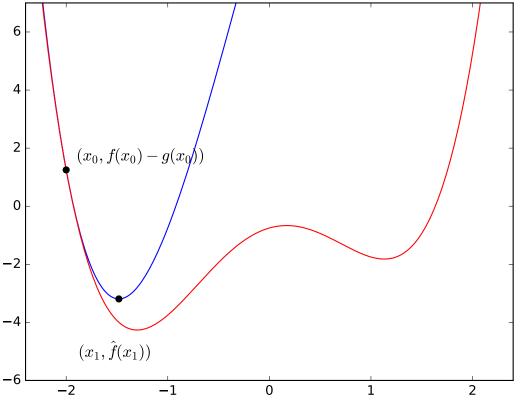
  - true objective at $\tilde{x}$ is better than convexified objective
  - true feasible set contains feasible set for convexified problem
- solve the convexified problem to get $x^{(k+1)}$ and repeat

Example

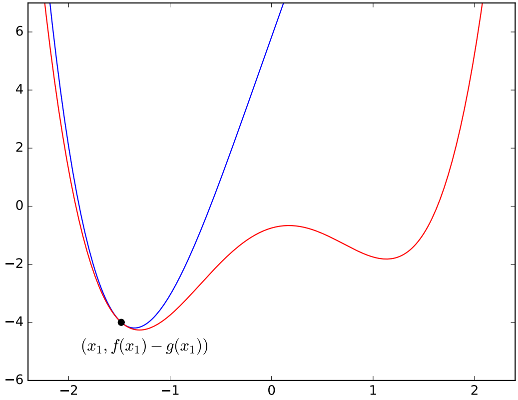
- unconstrained optimization on **R**



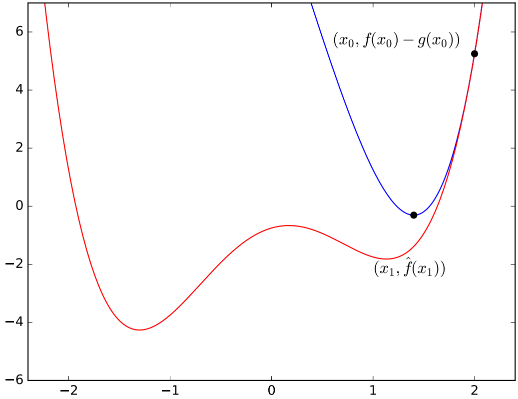
Example



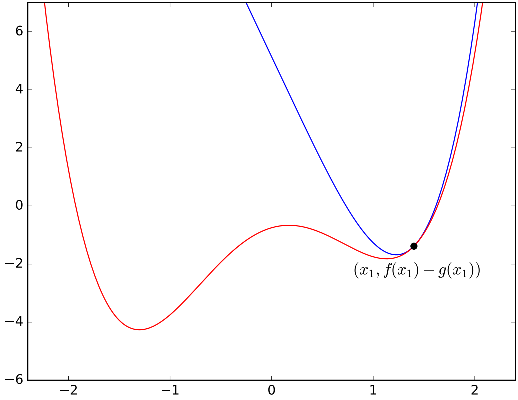
Example



Example



Example



Outline

Nonconvex Optimization Methods

- Difference of convex and multi-convex programming
- Quasiconvex programming

Formulating convex problems (wisely)

- Convex formulation from modeling
- Convexifying nonconvex problems

Miscellaneous topics on algorithms and solvers

Multi-convex programming

- given nonconvex problem with variable $(x_1, \ldots, x_n) \in \mathbf{R}^n$
- $\mathcal{I}_1, \ldots, \mathcal{I}_k \subset \{1, \ldots, n\}$ are index subsets with $\bigcup_j \mathcal{I}_j = \{1, \ldots, n\}$
- suppose problem is convex in subset of variables x_i, $i \in \mathcal{I}_j$, when x_i, $i \notin \mathcal{I}_j$ are fixed
- alternating convex optimization method: cycle through j, in each step optimizing over variables x_i, $i \in \mathcal{I}_j$
- special case: bi-convex problem
  - $x = (u, v)$; problem is convex in u (v) with v (u) fixed
  - alternate optimizing over u and v

Nonnegative matrix factorization

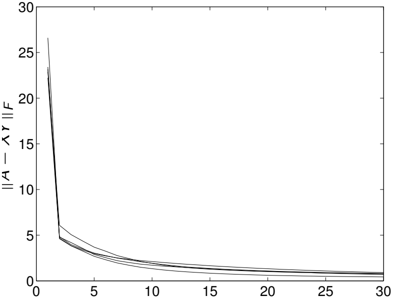
- NMF problem:

$$\begin{array}{ll} \text{minimize} & \|A - XY\|_F \\ \text{subject to} & X_{ij},\ Y_{ij} \geq 0 \end{array}$$

  variables $X \in \mathbf{R}^{m \times k}$, $Y \in \mathbf{R}^{k \times n}$, data $A \in \mathbf{R}^{m \times n}$
- difficult problem, except for a few special cases (*e.g.*, $k = 1$)
- alternating convex optimation: solve QPs to optimize over X, then Y, then X ...

Example

- convergence for example with $m = n = 50$, $k = 5$ (five starting points)



Outline

Nonconvex Optimization Methods

- Difference of convex and multi-convex programming
- Quasiconvex programming

Formulating convex problems (wisely)

- Convex formulation from modeling
- Convexifying nonconvex problems

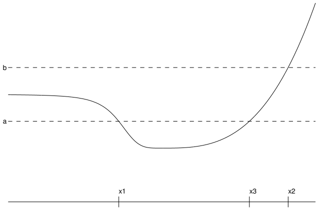
Miscellaneous topics on algorithms and solvers

Quasiconvex programming

Quasiconvex functions $f : \mathbf{R}^n \to \mathbf{R}$ is quasiconvex if $\mathbf{dom}\, f$ is convex and the sublevel sets

$$S_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}$$

are convex for all α



- f is quasiconcave if $-f$ is quasiconvex
- f is quasilinear if it is quasiconvex and quasiconcave

Quasiconvex programming

Examples

- $\sqrt{|x|}$ is quasiconvex on $\mathbf{R}$
- $\operatorname{ceil}(x) = \inf\{z \in \mathbf{Z} \mid z \geq x\}$ is quasilinear
- $\log x$ is quasilinear on $\mathbf{R}_{++}$
- $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbf{R}_{++}^2$
- linear-fractional function

  $$f(x) = \frac{a^T x + b}{c^T x + d}, \qquad \mathbf{dom}\, f = \{x \mid c^T x + d > 0\}$$

  is quasilinear
- distance ratio

  $$f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}, \qquad \mathbf{dom}\, f = \{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$$

  is quasiconvex

Quasiconvex programming

Internal rate of return

- cash flow $x = (x_0, \dots, x_n)$; x_i is payment in period i (to us if $x_i > 0$)
- we assume $x_0 < 0$ and $x_0 + x_1 + \cdots + x_n > 0$
- present value of cash flow x, for interest rate r:

$$\mathrm{PV}(x, r) = \sum_{i=0}^{n} (1 + r)^{-i} x_i$$

- internal rate of return is smallest interest rate for which $\mathrm{PV}(x, r) = 0$:

$$\mathrm{IRR}(x) = \inf\{r \geq 0 \mid \mathrm{PV}(x, r) = 0\}$$

Quasiconvex programming

Internal rate of return

- internal rate of return is smallest interest rate for which $\mathrm{PV}(x, r) = 0$:

$$\mathrm{IRR}(x) = \inf\{r \geq 0 \mid \mathrm{PV}(x, r) = 0\}$$

IRR is quasiconcave: superlevel set is intersection of open halfspaces

$$\mathrm{IRR}(x) \geq R \iff \sum_{i=0}^{n} (1+r)^{-i} x_i > 0 \text{ for } 0 \leq r < R$$

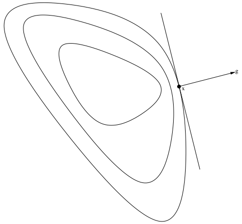
Quasiconvex programming

Properties modified Jensen inequality: for quasiconvex f

$$0 \le \theta \le 1 \quad \Longrightarrow \quad f(\theta x + (1-\theta)y) \le \max\{f(x), f(y)\}$$

first-order condition: differentiable f with cvx domain is quasiconvex iff

$$f(y) \le f(x) \quad \Longrightarrow \quad \nabla f(x)^T (y - x) \le 0$$



sums of quasiconvex functions are not necessarily quasiconvex

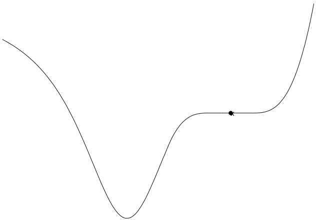
Quasiconvex programming

Problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

with $f_0 : \mathbf{R}^n \to \mathbf{R}$ quasiconvex, f_1, ..., f_m convex

can have locally optimal points that are not (globally) optimal



Quasiconvex programming

Convex representation of sublevel sets of f_0 if f_0 is quasiconvex, there exists a family of functions ϕ_t such that:

- $\phi_t(x)$ is convex in x for fixed t
- t-sublevel set of f_0 is 0-sublevel set of ϕ_t, *i.e.*,

$$f_0(x) \leq t \quad \Longleftrightarrow \quad \phi_t(x) \leq 0$$

example

$$f_0(x) = \frac{p(x)}{q(x)}$$

with p convex, q concave, and $p(x) \geq 0$, $q(x) > 0$ on **dom** f_0

can take $\phi_t(x) = p(x) - tq(x)$:

- for $t \geq 0$, ϕ_t convex in x
- $p(x)/q(x) \leq t$ if and only if $\phi_t(x) \leq 0$

Quasiconvex programming

Quasiconvex OPT via convex feasibility problems

$$\phi_t(x) \leq 0, \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m, \qquad Ax = b \tag{1}$$

- for fixed t, a convex feasibility problem in x
- if feasible, we can conclude that $t \geq p^\star$; if infeasible, $t \leq p^\star$

Bisection method for quasiconvex optimization

given $l \leq p^\star$, $u \geq p^\star$, tolerance $\epsilon > 0$.

repeat

1. $t := (l + u)/2$.
2. Solve the convex feasibility problem (1).
3. **if** (1) is feasible, $u := t$; **else** $l := t$.

until $u - l \leq \epsilon$.

Quasiconvex programming

Quasiconvex OPT via convex feasibility problems

Bisection method for quasiconvex optimization

given $l \le p^\star$, $u \ge p^\star$, tolerance $\epsilon > 0$.

repeat

1. $t := (l + u)/2$.
2. Solve the convex feasibility problem (1).
3. **if** (1) is feasible, $u := t$; **else** $l := t$.

until $u - l \le \epsilon$.

---

requires exactly $\lceil \log_2((u - l)/\epsilon) \rceil$ iterations (where u, l are initial values).

- Choose u and l: if infeasible for $t = u$, then $l = u$, $u = 2u$. If feasible for $t = l$, then $u = l$, $l = l/2$. Otherwise, start use current u and l.

Summary

- nonconvex problems are generally intractable
- these are heuristics with no optimality guarantee
  - but often works *very well* in practice
- CVXPY plugins are in the works
  - DCCP: difference of convex programming, solved via *convex-concave procedure*

Summary

- nonconvex problems are generally intractable
- these are heuristics with no optimality guarantee
  - but often works *very well* in practice
- CVXPY plugins are in the works
  - DCCP: difference of convex programming, solved via *convex-concave procedure*
  - DMCP: multi-convex optimization, solved via *block coordinate descent*

Summary

- nonconvex problems are generally intractable
- these are heuristics with no optimality guarantee
  - but often works *very well* in practice
- CVXPY plugins are in the works
  - DCCP: difference of convex programming, solved via *convex-concave procedure*
  - DMCP: multi-convex optimization, solved via *block coordinate descent*
  - QCQP: nonconvex QCQP (quadratically constrained quadratic programming) via *suggest and improve*

Summary

- nonconvex problems are generally intractable
- these are heuristics with no optimality guarantee
  - but often works *very well* in practice
- CVXPY plugins are in the works
  - DCCP: difference of convex programming, solved via *convex-concave procedure*
  - DMCP: multi-convex optimization, solved via *block coordinate descent*
  - QCQP: nonconvex QCQP (quadratically constrained quadratic programming) via *suggest and improve*
  - NCVX: mostly convex apart from decision variables from a non-convex set, solved via *NC-ADMM or relax-round-polish*
- main idea: automatically recognize the specific nonconvexity pattern and apply appropriate heuristics

Outline

Outline

Outline

- Nonconvex Optimization Methods
  - Difference of convex and multi-convex programming
  - Quasiconvex programming
- Formulating convex problems (wisely)
  - Convex formulation from modeling
  - Convexifying nonconvex problems
- Miscellaneous topics on algorithms and solvers

Bandlimited signal recovery from zero-crossings

Let $y \in \mathbf{R}^n$ denote a bandlimited signal $(t = 1, \ldots, n)$:

$$y_t = \sum_{j=1}^{B} a_j \cos\left(\frac{2\pi}{n}(f_{\min} + j - 1)t\right) + b_j \sin\left(\frac{2\pi}{n}(f_{\min} + j - 1)t\right).$$

Given: $f_{\min}$ the lowest frequency in the band, B the bandwidth, and the signs of y, *i.e.*, $s = \mathrm{sign}(y)$, with $s_t = 1$ if $y_t \geq 0$ and $s_t = -1$ otherwise.

Unknowns: the coefficients $a,\ b \in \mathbf{R}^B$ and the signal $y \in \mathbf{R}^n$.

Goal: find y and $a,\ b$ that minimizes $\|y\|_2$, and are consistent with the bandlimited assumption above, the signs and a normalization constraint $\|y\|_1 = n$ (as positive scaling does not change signs).

Bandlimited signal recovery from zero-crossings

Let $y \in \mathbf{R}^n$ denote a bandlimited signal $(t = 1, \ldots, n)$:

$$y_t = \sum_{j=1}^{B} a_j \cos\left(\frac{2\pi}{n}(f_{\min} + j - 1)t\right) + b_j \sin\left(\frac{2\pi}{n}(f_{\min} + j - 1)t\right).$$

Given: $f_{\min}$ the lowest frequency in the band, B the bandwidth, and the signs of y, *i.e.*, $s = \mathrm{sign}(y)$, with $s_t = 1$ if $y_t \geq 0$ and $s_t = -1$ otherwise.

Solution:

- bandlimited assumption: $\hat{y} = Ax$, $A = [C \;\; S]$, $x = (a, b)$.
  $C_{tj} = \cos(2\pi(f_{\min} + j - 1)t/n)$,
  $S_{tj} = \sin(2\pi(f_{\min} + j - 1)t/n)$.
- sign consistency: $s_t a_t^T x \geq 0$.
- normalization: $\|\hat{y}\|_1 = s^T A x = n$.

Bandlimited signal recovery from zero-crossings

Let $y \in \mathbf{R}^n$ denote a bandlimited signal $(t = 1, \ldots, n)$:

$$y_t = \sum_{j=1}^{B} a_j \cos\left(\frac{2\pi}{n}(f_{\min} + j - 1)t\right) + b_j \sin\left(\frac{2\pi}{n}(f_{\min} + j - 1)t\right).$$

Given: $f_{\min}$ the lowest frequency in the band, B the bandwidth, and the signs of y, *i.e.*, $s = \mathrm{sign}(y)$, with $s_t = 1$ if $y_t \geq 0$ and $s_t = -1$ otherwise.

Solution:

- We finally arrive at:

$$\begin{array}{ll} \text{minimize} & \|Ax\|_2 \\ \text{subject to} & s_t a_t^T x \geq 0, \quad t = 1, \ldots, n \\ & s^T A x = n. \end{array}$$

Matrix equilibration

We say that a matrix is ℓ_p equilibrated if each of its rows has the same ℓ_p norm, and each of its columns has the same ℓ_p norm.

Goal: given matrix $A \in \mathbf{R}^{m \times n}$, find diagonal invertible matrices $D \in \mathbf{R}^{m \times m}$ and $E \in \mathbf{R}^{n \times n}$ such that DAE is ℓ_p equilibrated.

Naive feasibility problem: find D, E, and two real numbers ν and ω, s.t.

$$\mathbf{1} D^p B E^p = -\nu \mathbf{1}^T, \quad D^p B E^p \mathbf{1} = -\omega \mathbf{1}.$$

Here $B_{ij} = |A_{ij}|^p$. Nonconvex!

Matrix equilibration

Naive feasibility problem: find D, E, and two real numbers ν and ω, s.t.

$$\mathbf{1} D^p B E^p = -\nu \mathbf{1}^T, \quad D^p B E^p \mathbf{1} = -\omega \mathbf{1}.$$

Here $B_{ij} = |A_{ij}|^p$.

- **Solution**: find an convex optimization problem with the feasibility problem as its KKT/optimality conditions.

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \sum_{j=1}^n B_{ij} e^{u_i + v_j} \\ \text{subject to} & \mathbf{1}^T u = 0, \quad \mathbf{1}^T v = 0. \end{array}$$

- Then $D = \mathbf{diag}(e^{u/p})$, $E = \mathbf{diag}(e^{v/p})$.

Matrix equilibration

- **Solution**: find an convex optimization problem with the feasibility problem as its KKT/optimality conditions.

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \sum_{j=1}^n B_{ij} e^{u_i + v_j} \\ \text{subject to} & \mathbf{1}^T u = 0, \quad \mathbf{1}^T v = 0. \end{array}$$

- Then $D = \mathbf{diag}(e^{u/p})$, $E = \mathbf{diag}(e^{v/p})$.
- Optimality conditions (ν, ω are multipliers of the constraints $\mathbf{1}^T u = 0$ and $\mathbf{1}^T v = 0$, resp.):

$$\sum_{j=1}^n B_{ij} e^{u_i + v_j} + \nu = 0, \quad i = 1, \ldots, m,$$

$$\sum_{i=1}^m B_{ij} e^{u_i + v_j} + \omega = 0, \quad j = 1, \ldots, n.$$

Outline

- Nonconvex Optimization Methods
  - Difference of convex and multi-convex programming
  - Quasiconvex programming
- Formulating convex problems (wisely)
  - Convex formulation from modeling
  - Convexifying nonconvex problems
- Miscellaneous topics on algorithms and solvers

Linear-fractional program

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

linear-fractional program

$$f_0(x) = \frac{c^T x + d}{e^T x + f}, \qquad \mathbf{dom}\, f_0(x) = \{x \mid e^T x + f > 0\}$$

- a quasiconvex optimization problem; can be solved by bisection

Linear-fractional program

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & Gx \preceq h \\
& Ax = b
\end{array}
$$

linear-fractional program

$$
f_0(x) = \frac{c^T x + d}{e^T x + f}, \qquad \textbf{dom}\, f_0(x) = \{x \mid e^T x + f > 0\}
$$

- also equivalent to the LP (variables y, z)

$$
\begin{array}{ll}
\text{minimize} & c^T y + dz \\
\text{subject to} & Gy \preceq hz \\
& Ay = bz \\
& e^T y + fz = 1 \\
& z \geq 0
\end{array}
$$

Linear-fractional program

Proof sketch of equivalence

$$
\begin{array}{ll}
\text{minimize} & f_0(x) = \frac{c^T x + d}{e^T x + f} \\
\text{subject to} & Gx \preceq h, \quad Ax = b
\end{array}
$$

$$
\begin{array}{ll}
\text{minimize} & c^T y + dz \\
\text{subject to} & Gy \preceq hz, \quad Ay = bz, \\
& e^T y + fz = 1, \quad z \geq 0
\end{array}
$$

- $y = x/(e^T x + f)$, $z = 1/(e^T x + f)$.
- $x = y/z$ if $z \neq 0$. Otherwise, consider $x = x_0 + ty$, then $f_0(x) \to c^T y + dz$.

Covariance estimation for Gaussian random variables

Let $y \in \mathcal{N}(0, \Sigma)$ ($y \in \mathbf{R}^n$), *i.e.*, $\mathbf{E}[yy^T] = \Sigma$. Then the density is

$$p_\Sigma(y) = (2\pi)^{-n/2} \det(R)^{-1/2} \exp(-y^T \Sigma y/2).$$

For samples $y_1, \ldots, y_m$, the negative log-likelihood function is

$$l(\Sigma) = (mn/2)\log(2\pi) + (m/2)\log\det\Sigma + (m/2)\mathbf{tr}(\Sigma^{-1}Y),$$

where $Y = \frac{1}{m}\sum_{k=1}^m y_k y_k^T$. Nonconvex!

Covariance estimation for Gaussian random variables

For samples $y_1, \ldots, y_m$, the negative log-likelihood function is

$$l(\Sigma) = (mn/2)\log(2\pi) + (m/2)\log\det\Sigma + (m/2)\mathbf{tr}(\Sigma^{-1}Y),$$

where $Y = \frac{1}{m}\sum_{k=1}^{m} y_k y_k^T$. Nonconvex!

Solution: change of variable to $S = \Sigma^{-1}$.

$$\tilde{l}(S) = (mn/2)\log(2\pi) - (m/2)\log\det S + (m/2)\mathbf{tr}(SY).$$

Now convex!

Maximum Sharpe ratio portfolio

Consider the following problem:

$$\begin{array}{ll} \text{minimize} & \mu^T x / \|\Sigma^{1/2} x\|_2 \\ \text{subject to} & \mathbf{1}^T x = 1, \quad \|x\|_1 \leq L^{\max}, \end{array}$$

where μ is the mean return, $\Sigma \succ 0$ is the return covariance, and $L^{\max}$ is the leverage limit. Assume that $\exists x$, s.t. $\mu^T x > 0$.

- This is **quasi-convex** – but can we do better?

Maximum Sharpe ratio portfolio

Consider the following problem:

$$\begin{array}{ll} \text{minimize} & \mu^T x / \|\Sigma^{1/2} x\|_2 \\ \text{subject to} & \mathbf{1}^T x = 1, \quad \|x\|_1 \leq L^{\max}, \end{array}$$

where μ is the mean return, $\Sigma \succ 0$ is the return covariance, and $L^{\max}$ is the leverage limit. Assume that $\exists x$, s.t. $\mu^T x > 0$.

- This is **quasi-convex** – but can we do better?
- Yes – via homogeneity in x of the objective function.

Maximum Sharpe ratio portfolio

Consider the following problem:

$$\begin{array}{ll} \text{maximize} & \mu^T x / \|\Sigma^{1/2} x\|_2 \\ \text{subject to} & \mathbf{1}^T x = 1, \quad \|x\|_1 \leq L^{\max}, \end{array}$$

- First step: rewrite leverage constraint as $\|x\|_1 \leq L^{\max} 1^T x$, and add redundant constraint $\mu^T x > 0$ – homogeneous.

$$\begin{array}{ll} \text{maximize} & \mu^T x / \|\Sigma^{1/2} x\|_2 \\ \text{subject to} & \mathbf{1}^T x = 1, \quad \|x\|_1 \leq L^{\max} 1^T x, \quad \mu^T x > 0. \end{array}$$

Maximum Sharpe ratio portfolio

- First step: rewrite leverage constraint as $\|x\|_1 \leq L^{\max}\mathbf{1}^T x$, and add redundant constraint $\mu^T x > 0$ – homogeneous.

$$\begin{array}{ll} \text{maximize} & \mu^T x / \|\Sigma^{1/2} x\|_2 \\ \text{subject to} & \mathbf{1}^T x = 1, \quad \|x\|_1 \leq L^{\max} 1^T x, \quad \mu^T x > 0. \end{array}$$

- Second step: change of variables $z = x/\mu^T x \Rightarrow \mu^T z = 1 \Rightarrow x = z/\mathbf{1}^T z$.

$$\begin{array}{ll} \text{maximize} & 1/\|\Sigma^{1/2} z\|_2 \\ \text{subject to} & \mu^T z = 1, \quad \|z\|_1 \leq L^{\max}\mathbf{1}^T z. \end{array}$$

Maximum Sharpe ratio portfolio

Consider the following problem:

$$\begin{array}{ll} \text{maximize} & \mu^T x / \|\Sigma^{1/2} x\|_2 \\ \text{subject to} & \mathbf{1}^T x = 1, \quad \|x\|_1 \leq L^{\max}, \end{array}$$

- Finally **convex**!

$$\begin{array}{ll} \text{minimize} & \|\Sigma^{1/2} z\|_2 \\ \text{subject to} & \mu^T z = 1, \quad \|z\|_1 \leq L^{\max} \mathbf{1}^T z. \end{array}$$

General convexification procedures

- transformation (change of variables)
- convex relaxation
- convex restriction

Outline

Outline

Nonconvex Optimization Methods

- Difference of convex and multi-convex programming
- Quasiconvex programming

Formulating convex problems (wisely)

- Convex formulation from modeling
- Convexifying nonconvex problems

Miscellaneous topics on algorithms and solvers

Algorithm design

- sub-differential/sub-gradient and proximal operators
- monotone operators
- first-order methods, quasi-Newton methods, Newton methods/interior point methods
- primal-dual methods, distributed optimization
- stochastic and online algorithms

Modeling language and solver choices

- Clarification: CVXPY is not a solver, but a modeling language
- How to choose solver: choose the most specialized solver whenever possible – automatically done in CVXPY 1.0, and keep improving

Questions?

Q&A time now!