Convex and Robust Optimization, with Applications in Finance

Miguel Sousa Lobo
Information Systems Lab., Stanford University

thanks: Stephen Boyd, Darrell Duffie, etc.

Apr/11/00 – EESOR Colloquium

Outline

1. **Convex optimization, LP, SOCP, SDP**

2. Robust optimization ($e.g.$, robust LP as an SOCP)

3. Worst-case variance of a portfolio, robust portfolio design

Mathematical optimization

– general mathematical programming problem, $x \in \mathbf{R}^n$:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, L \\ & h_i(x) = 0, \quad i = 1, \ldots, M \end{array}$$

– a convex program (in standard form) if $f, \; g_1, \; \ldots, \; g_L$ are convex, and $h_1, \; \ldots, \; h_M$ are affine

– either case has very wide application, engineering, OR, finance, etc.

– not as widely used as could be, somewhat of a bad name:
  – reliability of methods (to be used unsupervised, or by non-specialists)
  – only as good as model (and often worse)

LS, LP, QP

– reliability first solved for some convex problems (late 40's):

least-squares, linear programming, (convex) quadratic programming

– have a global optimum

– any (well designed) solver will find the same solution,
and work everytime w/o tuning

– by contrast, non-linear programming seen as art:
different solvers produce different answers,
don't work reliably unsupervised,
don't scale well with problem size

Convex optimization

- LP: from simplex to interior-point (IP) methods (Karmarkar 84)
  - *simplex:* sequence of points on vertices of polyhedral constraint
  - *interior-point:* sequence of points in interior of polyhedron
- from **linear vs. non-linear** to **convex vs. non-convex** (Nesterov and Nemirovski 94)
  - IP methods extend naturally to (many) non-linear convex programs
  - polynomial complexity (complete complexity theory, not the case for simplex)

Non-linear convex programs

- two classes for which there are particularly effective methods:
  - second-order cone, or conic quadratic programming (SOCP)
  - semidefinite programming (SDP)

- both have constraint set defined by intersection of hyperplane with self-dual convex cone (same goes for LP)

- algorithms:
  - are effective, reliable enough for unsupervised use
  - scale well with problem size ($\approx O(n^3)$)

LP, SDP, SOCP

- linear program (LP)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, L \end{array}$$

- semidefinite program (SDP)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & A_0 + A_1 x_1 + \cdots + A_n x_n \succeq 0, \end{array}$$

where $A_i \in \mathbf{R}^{m \times m}$,
and $A \succeq 0$ denotes matrix inequality, *i.e.*, $z^T A z \geq 0, \forall z \in \mathbf{R}^m$

- second-order cone program (SOCP), or conic quadratic program (CQP)

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & \|A_i x + b_i\| \leq c_i^T x + d_i, \quad i = 1, \ldots, L \end{array}$$

  $A_i \in \mathbf{R}^{(n_i-1)\times n}$, optimization variable is $x \in \mathbf{R}^n$

- SOCP, SDP are convex, nonlinear, nondifferentiable problem
  - SOCP includes LP, QP, QCQP as special cases
  - SDP includes SOCP as special case

- many IP methods available

- implementations available (SP, SOCP, SDPPACK, SeDuMi, MOSEK), work ongoing in sparse implementations

Portfolio optimization

single-period, maximize expected revenue

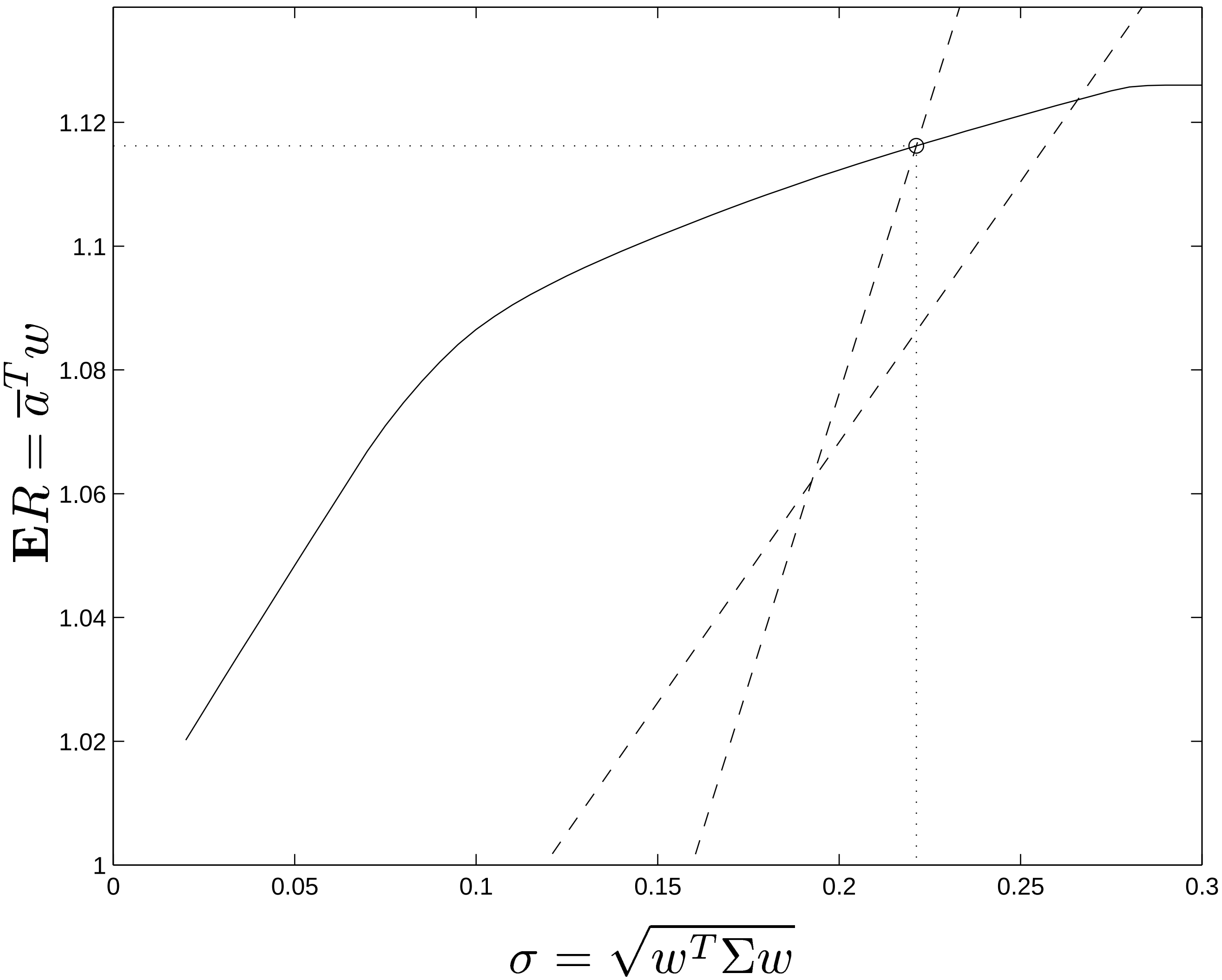
- convex problems (can be solved globally and efficiently)
  - linear transaction costs
  - constraints on: shorting, diversity, variance, shortfall risk
- fixed (plus linear) transaction costs is a non-convex problem
  - a hard combinatorial problem
  - with convex optimization, can get
    - upper bound on performance (convex relaxation)
    - approximate solution (iterative heuristic)
  - approximate solution often close to bound, can use branch-and-bound

Mean-variance portfolio optimization

$$
\begin{array}{ll}
\text{maximize} & \overline{a}^T w + \gamma w^T \Sigma w \\
\text{subject to} & \mathbf{1}^T w = 1 \\
 & w \geq 0
\end{array}
$$

where

w	$=$	holdings in each security (program variable),	$w \in \mathbf{R}^n$
$\overline{a}$	$=$	expected return on each security,	$\overline{a} \in \mathbf{R}^n$
Σ	$=$	covariance matrix of returns,	$\Sigma \in \mathbf{R}^{n \times n}$


1.12
1.1
1.08
1.06
1.04
1.02
1
0
0.05
0.1
0.15
0.2
0.25
0.3
$\mathbf{E}R = \bar{a}^T w$
$\sigma = \sqrt{w^T \Sigma w}$

Outline

1. Convex optimization, LP, SOCP, SDP
2. **Robust optimization (*e.g.*, robust LP as an SOCP)**
3. Worst-case variance of a portfolio, robust portfolio design

Robust optimization

- in practical problems, parameters are never exact
  (measurement and implementation errors, noise, modeling errors)
- optimization often has a way of "finding" these errors
- should solve robust version
  - minimize expected value over some parameter distribution
  - minimize worst-case over parameter uncertainty set
- robust versions of some common convex problems are still convex, and therefore tractable (Ben-Tal, Nemirovski, El Ghaoui, . . .)
- not true for all problems, sometimes robust version is numerically hard
  (to make tractable can do relaxations, *i.e.*, increase conservativeness)

Robust linear program with ellipsoidal uncertainty

- c and b_i fixed, a_i uncertain

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i \;\; \text{for all } a_i \in \mathcal{E}_i, \quad i = 1, \ldots, m \end{array}$$

- the uncertainty in a_i is described by an ellipsoid

$$\mathcal{E}_i = \{\overline{a}_i + P_i u \mid \|u\| \leq 1\}$$

with $P_i = P_i^T \geq 0$

comments:

– other approaches exist to deal with data uncertainty
($e.g.$, stochastic programming)

– robust optimization recovers nominal problem as uncertainty goes to zero

– different interpretations of robust optimization:

  – actual description of problem,
    hard constraints with bounded tolerance in parameters
    ($e.g.$, structural optimization)

  – "chance-constrained programming"

  – as a heuristic to reduce sensitivity of solution

– uncertainty set often from statistical considerations

– which problems are tractable? work in progress . . .
(for LP see Ben-Tal & Nemirovski, OR Letters, 1999)

– mathematical tractability vs. practical solvability

– does it work? good (but limited) evidence so far, mostly for LP

– same ideas have long been used (to very good effect) in robust control

– downside-risk (and mean-variance) portfolio optimization
can be re-interpreted as robust LP

Robust LP as an SOCP

- note that

$$\max\{\, a_i^T x \mid a_i \in \mathcal{E}_i \,\} = \overline{a}_i^T x + \|P_i x\|$$

  and therefore

$$a_i^T x \leq b_i \;\; \text{for all } a_i \in \mathcal{E}_i \quad \Longleftrightarrow \quad \overline{a}_i^T x + \|P_i x\| \leq b_i$$

- the robust LP can be expressed as the SOCP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \overline{a}_i^T x + \|P_i x\| \leq +b_i, \quad i = 1, \ldots, m \end{array}$$

- the additional norm terms act as 'regularization terms', discouraging large x in directions with considerable uncertainty in the parameters a_i

Probabilistic interpretation of robust LP

- a is an R.V. with Gaussian distribution, mean $\overline{a}$, variance Σ

  we want to solve (for $\pi \geq 0.5$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \text{Prob}\{a^T x \leq b\} \geq \pi \end{array}$$

  rewrite the constraint (Φ is the c.d.f. of standard Gaussian)

$$\text{Prob}\{a^T x \leq b\} \geq \pi \quad \iff \quad b - \overline{a}^T x \geq \Phi^{-1}(\pi)\|\Sigma^{1/2} x\|$$

- **downside-risk** approach to portfolio optimization is equivalent to robust counterpart of revenue maximization, $i.e.$, pre-Markovitz (add budget constraint $\mathbf{1}^T x = 1$, maximize $a^T x$, ...)

Using robust LP

- as many applications as there are LPs!
- coupled with efficient SOCP solvers, could (and should) be used in all LP solvers:
  - user specifies LP
  - user specifies parameter tolerances
  - if tolerances are zero, call LP solver
  - otherwise, generate SOCP and call SOCP solver

Robust least-squares

- "solve" overdetermined system $Ax \approx b$, with uncertainty in A
- rows of A lie within an ellipsoid: $a_i \in \mathcal{E}_i$, where

$$\mathcal{E}_i = \{\overline{a}_i + P_i u \mid \|u\| \leq 1\} \qquad (P_i = P_i^T > 0)$$

- problem: minimize the worst-case residual

$$\text{minimize} \quad \max_{a_i \in \mathcal{E}_i} \left(\sum_{i=1}^{n} \left(a_i^T x - b_i\right)^2 \right)^{1/2}$$

- work out the objective function in a closed form, and the problem can be formulated as

$$\text{minimize} \quad \left(\sum_{i=1}^{n} \left(\left| \overline{a}_i^T x - b_i \right| + \| P_i x \| \right)^2 \right)^{1/2}$$

which can be cast as the SOCP

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & \|t\| \leq s \\ & u_i + \|P_i x\| \leq t_i, \quad i = 1, \ldots, n \\ & \left| \overline{a}_i^T x - b_i \right| \leq u_i, \quad i = 1, \ldots, n \end{array}$$

- easily extended for uncertainty in b, can add other constraints

Outline

1. Convex optimization, LP, SOCP, SDP
2. Robust optimization ($e.g.$, robust LP as an SOCP)
3. **Worst-case variance of a portfolio, robust portfolio design**

Motivation

(Ju and Pearson, Journal of Risk, 1999)

– traders are rewarded for performance vs. exposure to risk

– penalty for portfolio large $w^T \Sigma w$ (where w are portfolio weights)

– $e.g.$: covariance matrix is 500 by 500, estimated with 250 data points

– covariance matrix estimate has a null-space!

– traders have an incentive to find portfolios with “zero variance”

– usual solutions: limit each trader to small number of assets,
impose diagonal-dominant structure on Σ

– a better solution?

Worst-case variance

- given uncertainty in the knowledge of Σ, how risky can a portfolio be?

$$\begin{array}{ll} \text{maximize} & w^T \Sigma\, w \\ \text{subject to} & \underline{\Sigma}_{ij} \leq \Sigma_{ij} \leq \overline{\Sigma}_{ij} \\ & \Sigma \succeq 0 \end{array}$$

$w \in \mathbf{R}^n$ is the portfolio (fixed and known)
$\Sigma \in \mathbf{R}^{n \times n}$ is the program variable

- linear objective, convex constraints; a *semi-definite program* (SDP)
- what other descriptions of the uncertainty in Σ are convex?

Constraints on the correlation coefficients

- constraint:

$$\underline{\rho}_{ij} \leq \rho_{ij} \leq \overline{\rho}_{ij} \qquad \Leftrightarrow \qquad \underline{\rho}_{ij}\sqrt{\Sigma_{ii}\Sigma_{jj}} \leq \Sigma_{ij} \leq \overline{\rho}_{ij}\sqrt{\Sigma_{ii}\Sigma_{jj}}$$

- linear if Σ_{ii} constant ($i.e.$, known exactly)

- convex if $\underline{\rho}_{ij} \leq 0$ and $\overline{\rho}_{ij} \geq 0$
  hyperbolic, can write as second-order cone constraint using:

$$\left\| \begin{bmatrix} \dfrac{2}{\overline{\rho}_{ij}}(\Sigma_{ij})_+ \\ \Sigma_{ii} - \Sigma_{jj} \end{bmatrix} \right\| \leq \Sigma_{ii} + \Sigma_{jj}$$

Wishart distribution and ellispsoidal constraint

– for $x_i \sim \mathcal{N}(0, \Sigma)$, the covariance estimate $\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} x_i x_i^T$ is Wishart distributed (and asymptotically normal for $N \to \infty$)

– second-moment of Wishart distribution is

$$Cov\left(vec(\hat{\Sigma})\right) = \frac{1}{N}(I + K)(\Sigma \otimes \Sigma)$$

where K is permutation matrix such that $K\, vec(C) = vec\left(C^T\right)$

– very large matrix ($\approx n^4$) but lots of structure
– use to form ellipsoidal uncertainty set for entries of $\hat{\Sigma}$
– use $\hat{\Sigma}$ for Σ? low-rank for small N still an issue

Robust design problem

- incorporate worst-case variance analysis in portfolio optimization
- goal is to select portfolio with highest expected return, subject to a given worst-case portfolio variance (plus budget constraint):

$$\begin{array}{ll} \text{maximize} & \bar{a}^T x \\ \text{subject to} & \max_{\Sigma \in \mathcal{S}} x^T \Sigma x \leq \sigma^2, \quad \mathbf{1}^T x \leq 1 \end{array}$$

- equivalent: find tradeoff curve by minimizing worst-case variance subject to constraint on return:

$$\begin{array}{ll} \text{minimize} & \max_{\Sigma \in \mathcal{S}} x^T \Sigma x \\ \text{subject to} & \bar{a}^T x \geq R, \quad \mathbf{1}^T x \leq 1 \end{array}$$

Min-max-max-min

- problem:

$$\min_{\bar{a}^T x \geq R,\ \mathbf{1}^T x \leq 1} \quad \max_{\Sigma \in \mathcal{S}} \quad x^T \Sigma\, x$$

- $x^T \Sigma\, x$ is convex in x, and linear (hence concave) in Σ, can switch min-max to max-min:

$$\max_{\Sigma \in \mathcal{S}} \quad \min_{\bar{a}^T x \geq R,\ \mathbf{1}^T x \leq 1} \quad x^T \Sigma\, x$$

- overall problem is convex, can do either min-max or max-min (pick the most convenient numerically)

Subgradient from primal-dual optimal point

- define function φ as the optimum objective:

$$\varphi(\Sigma) = \min_{\bar{a}^T x \geq R,\ \mathbf{1}^T x \leq 1} \ x^T \Sigma\, x$$

- compute $\varphi(\Sigma)$ by primal-dual interior-point method (a QP)

- subgradient comes for free: $\nabla_{\Sigma} \varphi = -x^*(x^*)^T$
  where x^* is the optimal x

- subgradient defines a cutting-plane, can ruleout half-space

Analytic centering, cutting-plane methods

problem:

$$\begin{array}{ll} \text{maximize} & \varphi(\Sigma) \\ \text{subject to} & \Sigma \in \mathcal{S} \end{array}$$

algorithm:

1. inner problem: compute φ, subgradient, cutting-plane
2. outer problem: find analytic center
   ($\Sigma \in \mathcal{S}$ plus linear constraint from cutting-plane)
3. repeat

inner problem is a QP (or SDP),
outer problem is similar to SDP (or QP),
in total, must solve N convex QPs and N semidefinite analytic centering problems (with N maybe in the range of 10 to 100?)

SDP dual

At least in some cases, can use SDP dual for simpler solution method:

$$\min_{\bar{a}^T x \geq R,\ \mathbf{1}^T x \leq 1} \quad \max_{\underline{\Sigma} \leq \Sigma \leq \overline{\Sigma},\ \Sigma \succeq 0} \quad x^T \Sigma\, x$$

using the dual of the max problem, get SDP:

$$\begin{array}{ll} \mbox{minimize} & \mathbf{Tr}(\overline{\Sigma}\,\overline{\Lambda} - \underline{\Sigma}\,\underline{\Lambda}) \\ \mbox{subject to} & \bar{a}^T x \geq R,\ \mathbf{1}^T x \leq 1,\ \overline{\Lambda} \geq 0,\ \underline{\Lambda} \geq 0 \\ & \left[\begin{array}{cc} -\overline{\Lambda} + \underline{\Lambda} & x \\ x^T & 1 \end{array} \right] \succeq 0 \end{array}$$

where the variables are $x \in \mathbf{R}^n$, and $\overline{\Lambda}, \underline{\Lambda} \in \mathbf{R}^{n \times n}$.

Other possible methods: iterated projections; barrier for max-min.

Extensions

- can handle uncertainty in expected returns $\bar{a}$
- can do the same for shorfall-risk (worst-case analysis & robust design)
- can do the same for robust ℓ_2 index tracking
  (ellipsoidal uncertainty in $\bar{a}$ can be handled through $\mathcal{S}$-procedure)

Review, conclusions

- convex programming, self-dual conic programs, SOCP
  - efficient methods for large, non-linear, non-differentiable problems
  - many applications, software available
- robust optimization
  - protect against errors in problem data
  - some robust optimization problems can be stated as convex programs
- portfolio selection, many convex problems, relaxations for non-convex
- significant uncertainty in data (expected returns, covariance matrix)
  - can compute worst-case risk (via SDP)
  - can do robust portfolio design (via cutting-plane)

Outline

1. Convex optimization, LP, SOCP, SDP

2. Robust optimization ($e.g.$, robust LP as an SOCP)

3. Worst-case variance of a portfolio, robust portfolio design

4. **Self-dual conic programs, interior-point methods**

Convex program, standard form

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& a_i^T x = b_i, \quad i = 1, \ldots, p
\end{array}
$$

where the optimization variable is $x \in \mathbf{R}^n$,

and the functions $f_0, \ f_1, \ \ldots, \ f_m$ are convex, that is:

$$\lambda\, f_i\,(x) + (1-\lambda)\, f_i\,(y) \geq f_i\,(\lambda\, x + (1-\lambda)\, y) \quad \text{for all} \ \ x, y \ \text{and} \ \lambda \in [0,1]$$

Self-dual conic program

- primal program

$$\begin{array}{ll} \text{minimize} & \langle c, x\rangle \\ \text{subject to} & A\,x = b,\ x \in \mathcal{K} \end{array}$$

- $\mathcal{K}$ is a convex cone if it is a convex set such that

$$x \in \mathcal{K} \quad \Rightarrow \quad \alpha x \in \mathcal{K},\ \forall \alpha \geq 0$$

- dual program

$$\begin{array}{ll} \text{maximize} & \langle b, z\rangle \\ \text{subject to} & c - A^* z \ \in\ \mathcal{K}^* \end{array}$$

- cone is self-dual if

$$\mathcal{K}^* = \{\ z \ \mid\ \langle z, x\rangle \geq 0 \text{ for all } x \in \mathcal{K}\} = \mathcal{K}$$

Non-negative orthant and semi-definite cone

- non-negative orthant (leads to LP)

$$\mathbf{R}^n_+ = \{\ x \ \mid x \in \mathbf{R}^n,\ x_i \geq 0,\ i = 1, \ldots, n\ \}$$

- positive semi-definite cone (leads to SDP)

$$\mathcal{K}_{\mathrm{PSD}} = \left\{\ X\ \middle|\ X \in \mathbf{R}^{n\times n},\ u^T X\, u \geq 0,\ \forall u \in \mathbf{R}^n\ \right\}$$

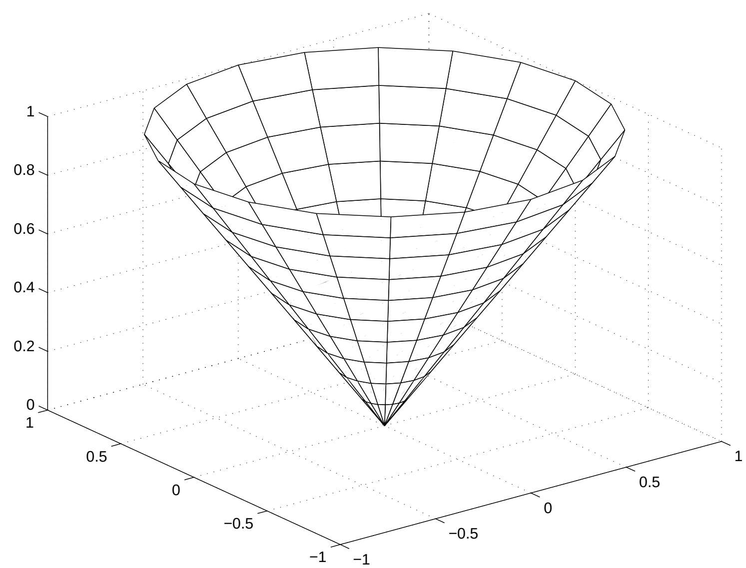
- both are self-dual

Second-order, or quadratic cone

standard second-order cone of dimension n

$$\mathcal{K}_{\mathrm{SO}} = \left\{ \left[\begin{array}{c} u \\ t \end{array} \right] \;\middle|\; u \in \mathbf{R}^{n-1},\ t \in \mathbf{R},\ \|u\| \leq t \right\}$$

(also quadratic, Lorentz, or ice-cream cone), self-dual



Duality

- *duality gap*: primal minus dual objective, $\eta\,(x, z) = \langle c, x \rangle - \langle b, z \rangle$

- *weak duality*: $\eta\,(x, z) = \langle x, c - A^* z \rangle \geq 0$, for any feasible x, z

- *strong duality*: if primal or dual is strictly feasible, $\eta\,(x^*, z^*) = 0$

- primal-dual methods: minimize gap
  - optimal gap known a priori: exploited for effective methods
  - if cone is self-dual ($\mathcal{K} = \mathcal{K}^*$), the dual program has same structure as primal program $\Rightarrow$ easy to write primal-dual methods

Primal-dual interior-point methods: key ideas

- barrier function:
  - bounded in interior of cone
  - goes to $+\infty$ at boundary
  - convex and smooth
  - self-concordant: third derivative bounded w.r.t. second derivative (guarantees effectiveness of Newton steps!)

- examples:
  - for LP, $\phi(x) = -\sum_i \log x_i$
  - for SOCP, $\phi(u,t) = -\log(t - \|u\|)$
  - for SDP, $\phi(X) = -\log\det X$

- can add barriers for constraints defined by different cones, and preserve properties

- iteration k, take Newton step over:

$$\phi(x) + \phi(z) + \alpha_k \eta(x, z)$$

- choose $\alpha_k = c/\eta(x_{k-1}, z_{k-1})$, same as minimizing **potential function**:

$$\psi(x, z) = \phi(x) + \phi(z) + c \log \eta(x, z)$$

- extra term $c \log \eta$:
  - concave, smooth
  - goes to $-\infty$ at optimum (and so does potential, for c large enough)