Optimization

4. Convex optimization: theory

Andrew Lesniewski

Baruch College
New York

Fall 2018

Outline

Why convex optimization?

- Optimization problems encountered in applications are typically nonlinear.
- As discussed in Lecture Note #1 and #2, there are no general methods to tackle nonlinear optimization problems, many of these problems are very difficult.
- The first line of attack for solving such problems is local optimization: we try to seek a point that is only locally optimal. For practical purposes, such a solution may be good enough.
- Many of these methods are fast and apply to wide varieties of situations.
- However, their efficiency is a function of the number of variables: a problem with 10 variables if often challenging, a problem with 100 variables may prove intractable.
- Tuning an optimization algorithm (adjusting the parameters of the algorithm, choosing an initial guess) is often more art than science.
- In many situations, turning to global optimization techniques is necessary. Global optimization is used for problems with a small number of variables, where computing time is not critical, and the value of finding the true global solution is high.

Why convex optimization?

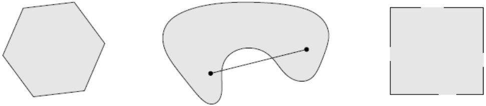
- *Convex optimization* offers a reasonable compromise between suitability of the mathematical formulation of the problem and its tractability.
- Formulating an optimization problem as a convex approximation problem may be a productive way to find a decent first solution.
- Starting with a hard nonconvex approximate problem, we may first try to find a convex approximation to the problem. By solving this approximate problem, which can be done easily and without an initial guess, we obtain the exact solution to the approximate convex problem. This point can then be used as the starting point for a local optimization method, applied to the original nonconvex problem.
- There are many effective methods for solving convex optimization problems. Key fact about these methods is that no information about the distance to the globally optimal solution is required.
- The message: if we can formulate an optimization problem as a convex optimization problem, then we can solve it efficiently.

Convex sets

- A set $C \subset \mathbb{R}^n$ is called *convex*, if for any $x, y \in C$, the line segment joining x and y is contained in C. In other words, if

$$\alpha x + (1 - \alpha) y \in C, \quad \text{for any } 0 \leq \alpha \leq 1. \tag{1}$$

- In other words, for any two points in a convex set set, the line segment connecting these points is contained in the set.
- The first of the sets below is convex, while the second and third are not.



Convex sets

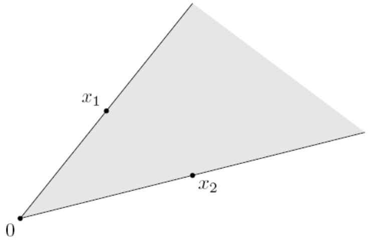
- Examples of convex sets:
  (i) Lines.
  (ii) Hyperplanes $\{x \in \mathbb{R}^n : a^{\mathrm{T}} x = b\}$.
  (iii) Halfspaces $\{x \in \mathbb{R}^n : a^{\mathrm{T}} x < b\}$.
  (iv) Polyhedra $\{x \in \mathbb{R}^n : a_i^{\mathrm{T}} x = b_i, i \in \mathcal{E}, \text{ and } a_i^{\mathrm{T}} x \leq b_i, i \in \mathcal{I}\}$.
  (v) Euclidean balls $\{x \in \mathbb{R}^n : \|x - x_0\| \leq r\}$, where $r > 0$. Indeed,

$$\begin{aligned}\|\alpha x + (1-\alpha)y - x_0\| &= \|\alpha(x - x_0) + (1-\alpha)(y - x_0)\| \\ &\leq \alpha\|x - x_0\| + (1-\alpha)\|y - x_0\| \\ &\leq \alpha r + (1-\alpha) r \\ &= r.\end{aligned}$$

  (vi) Ellipsoids $\{x \in \mathbb{R}^n : (x - x_0)A^{-1}(x - x_0) \leq 1\}$, where $A \in \mathrm{Mat}_n(\mathbb{R})$ is a positive definite matrix.

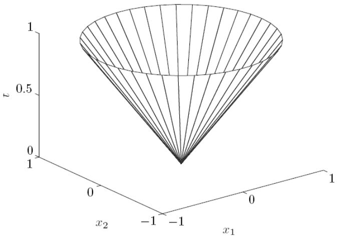
Cones

- A set $C \subset \mathbb{R}^n$ is called a *cone*, if for any $x \in C$ and $\theta > 0$, $\theta x \in C$.
- A set $C \subset \mathbb{R}^n$ is called a *convex cone*, if for any $x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$, $\theta_1 x_1 + \theta_2 x_2 \in C$. In other words, C is both a cone and a convex set.
- Here is a graphical representation of this condition:



Examples of cones

- Let $\|x\|$ be a norm on $\mathbb{R}^n$. The set $C = \{(x, t) \in \mathbb{R}^n \times \mathbb{R}_+ : \|x\| \leq t\}$, called a *norm cone*, is a convex cone.



- In the case when $\|x\|$ is the usual Euclidean norm, the norm cone is called the *second-order cone* and will be denoted by $\mathcal{K}_n$.

Examples of cones

- The set $\mathbb{P}_+^n$ of positive semidefinite $n \times n$ matrices is a convex cone.
- Indeed, for $A, B \in \mathbb{P}_+^n$, $\theta, \eta \geq 0$, and $u \in \mathbb{R}^n$,

$$\begin{aligned} u^{\mathrm{T}}(\theta A + \eta B)u &= \theta u^{\mathrm{T}} Au + \eta u^{\mathrm{T}} Bu \\ &\geq 0, \end{aligned}$$

i.e. $\theta A + \eta B \in \mathbb{P}_+^n$.

Proper cones

- A convex cone C is *proper*, if
  (i) it is closed (it contains its boundary),
  (ii) it is solid (has a nonempty interior),
  (iii) it is pointed (it does not contain a line.)
- Examples of proper cones include:
  (i) the *nonnegative orthant* $\{x \in \mathbb{R}^n : x_i \geq 0, i = 1, \ldots, n\}$,
  (ii) the cone $\mathbb{P}^n_+$ of positive semidefinite matrices,
  (iii) the second order cone $\mathcal{K}_n$.

Operations preserving convexity

- The intersection $\bigcap_{i=1}^{k} C_i$ of any number of convex sets C_i is convex.
- The image $f(C)$ of a convex set under an affine function $f(x) = Ax + b$ is convex. Indeed, if $y_1, y_2 \in f(C)$, then there exist $x_1, x_2 \in C$ such that

$$\begin{aligned}\alpha y_1 + (1-\alpha)y_2 &= \alpha(Ax_1 + b) + (1-\alpha)(Ax_2 + b)\\ &= A(\alpha x_1 + (1-\alpha)x_2) + b\end{aligned}$$

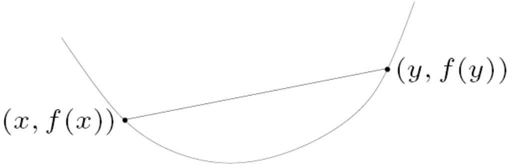
  which is an element of $f(C)$ since $\alpha x_1 + (1-\alpha)x_2 \in C$.
- The inverse image $f^{-1}(C)$ of a convex set under an affine function $f(x) = Ax + b$ is convex.

Convex functions

- Let C be a convex set. A function $f(x)$ defined on C is *convex*, if

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y), \tag{2}$$

for any $x, y \in C$ and $0 \leq \alpha \leq 1$.



- The inequality above defining a convex function is known as *Jensen's inequality*.

Convex functions

- A function $f(x)$ defined on a convex set C is *strictly convex*, if

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y), \tag{3}$$

  for any $x, y \in C$ and $0 \leq \alpha \leq 1$.
- Let C be a convex set. A function $f(x)$ defined on C is concave, if

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y), \tag{4}$$

  for any $x, y \in C$ and $0 \leq \alpha \leq 1$.
- Note that $f(x)$ is concave if and only if $-f(x)$ convex.
- An affine function is both convex and concave.

Properties of convex functions

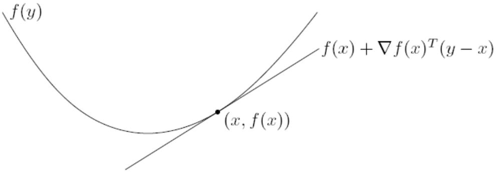
- The definition of a convex function is not always easy to verify in practice. Under additional smoothness assumptions on this may become more manageable. Below we let $\mathrm{dom}(f)$ denote the domain of the function $f(x)$.
- *First order conditions.* Suppose that $f(x)$ is (once) differentiable. Then it is convex if and only if the following conditions are satisfied:
  - (i) $\mathrm{dom}(f)$ is a convex set,
  - (ii) for all $x, y \in \mathrm{dom}(f)$,

$$f(y) \geq f(x) + \nabla f(x)^{\mathrm{T}}(y - x). \tag{5}$$

- Note that the right hand side in the inequality above is the first order Taylor expansion of $f(y)$ around $y = x$. It represents the hyperplane in $\mathbb{R}^n$ tangent to $f(y)$ at $y = x$.

Properties of convex functions

- Geometrically, this condition can be visualized as follows.



Properties of convex functions

- *Second order conditions.* Suppose now that $f(x)$ is *twice* differentiable. Then it is convex if and only if the following conditions are satisfied:
  (i) $\mathrm{dom}(f)$ is a convex set,
  (ii) for all $x \in \mathrm{dom}(f)$, the Hessian of $f(x)$ is a positive-semidefinite matrix

$$\nabla^2 f(x) \geq 0. \tag{6}$$

- The function is strictly convex, if its Hessian is a positive definite matrix.

Examples of convex functions

- The exponential function $f(x) = \exp(ax)$, where $a \in \mathbb{R}$, is convex on $\mathbb{R}$.
- The power function $f(x) = x^a$, $x > 0$, is convex, when $a \leq 0$ or $a \geq 1$, and concave, when $0 \leq a \leq 1$.
- Negative entropy $f(x) = x \log(x)$ is convex on its domain.
- The Rosenbrock function introduced in Lecture Notes #1 is neither convex nor concave over $\mathbb{R}^2$.
- $f(x, y) = x^2/y$ is convex over $\mathbb{R} \times \mathbb{R}_+$.
- The quadratic form $f(x) = \frac{1}{2}x^{\mathrm{T}}Px$, x$\in \mathbb{R}$, where is a positive semi-definite matrix $P \geq 0$ is convex. Indeed, its Hessian is given by

$$\nabla^2 f(x) = P.$$

It is strictly convex, if P is positive definite.

Properties of convex functions

- The following operations preserve convexity:
  (i) If $f_i(x)$, $i = 1, \dots, k$, are convex, and $\beta_i > 0$, then
  $$f(x) = \beta_1 f_1(x) + \ldots + \beta_k f_k(x)$$
  is convex.
  (ii) If $f(x)$ is convex and $A \in \mathrm{Mat}_{nm}(\mathbb{R})$, $b \in \mathbb{R}^m$, then the composed function
  $$g(x) = f(Ax + b)$$
  is convex.
  (iii) If $f_i(x)$, $i = 1, \dots, k$ are convex, then
  $$f(x) = \max\left((f_1(x), \dots, f_k(x)\right)$$
  is convex.
  (iv) If $f(x)$ is convex and $g(y)$, $y \in \mathbb{R}$, is convex and nondecreasing, then
  $$h(x) = g\big(f(x)\big)$$
  is convex.

Convex optimization problems

- A *standard form convex optimization problem* is formulated as follows:

$$\min_{x \in \mathbb{R}^n} f(x), \quad \text{subject to } \begin{cases} a_i^{\mathrm{T}} x = b_i, & \text{if } i \in \mathcal{E}, \\ c_i(x) \leq 0, & \text{if } i \in \mathcal{I}. \end{cases} \tag{7}$$

  where the objective function $f(x)$ and the constraints $c_i(x)$ are convex functions on $\mathbb{R}^n$, and where a_i and b_i are constant vectors.
- Note that, in a convex optimization problem, the equality constraints are assumed to be linear.
- This condition guarantees that the feasible set of feasible points of a convex optimization problem is convex.

Convex optimization problems

- *Example.* The following problem

$$\min x_1^2 + x_2^2, \quad \text{subject to } \begin{cases} \frac{x_1}{1+x_2^2} \leq 0, \\ (x_1 + x_2)^2 = 0, \end{cases}$$

  is not convex according to our definition (because the first constraint is not a convex function, and the second constraint is not linear). Perversely, the feasible set $\{(x_1, x_2) : x_1 \leq 0, x_2 = -x_1\}$ is convex.
- The following, equivalent but not identical, problem is convex:

$$\min x_1^2 + x_2^2, \quad \text{subject to } \begin{cases} x_1 \leq 0, \\ x_1 + x_2 = 0. \end{cases}$$

Examples of convex optimization problems

- The unconstraint least square problem

$$f(x)=\frac{1}{2}\sum_{i=1}^{n}\left((Ax)_i-v_i\right)^2,\ x\in\mathbb{R}^n, \tag{8}$$

where $v_i\in\mathbb{R}$, A$\in\mathrm{Mat}_n(\mathbb{R})$, is convex. Indeed, we verify that $\nabla^2 f=A^{\mathrm{T}}A$ which is positive semidefinite.

- On the other hand, the unconstraint nonlinear least square (NLS) problem

$$f(x)=\frac{1}{2}\sum_{i=1}^{n}\left(f_i(x)-v_i\right)^2,\ x\in\mathbb{R}^n, \tag{9}$$

may or may not be convex, depending on the functions $f_i(x)$.

Fundamental properties of convex optimization problems

- *Fundamental Property 1.* Any local minimum of a convex function $f(x)$ in a convex set C is also a global minimum.
- The *proof* goes by contradiction. If a local minimum x^* is not global, then we can find $y \in C$, such that $f(y) < f(x^*)$. Thus, by convexity,

$$\begin{aligned} f(\alpha x^* + (1-\alpha)y) &\leq \alpha f(x^*) + (1-\alpha)f(y) \\ &< f(x^*), \end{aligned}$$

  for all $0 < \alpha < 1$.
- In other words, since C is convex, $f(x)$ is strictly less than $f(x^*)$ along the line segment connecting y to x^*, regardless of how close x is to x^*. This contradicts the assumption that x^* is a local minimum, which proves the claim.
- In the following, we will denote the optimal value of $f(x)$ by f^*:

$$f^* = f(x^*). \tag{10}$$

Convex optimization problems

- Fundamental Property 1 does not assert that local minima of convex functions exist.
- There are various results guaranteeing the existence of minima of a convex function that can be found in the literature. An example is the classic Weierstrass' theorem.
- *Fundamental Property 2.* If the domain $\mathrm{dom}(f)$ of a continuous function $f(x)$ is closed and bounded, then it attains its minima.
- In particular, a convex function with bounded $\mathrm{dom}(f)$ has a unique global minimum.

Example: unconstrained quadratic optimization

- Consider the problem of minimizing the convex function

$$f(x) = \frac{1}{2}\, x^{\mathrm{T}} A x + a^{\mathrm{T}} x + b,\ \ x \in \mathbb{R}^n,$$

where A is a positive semidefinite matrix, $a \in \mathbb{R}^n$, and $b \in \mathbb{R}$.

- The optimality condition reads:

$$Ax + a = 0,$$

and we have the following possibilities:

(i) if A is invertible (i.e. it is positive definite), there is a unique solution $x^* = -A^{-1} a$.

(ii) if A is not invertible and a is not in the range of A, there is no solution, and $f(x) \to -\infty$, as $\|x\| \to \infty$ along certain directions.

(ii) if A is not invertible and a is in the range of A, there are infinitely many solutions.

Dual Lagrange function

- Recall from Lecture Notes #2 that the Lagrange function corresponding to the optimization problem (7) is defined by:

$$L(x, \lambda) = f(x) + \lambda^{\mathrm{T}} c(x), \tag{11}$$

where $\lambda \in \mathbb{R}^m$ is a vector of Lagrange multipliers corresponding to the (equality and inequality) constraints defining the feasible set Ω.

- The *Lagrange dual function* is defined as

$$\begin{aligned} q(\lambda) &= \inf_{x \in \Omega} L(x, \lambda) \\ &= \inf_{x \in \Omega} \big(f(x) + \lambda^{\mathrm{T}} c(x) \big). \end{aligned} \tag{12}$$

- Notice that $q(\lambda)$ has the following properties:
  (i) It is concave.
  (ii) It is a lower bound for f^*: if $\lambda_i \geq 0$, for $i \in \mathcal{I}$, then

$$q(\lambda) \leq f^*. \tag{13}$$

Dual Lagrange function

- (i) is true, because for $0 \leq \alpha \leq 1$

$$
\begin{aligned}
q(\alpha\lambda + (1-\alpha)\mu) &= \inf_{x\in\Omega} \big(f(x) + (\alpha\lambda + (1-\alpha)\mu)^{\mathrm{T}} c(x)\big) \\
&= \inf_{x\in\Omega} \big(\alpha(f(x) + \lambda^{\mathrm{T}} c(x)) + (1-\alpha)(f(x) + \mu^{\mathrm{T}} c(x))\big) \\
&\geq \alpha q(\lambda) + (1-\alpha) q(\mu),
\end{aligned}
$$

  as $\inf \big(u(x) + v(x)\big) \geq \inf u(x) + \inf v(x)$.

- (i) is true, because if x is any feasible point, then

$$
\begin{aligned}
f(x) &\geq f(x) + \lambda^{\mathrm{T}} c(x) \\
&\geq \inf_{x\in\Omega} L(x, \lambda) \\
&= q(\lambda).
\end{aligned}
$$

  Now we take the minimum over all feasible points.

Dual Lagrange function: example

- *Example.* Consider the problem:

$$\min \frac{1}{2}\, x^{\mathrm{T}} x \quad \text{subject to } Ax = b.$$

- The Lagrange function is $L(x, \lambda) = \frac{1}{2}\, x^{\mathrm{T}} x + \lambda^{\mathrm{T}}(Ax - b)$.
- Its minimum over x is at $x = -A^{\mathrm{T}}\lambda$. Plugging this into $L(x, \lambda)$, we find that

$$q(\lambda) = -\frac{1}{2}\lambda^{\mathrm{T}} A A^{\mathrm{T}} \lambda - b^{\mathrm{T}}\lambda.$$

- From the lower bound property of the dual Lagrange function we conclude that

$$f^* \geq -\frac{1}{2}\lambda^{\mathrm{T}} A A^{\mathrm{T}} \lambda - b^{\mathrm{T}}\lambda,$$

for all λ.

Dual Lagrange function: entropy maximization

- *Example.* Consider the problem of entropy maximization:

$$\min_p \sum_{i=1}^n p_i \log(p_i) \quad \text{subject to} \begin{cases} Ap \leq b, \\ \sum_{i=1}^n p_i = 1. \end{cases}$$

- The Lagrange function is

$$L(p, \lambda) = \sum_{i=1}^n p_i \log(p_i) + \sum_{i=1}^n \lambda_i (Ap - b)_i + \lambda_{n+1} \Big(\sum_{i=1}^n p_i - 1 \Big).$$

- Its minimum over p is at $p_i = \exp\big(-1 - \lambda_{n+1} - (A^{\mathrm{T}}\lambda)_i\big)$. Plugging this into $L(p, \lambda)$, we find that the dual Lagrange function is given by

$$q(\lambda) = -\sum_{i=1}^n \lambda_i b_i - \lambda_{n+1} - e^{-\lambda_{n+1} - 1} \sum_{i=1}^n e^{-(A^{\mathrm{T}}\lambda)_i}.$$

The dual problem

- The Lagrange dual problem is

$$\max_{\lambda} q(\lambda) \quad \text{subject to } \lambda_i \geq 0, \text{ for } i \in \mathcal{I}. \tag{14}$$

- This is a convex optimization problem, as $g(\lambda)$ is a concave function.
- In fact, this is a convex optimization problem, regardless of whether the original (primal) problem is convex or not.
- Its solution q^* provides the best lower bound for the primal problem.
- Recall from Lecture Notes #3 that the primal and dual problems in LP read:

$$\min c^{\mathrm{T}} x, \quad \text{subject to } \begin{cases} Ax = b, \\ x_i \geq 0, \text{ for } i = 1, \ldots, n. \end{cases}$$

and

$$\max b^{\mathrm{T}} y, \quad \text{subject to } \begin{cases} A^{\mathrm{T}} y + s = c, \\ s_i \geq 0, \text{ for } i = 1, \ldots, n. \end{cases}$$

Weak duality theorem

- The optimal value q^* of the dual problem is the best lower bound on f^* that can be obtained from the Lagrange dual function.
- In particular, we have the following important inequality:

$$q^* \leq f^* \tag{15}$$

  This property is called *weak duality*. It holds even if the original problem is not convex.
- Weak duality holds also when f^* and q^* are infinite.
- The difference $f^* - q^* \geq 0$ is called the *duality gap*.

Strong duality theorem and constraints qualification

- If the equality
$$q^* = f^* \tag{16}$$
holds, i.e., the duality gap is zero, then we say that *strong duality* holds.
- This means that the best bound that can be obtained from the Lagrange dual function is saturated.
- In general, strong duality does not hold. But if the primal problem is convex, strong duality holds under some additional conditions, called *constraint qualifications*.
- There are many known constraint qualifications. One simple example that we will discuss below is *Slater's condition*.

Constraint qualification: Slater's condition

- *Slater's condition:* There is a point x in the interior $\operatorname{int}\Omega$ of the feasible set Ω, such that

$$\begin{aligned} Ax &= b, \\ c_i(x) &< 0, \ \text{for } i \in \mathcal{I}. \end{aligned} \tag{17}$$

  Such a point $x \in \operatorname{int}\Omega$ is called *strictly feasible*.

- Assuming Slater's condition:
  (i) Strong duality holds.
  (ii) If $q^* > -\infty$, then the optimal value is attained: there exists λ^* such that $q(\lambda^*) = q^* = f^*$.

Constraint qualification: Slater's condition

- It is easy to see that Slater's condition holds for the last two examples.
- For the least square solution of a linear system, it just states that the primal system is feasible, provided that the vector b is in the range of the matrix A.
- For the entropy maximization example, Slater's condition says that there exists $p \in \mathbb{R}^n$ with $p \geq 0$, $Ap \leq b$ and $\sum_{i=1}^{n} p_i = 1$.

Constraint qualification: Slater's condition

- Recall that the necessary first order KKT conditions read: Let x^* be the solution to (7), and assume that x^* is regular. Then there exists a unique vector of Lagrange multipliers λ_i^*, $i = 1, \ldots m$, such that

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* c_i(x^*) &= 0, \\ c_i(x^*) &= 0, \text{ for } i \in \mathcal{E}, \\ c_i(x^*) &\leq 0, \text{ for } i \in \mathcal{I}, \\ \lambda_i^* c_i(x^*) &= 0, \text{ for } i \in \mathcal{I}, \\ \lambda_i^* &\geq 0, \text{ for } i \in \mathcal{I}. \end{aligned} \tag{18}$$

- Note that the second order condition is automatically satisfied, since the objective function and constraints are convex.
- Question: for convex problems, are these conditions also sufficient?

Constraint qualification: Slater's condition

- The answer is yes, if the problem satisfies Slater's condition.
- Under Slater's condition, x^* is optimal if and only if there exist λ^* that satisfy the KKT conditions.
- Slater's condition implies strong duality, and the dual optimum is attained.
- The first order KKT conditions generalize the optimality condition $\nabla f(x^*) = 0$ for unconstrained problems. We will see in the following that they are of great practical relevance.

Optimality criteria for convex optimization problems.

- In summary, the following optimality criteria follow from the KKT criteria.
  (i) *No constraints.* x^* is optimal if and only if
  $$\nabla f(x^*) = 0. \tag{19}$$
  (ii) *Equality constraints only.* x^* is optimal if and only if there exists $\lambda^* \in \mathbb{R}^m$ such that
  $$\begin{aligned} \nabla f(x^*) + A^{\mathrm{T}} \lambda^* &= 0, \\ Ax &= b. \end{aligned} \tag{20}$$
  (iii) *Equality and inequality constraints with Slater's condition.* x^* is optimal if and only if there exists $\lambda^* \in \mathbb{R}^m$ such that the KKT conditions are satisfied.

Example: optimization over the nonnegative orthant

- *Example.* Consider the minimization problem over the nonnegative orthant:

$$\min f(x) \quad \text{subject to } x_i \geq 0, i = 1, \ldots, n. \tag{21}$$

- The KKT conditions read

$$\begin{aligned} \nabla f(x^*) &= \lambda^*, \\ x_i^* &\geq 0, \\ \lambda_i^* &\geq 0, \\ \lambda_i^* x_i^* &= 0, \end{aligned} \tag{22}$$

for $i = 1, \ldots, n$.

Example: water-filling

- Consider the convex optimization problem:

$$\min \ -\sum_{i=1}^{n} \log(x+\alpha_i), \quad \text{subject to} \ \begin{cases} \sum_{i=1}^{n} x_i = 1, \\ x_i \geq 0, \ \text{for } i = 1, \ldots, n. \end{cases}$$

- The Lagrange function is

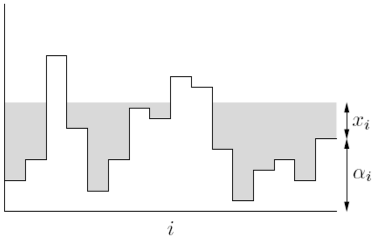
$$L(x,\lambda) = -\sum_{i=1}^{n} \log(x+\alpha_i) - \sum_{i=1}^{n} \lambda_i x_i + \lambda_{n+1}\Big(\sum_{i=1}^{n} x_i - 1\Big),$$

and the KKT conditions read: for $i = 1, \ldots, n$,

$$\begin{aligned} x_i^* &\geq 0, \\ \sum_{i=1}^{n} x_i^* &= 1, \\ \lambda_i^* &\geq 0, \\ \lambda_i^* x_i^* &= 0, \\ -\frac{1}{x_i^* + \alpha_i} + \lambda_i^* + \lambda_{n+1} &= 0. \end{aligned}$$

Example: water-filling

- Solving this system yields:
  (i) $\lambda_i^* = 0$ and $x_i^* = 1/\lambda_{n+1}^* - \alpha_i$, if $\lambda_{n+1}^* < 1/\alpha_i$.
  (ii) $\lambda_i^* = \lambda_{n+1}^* - 1/\alpha_i$ and $x_i^* = 0$, if $\lambda_{n+1}^* \geq 1/\alpha_i$.
  (iii) λ_{n+1}^* is uniquely determined from the condition $\sum_{i=1}^{n}(1/\lambda_{n+1}^* - \alpha_i)^+ = 1$.
- One way of interpreting this problem is as follows. α_i is the ground level above patch i. We flood the region with water to a depth $1/\lambda_{n+1}^*$, as shown in the figure below. The total amount of water is $\sum_{i=1}^{n}(1/\lambda_{n+1} - \alpha_i)^+$. We then increase the flood level until a total amount of water, equal to one, is used. The depth of water above patch i is then the optimal value x_i^*.



Formulation of the problem

- *Conic optimization* are among the simplest (and frequently showing up in applications) convex optimization problems, which have a linear objective function, and a single generalized inequality constraint function.
- Specifically, a conic optimization problem in standard form is formulated as follows:

$$\min \ f(x) \quad \text{subject to} \begin{cases} Ax = b, \\ x \in C, \end{cases} \tag{23}$$

  where C is a *convex cone* in $\mathbb{R}^n$.
- Note that in the special case of $C = \mathbb{R}^n_+$, (23) becomes an LP problem!
- In addition to LP, conic optimization includes two other important categories of optimization problems: second-order programming and semidefinite programming.

Second-order programming

- What happens when we the second-order cone with a hyperplane? We obtain an ellipsoid. This leads to the following class of optimization problems.
- Let $\|x\|$ denote the Euclidean norm. We call a constraint of the form

$$\|Ax + b\| \leq c^{\mathrm{T}} x + b, \tag{24}$$

where $A \in \mathrm{Mat}_{kn}(\mathbb{R})$ a *second-order cone constraint*. The condition is the same as requiring the vector $(Ax + b, c^{\mathrm{T}} x + d)$ to lie in the second-order cone $\mathbb{K}_k$ in $\mathbb{R}^{k+1}$.
- A *second-order cone program* (SOCP) is formulated as follows:

$$\min \ f^{\mathrm{T}} x \quad \text{subject to} \begin{cases} Fx = g, \\ \|A_i x + b_i\| \leq c_i^{\mathrm{T}} x + b_i, \ \text{for } i \in \mathcal{I}, \end{cases} \tag{25}$$

where $f \in \mathbb{R}^n$, $A_i \in \mathrm{Mat}_{n_i n}(\mathbb{R})$.

Second-order programming

- In other words, the inequality constraints in an SOCP are second order cone constraints.
- Convex optimization programs with quadratic inequality constraints can be converted into an SOCP.
- Consider a constraint of the form

$$x^{\mathrm{T}} P x + q^{\mathrm{T}} x + r \leq 0, \tag{26}$$

  where P is positive definite, and such that $q^{\mathrm{T}} P^{-1} q - 4r \geq 0$.
- Then there exists a non-singular matrix R such that $P = RR^{\mathrm{T}}$ (take e.g. the Cholesky decomposition), and the constraint reads

$$(R^{\mathrm{T}} x)^{\mathrm{T}} R^{\mathrm{T}} x + q^{\mathrm{T}} x + r \leq 0.$$

Second-order programming

- Completing the squares, we rewrite the inequality above as

$$(R^{\mathrm{T}}x + \frac{1}{2}R^{-1}q)^{\mathrm{T}}(R^{\mathrm{T}}x + \frac{1}{2}R^{-1}q) \leq \frac{1}{4}q^{\mathrm{T}}P^{-1}q - r.$$

- Introducing new variables $y = R^{\mathrm{T}}x + R^{-1}q/2$, we see that (26) is equivalent to the following second-order constraint (along with equality constraints):

$$\begin{aligned} y &= R^{\mathrm{T}}x + \frac{1}{2}R^{-1}q, \\ y_0 &= \frac{1}{2}\sqrt{q^{\mathrm{T}}P^{-1}q - 4r}, \\ (y_0, y) &\in \mathbb{K}_n. \end{aligned} \tag{27}$$

- SOCPs arise in portfolio management, we will discuss an application in Lecture Notes #5.

Semidefinite programming

- *Semidefinite programming* problems (SDP) arise whenever there is need to estimate a covariance matrix from the data. For example:
  - (i) In portfolio management, we need to estimate, from the data, the covariance matrix of returns of the assets.
  - (ii) In a multi-factor terms structure of interest rates, we need to estimate the covariance matrix of the factors driving the interest rate process.
- Specifically, a semidefinite problem arises when the cone C in (23) is the cone $\mathbb{P}_+^n$ of positive semidefinite matrices:

$$\min \ c^{\mathrm{T}} x \quad \text{subject to} \begin{cases} Ax = b, \\ x_1 F_1 + \ldots + x_n F_n + G \in \mathbb{P}_+^n, \end{cases} \tag{28}$$

where $F_i \in \mathbb{S}^k$, $i = 1, \ldots, n$ and $G \in \mathbb{S}^k$ are symmetric matrices of dimension k (their set is denoted by $\mathbb{S}^k$).

Semidefinite programming

- A standard form semidefinite problem is formulated as follows:

$$\min \ \mathrm{tr}(CX) \quad \text{subject to} \begin{cases} \mathrm{tr}(A_i X) = b_i, \ \text{for } i \in \mathcal{I}, \\ X \in \mathbb{P}^n_+, \end{cases} \tag{29}$$

  where $C \in \mathbb{S}^n$ and $A_i \in \mathbb{S}^n$ are symmetric.
- As in the case of LP, in the standard form SDP, we minimize a linear function of the variable, subject to linear equality constraints, and a nonnegativity constraint on the variables.

References

Boyd, S., and Vanderberghe, L.: *Convex Optimization*, Cambridge University Press (2004).