Convexity, duality, optimization

Torkel A. Haufmann
SINTEF Digital

Outline

1. Continuous optimization
2. Convexity
3. Convex optimization theory
4. Applications

Continuous problems

Today we will talk about *continuous* problems of the form

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in C, \end{aligned}$$

where $C \subseteq \mathbb{R}^n$ for some n. Assumptions:

- f is continuous.
- C is connected – and closed.

Continuous problems

Today we will talk about *continuous* problems of the form

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in C, \end{aligned}$$

where $C \subseteq \mathbb{R}^n$ for some n.
An example you've seen: linear programming!

Question

How can we solve problems on this form?

Continuous problems

Let's talk classification again: Given $(\mathbf{x}_i, y_i)$ for $i \in [n]$ and $y_i \in \{-1, 1\}, \mathbf{x}_i \in \mathbb{R}^p$ for all i, find β, α which best classifies the $\mathbf{x}_i$. That is to say, we'd like $\beta^T \mathbf{x}_i - \alpha = y_i$ for $i \in [n]$ – or at least that they have the same sign.

Question

How do we get this?

Continuous problems

Let's talk classification again: Given $(\mathbf{x}_i, y_i)$ for $i \in [n]$ and $y_i \in \{-1, 1\}, \mathbf{x}_i \in \mathbb{R}^p$ for all i, find β, α which best classifies the $\mathbf{x}_i$. That is to say, we'd like $\beta^T \mathbf{x}_i - \alpha = y_i$ for $i \in [n]$ – or at least that they have the same sign.

Question

How do we get this?

We want $y_i(\beta^T \mathbf{x}_i - \alpha) > 0$ for all $i \in [n]$.
Let l be a *loss* function, then we want to

$$\text{minimize} \quad \frac{1}{n} \sum_{i=1}^{n} l(y_i, \beta^T \mathbf{x}_i - \alpha) + \frac{1}{2} \|\beta\|^2.$$

Continuous problems

Exponential loss

$$l(y, u) = e^{-yu},$$

Logistic loss

$$l(y, u) = \log(1 + e^{-yu}),$$

Hinge loss

$$l(y, u) = \max\{0, 1 - yu\}.$$

Continuous problems

Exponential loss

$$l(y, u) = e^{-yu},$$

Logistic loss

$$l(y, u) = \log(1 + e^{-yu}),$$

Hinge loss

$$l(y, u) = \max\{0, 1 - yu\}.$$

Support Vector Machine:

$$\text{minimize} \quad \tfrac{1}{2}\|\beta\|^2 + \tfrac{1}{n}\sum_{i=1}^{n} \max(0, 1 - y_i(\beta^T \mathbf{x}_i - \alpha))$$

Convexity

Definition

A set C is *convex* if, for all $x, y \in C$, the line segment between x and y is entirely in C.

Or put differently,

$$\lambda x + (1 - \lambda) y \in C$$

for all $\lambda \in [0, 1]$.

Convexity

Definition

A set C is *convex* if, for all $x, y \in C$, the line segment between x and y is entirely in C.

Or put differently,

$$\lambda x + (1 - \lambda)y \in C$$

for all $\lambda \in [0, 1]$.

Definition

A function f is *convex* if, for all x, y in its domain, the line segment between $(x, f(x))$ and $(y, f(y))$ is above the graph of f.

Which is to say,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

for all $\lambda \in [0, 1]$.

Convexity

Definition

The *epigraph* of f is the set $\{(x, y) : y \geq f(x)\}$.

Convexity

Definition

The *epigraph* of f is the set $\{(x, y) : y \geq f(x)\}$.

Theorem

f is convex if and only if its epigraph is.

Convexity

Definition

The *epigraph* of f is the set $\{(x, y) : y \geq f(x)\}$.

Theorem

f is convex if and only if its epigraph is.

Exercise

Prove it!

Properly understood, every convex set has an *interior*.

Convexity

Definition

The *epigraph* of f is the set $\{(x, y) : y \geq f(x)\}$.

Theorem

f is convex if and only if its epigraph is.

Exercise

Prove it!

Properly understood, every convex set has an *interior*.

Theorem

If f is differentiable and defined on a convex set C, f is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^T (x - y)$$

for all $x, y \in C$.

Convexity

Definition

A *convex combination* of points $x_1, \ldots, x_n \in C$ is a combination

$$\sum_{i=1}^{n} \lambda_i x_i,$$

where $\sum_{i=1}^{n} \lambda_i = 1$ and each $\lambda_i \geq 0$.

Convexity

Definition

A *convex combination* of points $x_1, \ldots, x_n \in C$ is a combination

$$\sum_{i=1}^{n} \lambda_i x_i,$$

where $\sum_{i=1}^{n} \lambda_i = 1$ and each $\lambda_i \geq 0$.

Theorem (Jensen's inequality)

If f is convex and we apply it to a convex combination, we get

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i).$$

Convexity

Convex sets:

- Half-space: $\{x \in \mathbb{R}^p : \mathbf{a}^T\mathbf{x} \leq b\}$.
- Ball: $\{x \in \mathbb{R}^p : \|\mathbf{x} - \mathbf{c}\| \leq d\}$.

Convexity

Convex sets:

- Half-space: $\{x \in \mathbb{R}^p : \mathbf{a}^T\mathbf{x} \leq b\}$.
- Ball: $\{x \in \mathbb{R}^p : \|\mathbf{x} - \mathbf{c}\| \leq d\}$.

Combination:

- Convexity is preserved by intersection.
- $\{A\mathbf{x} + \mathbf{b} : \mathbf{x} \in C\}$, where C is convex.

Convexity

Convex sets:

- Half-space: $\{x \in \mathbb{R}^p : \mathbf{a}^T\mathbf{x} \leq b\}$.
- Ball: $\{x \in \mathbb{R}^p : \|\mathbf{x} - \mathbf{c}\| \leq d\}$.

Combination:

- Convexity is preserved by intersection.
- $\{A\mathbf{x} + \mathbf{b} : \mathbf{x} \in C\}$, where C is convex.

So for example polyhedra are convex.

Convexity

Convex functions:

- Norms.
- Powers ≥ 1.
- $-\log$ is convex on the positive reals.

Convexity

Convex functions:

- Norms.
- Powers ≥ 1.
- $-\log$ is convex on the positive reals.

Combinations:

- Nonnegative sums of convex functions are convex.
- If f is convex, then $f(Ax + b)$ is convex in x.
- $\max_i f_i(x)$ is convex if the f_i are convex.
- $f \circ g$ is convex if g is convex and f is convex and non-decreasing.

Convexity

Convex functions:

- Norms.
- Powers ≥ 1.
- $-\log$ is convex on the positive reals.

Combinations:

- Nonnegative sums of convex functions are convex.
- If f is convex, then $f(Ax + b)$ is convex in x.
- $\max_i f_i(x)$ is convex if the f_i are convex.
- $f \circ g$ is convex if g is convex and f is convex and non-decreasing.

So for example the SVM objective is convex.

Convexity

Convex functions:

- Norms.
- Powers ≥ 1.
- $-\log$ is convex on the positive reals.

Combinations:

- Nonnegative sums of convex functions are convex.
- If f is convex, then $f(Ax + b)$ is convex in x.
- $\max_i f_i(x)$ is convex if the f_i are convex.
- $f \circ g$ is convex if g is convex and f is convex and non-decreasing.

So for example the SVM objective is convex.

Theorem

A function f is convex if and only if $f(x + \lambda y)$ is convex for any given $x, y, \lambda \in \mathbb{R}$.

Convexity

Convex functions:

- Norms.
- Powers ≥ 1.
- $-\log$ is convex on the positive reals.

Combinations:

- Nonnegative sums of convex functions are convex.
- If f is convex, then $f(Ax + b)$ is convex in x.
- $\max_i f_i(x)$ is convex if the f_i are convex.
- $f \circ g$ is convex if g is convex and f is convex and non-decreasing.

So for example the SVM objective is convex.

Theorem

A function f is convex if and only if $f(x + \lambda y)$ is convex for any given $x, y, \lambda \in \mathbb{R}$.

Convexity

Why do we care about convexity?

Let $f : C \to \mathbb{R}$, and assume C is closed.

Definition

A point $x^* \in C$ is a *local minimum* of f if there is an ϵ such that for all $x \in C$,

$$\|x - x^*\| < \epsilon \Rightarrow f(x) \geq f(x^*).$$

Convexity

Why do we care about convexity?

Let $f : C \to \mathbb{R}$, and assume C is closed.

Definition

A point $x^* \in C$ is a *local minimum* of f if there is an ϵ such that for all $x \in C$,

$$\|x - x^*\| < \epsilon \Rightarrow f(x) \geq f(x^*).$$

Definition

A point $x^* \in C$ is a *global minimum* of f if, for all $x \in C$,

$$f(x^*) \leq f(x).$$

Convexity

Why do we care about convexity?

Let $f : C \to \mathbb{R}$, and assume C is closed.

Definition

A point $x^* \in C$ is a *local minimum* of f if there is an ϵ such that for all $x \in C$,

$$\|x - x^*\| < \epsilon \Rightarrow f(x) \geq f(x^*).$$

Definition

A point $x^* \in C$ is a *global minimum* of f if, for all $x \in C$,

$$f(x^*) \leq f(x).$$

Alternatively, x^* is a global minimum if it is a local minimum and all other local minima x satisfy $f(x) \geq f(x^*)$.

Question

What if C is open?

Convexity

Now we assume f, C are convex.

Theorem

f obtains its minimum on C.

Convexity

Now we assume f, C are convex.

Theorem

f obtains its minimum on C.

In fact we can do better:

Theorem

If $x^* \in C$ is a local minimum of f it is a global minimum.

Question

Why?

Theorem

If f is C^1, x^* is its global minimum if and only if $\nabla f(x^*) = 0$.

Convexity

Recall:

Definition (Extreme point)

$z \in C$ is an **extreme point** of the set C if we cannot find $x, y \in C$, with $x, y \neq z$, and scalar $\lambda \in [0, 1]$ such that $z = \lambda x + (1 - \lambda)y$.

Convexity

Recall:

Definition (Extreme point)

$z \in C$ is an **extreme point** of the set C if we cannot find $x, y \in C$, with $x, y \neq z$, and scalar $\lambda \in [0, 1]$ such that $z = \lambda x + (1 - \lambda)y$.

Definition (Convex hull)

The *convex hull* of a set of points is the smallest convex set containing the points.

Convexity

Recall:

Definition (Extreme point)

$z \in C$ is an **extreme point** of the set C if we cannot find $x, y \in C$, with $x, y \neq z$, and scalar $\lambda \in [0, 1]$ such that $z = \lambda x + (1 - \lambda)y$.

Definition (Convex hull)

The *convex hull* of a set of points is the smallest convex set containing the points.

Theorem (Krein-Milman)

If C is a convex, compact set, it has extreme points and is the closed convex hull of said extreme points.

Polyhedra, again

Recall that a *polyhedron* is a (convex) set of the form

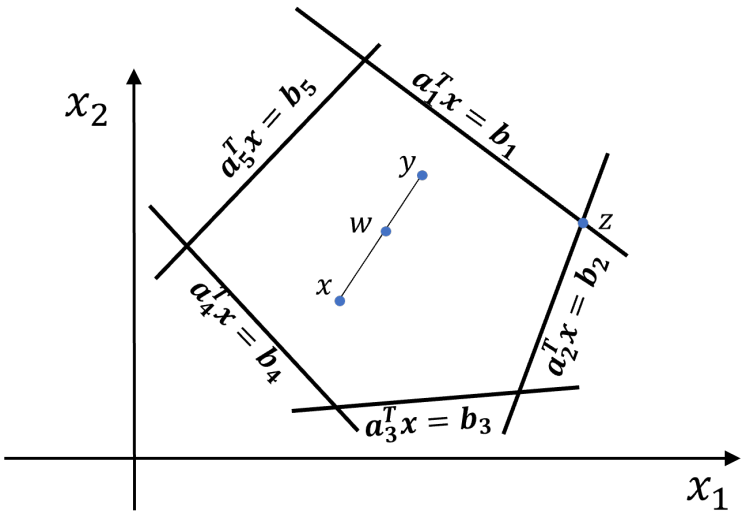
$$P = \{\mathbf{x} \in \mathbb{R}^p : A\mathbf{x} \leq \mathbf{b}\}.$$

Polyhedra, again

Recall that a *polyhedron* is a (convex) set of the form

$$P = \{\mathbf{x} \in \mathbb{R}^p : A\mathbf{x} \leq \mathbf{b}\}.$$

An extreme point of P satisfies p inequalities with equality*.



In this way, a *combinatorial aspect* of linear optimization appears.

Convex optimization

Let's go back:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in C, \end{aligned}$$

If f, C are convex this is known as a *convex optimization* problem.
Some over-simplifications:

- *Linear optimization* problems can be solved in polynomial time and have good theoretical properties.
- *Convex optimization* problems can *sometimes* be solved efficiently and have good theoretical properties.
- *General continuous optimization* problems do not have good theoretical properties and can only be solved in special cases.

Question

What does *solving* mean here?

Convex optimization

Methods for convex optimization problems include[1]:

- The ellipsoid method.
- Gradient methods.
- Interior-point methods.

[1] Recall lecture 1!

Convex optimization

Methods for convex optimization problems include[1]:

- The ellipsoid method.
- Gradient methods.
- Interior-point methods.

Let's say that C is defined by inequalities:

$$C = \{x \in \mathbb{R}^n : c_i(x) \leq 0 \text{ for } i \in [m]\}.$$

Hence the problem is:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & c_i(x) \leq 0, (i \in [m]) \end{aligned}$$

Question

Which conditions must the c_i satisfy?

[1] Recall lecture 1!

Descent directions

Now say that $x \in C$ and let s be some small step. For $x + s$ to be better than x, we must have

$$f(x + s) < f(x) \Rightarrow f(x + s) - f(x) < 0.$$

This is "approximately"

$$\nabla f(x)^T s < 0.$$

Question

Is this intuitive?

Descent directions

Now say that $x \in C$ and let s be some small step. For $x + s$ to be better than x, we must have

$$f(x + s) < f(x) \Rightarrow f(x + s) - f(x) < 0.$$

This is "approximately"

$$\nabla f(x)^T s < 0.$$

Question

Is this intuitive?

Similarly, for each i we need

$$0 \geq \nabla c_i(x)^T s.$$

Descent directions

So does there exist such an s?

- If x is in the interior, then yes if x is not a local minimum.
- If x is *on the boundary*, it is not so easy.

Descent directions

So does there exist such an s?

- If x is in the interior, then yes if x is not a local minimum.
- If x is *on the boundary*, it is not so easy.

Let's say $c_i(x) = 0$ for $i \in [m']$. Then we need

$$\nabla c_i(x)^T s \geq 0,$$

for $i \in [m']$. So can we find s with $x + s \in C$ and

$$\nabla f(x)^T s < 0, \qquad \nabla c_i(x)^T s \leq 0 \text{ for } i \in [m']?$$

If not x is a local minimum*.

The Lagrangian

Our *primal* problem is

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & c_i(x) \leq 0, (i \in [m]). \end{aligned}$$

We define the *Lagrangian*

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i c_i(x),$$

and the *Lagrange dual*

$$g(\lambda) = \inf_x L(x, \lambda).$$

Duality

Let p^* be the optimal value of the primal problem.

Theorem (Weak duality)

For $\lambda \geq 0$,

$$g(\lambda) \leq p^*$$

Duality

Let p^* be the optimal value of the primal problem.

Theorem (Weak duality)

For $\lambda \geq 0$,

$$g(\lambda) \leq p^*$$

Indeed, for any $x \in C$,

$$g(\lambda) = \inf_x L(x, \lambda) \leq f(x) + \sum_{i=1}^{m} \lambda_i c_i(x),$$

and as $x \in C$, $\lambda \geq 0$ we see that

$$g(\lambda) \leq f(x).$$

Duality

This motivates us to define the *dual problem*:

$$\begin{aligned} \text{maximize} \quad & g(\lambda) \\ \text{subject to} \quad & \lambda_i \geq 0, (i \in [m]). \end{aligned}$$

This is a convex optimization problem!*

Duality

This motivates us to define the *dual problem*:

$$\begin{aligned} \text{maximize} \quad & g(\lambda) \\ \text{subject to} \quad & \lambda_i \geq 0, (i \in [m]). \end{aligned}$$

This is a convex optimization problem!*

If we let d^* be the optimal value of the dual, we see that weak duality simply says

$$d^* \leq p^*.$$

But can we go further?

Duality

Definition

We say that *strong duality* holds for a primal-dual pair if

$$p^* = d^*.$$

Duality

Definition

We say that *strong duality* holds for a primal-dual pair if

$$p^* = d^*.$$

Definition

We say that *Slater's condition* holds if there is a *strictly feasible* x in the primal, i.e. an x such that

$$c_i(x) < 0,$$

for $i \in [m]$.

Duality

Definition

We say that *strong duality* holds for a primal-dual pair if

$$p^* = d^*.$$

Definition

We say that *Slater's condition* holds if there is a *strictly feasible* x in the primal, i.e. an x such that

$$c_i(x) < 0,$$

for $i \in [m]$.

Theorem

If the primal problem is convex and Slater's condition holds, we have strong duality.

KKT conditions

Assume the c_i are differentiable*, and that x^* is primal optimal and λ^* is dual optimal.
Then we have (why?):

$$\nabla L(x^*, \lambda^*) = 0.$$

That is,

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla c_i(x) = 0.$$

We also have feasibility for both x^* and λ^*.

KKT conditions

We have derived (almost!) the Karush-Kuhn-Tucker conditions:

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) &= 0, \\ c_i(x^*) &\leq 0, \quad (i \in [m]), \\ \lambda_i^* &\geq 0, \quad (i \in [m]), \\ \lambda_i^* c_i(x^*) &= 0, \quad (i \in [m]). \end{aligned}$$

KKT conditions

We have derived (almost!) the Karush-Kuhn-Tucker conditions:

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) &= 0, \\ c_i(x^*) &\leq 0, \quad (i \in [m]), \\ \lambda_i^* &\geq 0, \quad (i \in [m]), \\ \lambda_i^* c_i(x^*) &= 0, \quad (i \in [m]). \end{aligned}$$

Exercise

Show that such x^*, λ^* must be optimal to the original problems.

KKT conditions

We have derived (almost!) the Karush-Kuhn-Tucker conditions:

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) &= 0, \\ c_i(x^*) &\leq 0, \quad (i \in [m]), \\ \lambda_i^* &\geq 0, \quad (i \in [m]), \\ \lambda_i^* c_i(x^*) &= 0, \quad (i \in [m]). \end{aligned}$$

Exercise

Show that such x^*, λ^* must be optimal to the original problems.

Theorem

If

- *the primal problem is convex,*
- *the objective and constraints are differentiable,*
- *and Slater's condition applies,*

the KKT conditions are necessary and sufficient for optimality.

Support Vector Machine

Recall:

- We are given $(\mathbf{x}_i, y_i)$ for $i \in [n]$ and $y_i \in \{-1, 1\}, \mathbf{x}_i \in \mathbb{R}^p$ for all i.
- We want to find β, α such that for all i, $y_i(\beta^T \mathbf{x}_i - \alpha) > 0$.

For this purpose we wish to solve

$$\text{minimize} \quad \frac{1}{2}\|\beta\|^2 + \frac{1}{n}\sum_{i=1}^{n} \max(0, 1 - y_i(\beta^T \mathbf{x}_i - \alpha))$$

Question

Are the KKT conditions necessary and sufficient for optimality here?

Support Vector Machine

$$\text{minimize} \quad \frac{1}{2}\|\beta\|^2 + \frac{1}{n}\sum_{i=1}^{n} \max(0, 1 - y_i(\beta^T \mathbf{x}_i - \alpha))$$

Equivalent formulation:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|\beta\|^2 + \frac{1}{n}\sum_{i=1}^{n} w_i \\ \text{subject to} \quad & y_i(\beta^T x_i - \alpha) \geq 1 - w_i, \\ & w_i \geq 0. \end{aligned}$$

Question

Why is this equivalent?

Question

Are the KKT conditions necessary and sufficient for optimality here?

Exercises

Exercise

Prove that for a closed, compact convex set, the convex hull is the set of convex combinations of its extreme points.

Exercise

Prove Jensen's inequality.

Exercise

What is the dual of a linear program?

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \end{aligned}$$

Exercises

Exercise

From Boyd, Vandenberghe.

What is the dual of this problem?

$$\begin{aligned} \text{minimize} \quad & x^T x \\ \text{subject to} \quad & Ax = b \end{aligned}$$

Hint: The Lagrangian is the same, but equality constraints get "free" dual variables.

Exercise

Derive the dual of the following formulation:

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2}\|\beta\|^2 + \tfrac{1}{n}\sum_{i=1}^{n} w_i \\ \text{subject to} \quad & y_i(\beta^T x_i - \alpha) \geq 1 - w_i, \\ & w_i \geq 0. \end{aligned}$$

References

- Bertsimas, D. and Dunn, J., 2019. Machine Learning Under a Modern Optimization Lens.
- Boyd, S. and Vandenberghe, L., 2004. Convex Optimization.
- Nocedal, J. and Wright, S. J., 2006. Numerical Optimization - Second Edition.