Computational Method of Optimization

Convex optimization problems - Lecture V

Convex optimization problems

- Agenda
- optimization problem in standard form
- convex optimization problems
- quasiconvex optimization
- linear optimization

Source: Convex Optimization — Boyd & Vandenberghe

Optimization problem in standard form

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

- $x \in \mathbf{R}^n$ is the optimization variable
- $f_0 : \mathbf{R}^n \to \mathbf{R}$ is the objective or cost function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$, are the inequality constraint functions
- $h_i : \mathbf{R}^n \to \mathbf{R}$ are the equality constraint functions

optimal value:

$$p^\star = \inf\{f_0(x) \mid f_i(x) \leq 0, \ i = 1, \ldots, m, \ h_i(x) = 0, \ i = 1, \ldots, p\}$$

- $p^\star = \infty$ if problem is infeasible (no x satisfies the constraints)
- $p^\star = -\infty$ if problem is unbounded below

We'll start with just the boring nomenclature of an optimization problem. So an optimization problem looks like this. This is a notation for it is to minimize an objective, $f_0(x)$ subject to $f_i(x) \leq 0$. That's from i = 1,...,m, and $h_i(x) = 0$. So, this is sort of a standard form for an optimization problem.

Now here x is the optimization variable. This is called the cost. It's sometimes called the objective. It's got all sorts of other names or – if it's in a specific application area, it'll have a name, too.

Then these are called the inequality constraint functions, and the h's are the equality constraint functions. Now here the right hand sides are zero, if the right hand side weren't zero, how you could transform it. So it's just useful to have a standard form like this.

Optimization problem in standard form

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

- $x \in \mathbf{R}^n$ is the optimization variable
- $f_0 : \mathbf{R}^n \to \mathbf{R}$ is the objective or cost function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$, are the inequality constraint functions
- $h_i : \mathbf{R}^n \to \mathbf{R}$ are the equality constraint functions

optimal value:

$$p^\star = \inf\{f_0(x) \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\}$$

- $p^\star = \infty$ if problem is infeasible (no x satisfies the constraints)
- $p^\star = -\infty$ if problem is unbounded below

Now the optimal value of this problem – so the type of object or data what this is that's an optimization problem without the min, or with the min.

But you can talk about the optimal value of a problem, and that's simply the infimum or minimum of $f_0(x)$ over all x that satisfy the equality constraints. The infimum of an empty set is infinity, so that's standard.

Optimization problem in standard form

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

- $x \in \mathbf{R}^n$ is the optimization variable
- $f_0 : \mathbf{R}^n \to \mathbf{R}$ is the objective or cost function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$, are the inequality constraint functions
- $h_i : \mathbf{R}^n \to \mathbf{R}$ are the equality constraint functions

optimal value:

$$p^\star = \inf\{f_0(x) \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\}$$

- $p^\star = \infty$ if problem is infeasible (no x satisfies the constraints)
- $p^\star = -\infty$ if problem is unbounded below

The optimal value of this problem is infinity. If there's no feasible x, if there's no x that satisfies these inequalities and equalities.
And that makes sense because if your objective is to minimize something, then the value plus infinity means you minimize something if the value is infinity. And that is after all what infeasibility means. It means that there weren't any the hard constraints that met.

Now on the other hand, this could be minus infinity, and if there were a sequence of x's all feasible with f_0(x) k going to minus infinity.

For example, in risk sensitive estimation, they have a name for that when the problem is unbounded below.

Optimal and locally optimal points

x is **feasible** if $x \in \mathbf{dom}\, f_0$ and it satisfies the constraints

a feasible x is **optimal** if $f_0(x) = p^\star$; $X_{\rm opt}$ is the set of optimal points

x is **locally optimal** if there is an $R > 0$ such that x is optimal for

$$\begin{array}{ll} \text{minimize (over } z) & f_0(z) \\ \text{subject to} & f_i(z) \le 0, \quad i = 1, \ldots, m, \quad h_i(z) = 0, \quad i = 1, \ldots, p \\ & \|z - x\|_2 \le R \end{array}$$

examples (with $n = 1$, $m = p = 0$)

- $f_0(x) = 1/x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = 0$, no optimal point
- $f_0(x) = -\log x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -\infty$
- $f_0(x) = x \log x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -1/e$, $x = 1/e$ is optimal
- $f_0(x) = x^3 - 3x$, $p^\star = -\infty$, local optimum at $x = 1$

The idea of optimal and locally optimal points. You say a point is feasible if it's in the domain of the objective and it satisfies the constraint. If it satisfies the constraints, it must also be in the domain of the constraint functions because you wouldn't say it satisfies the constraints if it didn't also satisfy the constraints

$f_i(x)$ has to return a number and 'the not in domain token' because there's no way, if that returned not in domain, that you could say that the inequality was valid.

And a point is optimal if it's feasible and if its objective value is equal to the optimal value of the problem. That's optimal. X_{opt} is the set of optimal points. And this can be empty, for example if the problem is infeasible, it's empty. If the problem is unboundable below, this is empty. And when p* is finite, this can be empty.

Optimal and locally optimal points

x is **feasible** if $x \in \textbf{dom}\, f_0$ and it satisfies the constraints

a feasible x is **optimal** if $f_0(x) = p^\star$; X_{opt} is the set of optimal points

x is **locally optimal** if there is an $R > 0$ such that x is optimal for

$$\begin{array}{ll} \text{minimize (over } z) & f_0(z) \\ \text{subject to} & f_i(z) \leq 0, \quad i = 1, \ldots, m, \quad h_i(z) = 0, \quad i = 1, \ldots, p \\ & \|z - x\|_2 \leq R \end{array}$$

examples (with $n = 1$, $m = p = 0$)

- $f_0(x) = 1/x$, $\textbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = 0$, no optimal point
- $f_0(x) = -\log x$, $\textbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -\infty$
- $f_0(x) = x \log x$, $\textbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -1/e$, $x = 1/e$ is optimal
- $f_0(x) = x^3 - 3x$, $p^\star = -\infty$, local optimum at $x = 1$

X_{opt} is the empty set. Let's just minimize 1/x on the positive axis. If you minimize 1/x, the optimal value is zero, but it's not achieved because there's no x for which 1/x – there's no x in R_{++} for which 1/x is equal to zero. So in this case, you would have X_{opt} is empty, and that's a problem with optimal value zero which is not achieved

Now you say a problem is locally optimal if when you add a constraint – notice that this depends on x – if you add a constraint, that you not stray far in norm from x. R has to be positive here. If you can add such a constraint, and x becomes optimal, then it's locally optimal.

Optimal and locally optimal points

x is **feasible** if $x \in \mathbf{dom}\, f_0$ and it satisfies the constraints

a feasible x is **optimal** if $f_0(x) = p^\star$; $X_{\rm opt}$ is the set of optimal points

x is **locally optimal** if there is an $R > 0$ such that x is optimal for

$$\begin{array}{ll} \text{minimize (over } z\text{)} & f_0(z) \\ \text{subject to} & f_i(z) \le 0, \quad i = 1, \ldots, m, \quad h_i(z) = 0, \quad i = 1, \ldots, p \\ & \|z - x\|_2 \le R \end{array}$$

examples (with $n = 1$, $m = p = 0$)

- $f_0(x) = 1/x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = 0$, no optimal point
- $f_0(x) = -\log x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -\infty$
- $f_0(x) = x \log x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -1/e$, $x = 1/e$ is optimal
- $f_0(x) = x^3 - 3x$, $p^\star = -\infty$, local optimum at $x = 1$

These examples are very straightforward. So, if you minimize minus log, it's unbounded below obviously because I just take infinitely large x's and the sequence of x's getting larger and larger, negative log gets smaller and smaller.

Entropy here – that negative entropy – that has a perfectly well-defined minimum that occurs at 1/e. You just differentiate, set it equal to zero, and the optimal value is -1/e.

Optimal and locally optimal points

x is **feasible** if $x \in \mathbf{dom}\, f_0$ and it satisfies the constraints

a feasible x is **optimal** if $f_0(x) = p^\star$; $X_{\rm opt}$ is the set of optimal points

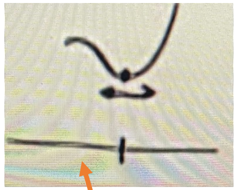
x is **locally optimal** if there is an $R > 0$ such that x is optimal for

$$\begin{array}{ll} \text{minimize (over } z) & f_0(z) \\ \text{subject to} & f_i(z) \le 0, \quad i = 1, \dots, m, \quad h_i(z) = 0, \quad i = 1, \dots, p \\ & \|z - x\|_2 \le R \end{array}$$

examples (with $n = 1$, $m = p = 0$)

- $f_0(x) = 1/x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = 0$, no optimal point
- $f_0(x) = -\log x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -\infty$
- $f_0(x) = x \log x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -1/e$, $x = 1/e$ is optimal
- $f_0(x) = x^3 - 3x$, $p^\star = -\infty$, local optimum at $x = 1$

If you have f0(x) = x^3 – 3x, here it’s unbounded below. Obviously, as x goes to minus infinity, this thing goes to minus infinity, so it’s unbounded below, but you have a local optimum at x = 1.



You will find that around x = 1, this thing goes like that. The derivative is zero, and in a little neighborhood around it, it’s locally convex there, and that means as you stray from x = 1, the objective value goes up, so that’s locally optimal.

Implicit constraints

the standard form optimization problem has an **implicit constraint**

$$x \in \mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \;\cap\; \bigcap_{i=1}^{p} \mathbf{dom}\, h_i,$$

- we call $\mathcal{D}$ the **domain** of the problem
- the constraints $f_i(x) \le 0$, $h_i(x) = 0$ are the explicit constraints
- a problem is **unconstrained** if it has no explicit constraints ($m = p = 0$)

example:

$$\text{minimize} \quad f_0(x) = -\sum_{i=1}^{k} \log(b_i - a_i^T x)$$

is an unconstrained problem with implicit constraints $a_i^T x < b_i$

There's an implicit constraint that the only x's you'll consider are in the domain of all functions, all objective and constraint functions, the objective and all constraints. So that's sometimes called an implicit constraint because you don't write it down explicitly

Merely by writing down f_i and then called on x, you are imposing the constraint that x is in the domain of f_i. So that's why it's implicit.

And you say a problem is unconstrained if it has no explicit constraints. Here would be a very common example, one in fact that we'll see a great deal of. It minimizes the following function. It's the sum of the negative $\log(b_i - a_i^T x)$. Now to talk about log of something, this has to be positive.

Implicit constraints

the standard form optimization problem has an **implicit constraint**

$$x \in \mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \;\cap\; \bigcap_{i=1}^{p} \mathbf{dom}\, h_i,$$

- we call $\mathcal{D}$ the **domain** of the problem
- the constraints $f_i(x) \leq 0$, $h_i(x) = 0$ are the explicit constraints
- a problem is **unconstrained** if it has no explicit constraints ($m = p = 0$)

example:

$$\text{minimize} \quad f_0(x) = -\sum_{i=1}^{k} \log(b_i - a_i^T x)$$

is an unconstrained problem with implicit constraints $a_i^T x < b_i$

Now to say that $b_i - a_i^T x$ is positive says that there's a positive slack in the inequalities $b_i < a_i^T x$ like that. So the set of x that satisfies this is an open polyhedron. Now on that open polyhedron, this function here is convex, smooth, everything.

Feasibility problem

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

can be considered a special case of the general problem with $f_0(x) = 0$:

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

- $p^\star = 0$ if constraints are feasible; any feasible x is optimal
- $p^\star = \infty$ if constraints are infeasible

Let's talk about the feasibility problem. One variation on the optimization problem is this. This just says there is no objective, and what that means is that any x that satisfies these constraints is equally attractive. That's called a feasibility Problem. You make it minimize the constant objective function zero subject to, and then these inequality constraints.

So what it means is this: the optimal value of this problem is either zero or plus infinity. If there is a feasible point, then – any feasible point here is optimal because if I have a feasible point, I could ask what is its objective value, and the answer would be zero, but that's as small as the objective value gets among feasible points.

If there is one, that's p*. Therefore, any feasible point is optimal here. On the other hand, if it's infeasible then the p* is you take the infimum of zero over the empty set, and that's plus infinity so everything works out. The optimal set here coincides with the feasible set, so basically in this problem, in this formulation if you're feasible, you're optimal.

Convex optimization problem

standard form convex optimization problem

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

- $f_0, f_1, \dots, f_m$ are convex; equality constraints are affine
- problem is *quasiconvex* if f_0 is quasiconvex (and $f_1, \dots, f_m$ convex)

often written as

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

important property: feasible set of a convex optimization problem is convex

what a convex optimization problem is? There's a restriction, and the restriction is this: the objective must be convex function and the inequality constraints must also be convex. The equality constraints must be affine. And that means I can write them this way. we can also just write it as Ax = b in very compact notation

Now if the objective is quasi-convex, then this is called a quasi-convex optimization problem, and you might just write it out – this is where I've just substituted for the equality constraints a single matrix vector equality constraint. Now the feasible set of a convex optimization problem is convex.

Convex optimization problem

standard form convex optimization problem

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

- $f_0, f_1, \ldots, f_m$ are convex; equality constraints are affine
- problem is *quasiconvex* if f_0 is quasiconvex (and $f_1, \ldots, f_m$ convex)

often written as

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

important property: feasible set of a convex optimization problem is convex

Now I can say what a convex optimization problem so here it is. There's a restriction: the objective must be convex function and the inequality constraints must also be convex. The equality constraints, must be affine. And that means I can write them this way. I can also just write it as **Ax = b** in compact notation.

Now if the objective is quasi-convex, then this is called a quasi-convex optimization problem, and you might just write it out – this is where you substituted for the equality constraints a single matrix vector equality constraint.

Convex optimization problem

standard form convex optimization problem

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

- $f_0, f_1, \ldots, f_m$ are convex; equality constraints are affine
- problem is *quasiconvex* if f_0 is quasiconvex (and $f_1, \ldots, f_m$ convex)

often written as

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

important property: feasible set of a convex optimization problem is convex

"What's a convex problem?" you would say this: it's minimizing a convex function over a convex set. What could be more natural? And in fact, that's called the abstract of the loose definition . And in fact, it's okay but I want to point out here that for us an attribute– so to say it's a convex optimization problem, it is not an attribute of the feasible set. It's an attribute of the problem description.

So there's other ways to write Ax = b, and they will not work with these things. For example, we could write this. we could write || Ax – b || = 0. You have to admit if || Ax – b || = 0, Ax = b. So We haven't changed anything here. That's not a convex problem in the way we'll describe it, even though it's equivalent to one with Ax = b. Whether or not you're a convex optimization problem is an attribute of the description of the problem and not for example the feasible set

Example

$$\begin{array}{ll} \text{minimize} & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} & f_1(x) = x_1/(1+x_2^2) \le 0 \\ & h_1(x) = (x_1+x_2)^2 = 0 \end{array}$$

- f_0 is convex; feasible set $\{(x_1, x_2) \mid x_1 = -x_2 \le 0\}$ is convex
- not a convex problem (according to our definition): f_1 is not convex, h_1 is not affine
- equivalent (but not identical) to the convex problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & x_1 \le 0 \\ & x_1 + x_2 = 0 \end{array}$$

Here's a simple example. Let's scan this problem for convexity and see what happens. The first constraint – it's nothing but x_1 = 0 and 1 + x_2^2 is positive. f_1 is not a convex function, and so this problem is rejected. It is not convex. And the second one also is not convex because – that is indeed a convex function of x_1 and x_2, but you're not allowed to have convex function equals zero.

You can have a convex function less than or equal to zero. So this – I would have to call h_1 that $f_2 = (x_1+x_2)^2 \le 0$ now. That would've been fine. That'd be a convex problem because you have a convex function here less than or equal to zero.

Example

$$\begin{array}{ll} \text{minimize} & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} & f_1(x) = x_1/(1 + x_2^2) \leq 0 \\ & h_1(x) = (x_1 + x_2)^2 = 0 \end{array}$$

- f_0 is convex; feasible set $\{(x_1, x_2) \mid x_1 = -x_2 \leq 0\}$ is convex
- not a convex problem (according to our definition): f_1 is not convex, h_1 is not affine
- equivalent (but not identical) to the convex problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & x_1 \leq 0 \\ & x_1 + x_2 = 0 \end{array}$$

But the point is here is you take these and you rewrite it in an equivalent way. By the way, the problem – these are not identical problems. The problems are identical only if the objective functions and constraint functions are identical. Then the two problems are identical.

However, they're equivalent. Equivalent means that by solving one, you can construct a solution of the other and vice versa

Local and global optima

any locally optimal point of a convex problem is (globally) optimal

proof: suppose x is locally optimal, but there exists a feasible y with $f_0(y) < f_0(x)$

x locally optimal means there is an $R > 0$ such that

$$z \text{ feasible}, \quad \|z - x\|_2 \le R \quad \Longrightarrow \quad f_0(z) \ge f_0(x)$$

consider $z = \theta y + (1 - \theta)x$ with $\theta = R/(2\|y - x\|_2)$

- $\|y - x\|_2 > R$, so $0 < \theta < 1/2$
- z is a convex combination of two feasible points, hence also feasible
- $\|z - x\|_2 = R/2$ and

$$f_0(z) \le \theta f_0(y) + (1 - \theta)f_0(x) < f_0(x)$$

which contradicts our assumption that x is locally optimal

Let's say a little bit about local and global optima. If you have an optimization problem – it says if you were to restrict your search – you have a point which is a candidate for being optimal, and it says if you restrict your search arbitrarily closely, locally – but if you do a full search in there and find that there's no better point locally, you can make the conclusion from having examined this R ball, it can be as small– you can make conclusion that, there'd be nothing better.

Suppose you have a point that's locally optimal but not globally optimal. What that means is there's a point – another point that's optimal – in fact, you don't even need the point to be optimal. So you just need a point that's feasible and has the objective value better than your locally optimal point.

Local and global optima

any locally optimal point of a convex problem is (globally) optimal

proof: suppose x is locally optimal, but there exists a feasible y with $f_0(y) < f_0(x)$

x locally optimal means there is an $R > 0$ such that

$$z \text{ feasible}, \quad \|z - x\|_2 \leq R \quad \Longrightarrow \quad f_0(z) \geq f_0(x)$$

consider $z = \theta y + (1 - \theta)x$ with $\theta = R/(2\|y - x\|_2)$

- $\|y - x\|_2 > R$, so $0 < \theta < 1/2$
- z is a convex combination of two feasible points, hence also feasible
- $\|z - x\|_2 = R/2$ and

$$f_0(z) \leq \theta f_0(y) + (1 - \theta) f_0(x) < f_0(x)$$

which contradicts our assumption that x is locally optimal

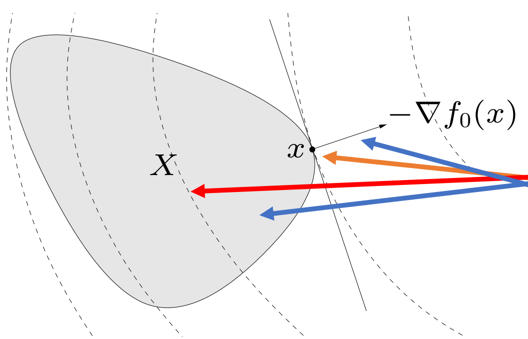
So suppose that happens. Then it's very simple what to do. You simply construct the line segment between your local optimum and this better but feasible point. You construct the line segment, and you start moving towards – from where you are locally optimal to this point that's better. What happens is – of course, as you move on that line, you remain feasible because x is feasible, y is feasible, the feasible set is convex. Therefore, all along that line segment you will be feasible.

Then what can you say? Well, now you have a convex function that basically is locally optimal at first, but then later actually achieves a value lower. And of course, that's impossible. So that's the idea. It's very simple to show this.

Optimality criterion for differentiable f_0

x is optimal if and only if it is feasible and

$$\nabla f_0(x)^T(y - x) \geq 0 \quad \text{for all feasible } y$$



if nonzero, $\nabla f_0(x)$ defines a supporting hyperplane to feasible set X at x

Let's look at an optimality criterion. A point x is optimal – this is for a convex optimization problem – if and only if it's feasible, and the following is true: it says that the gradient of the objective evaluated at the point $(x)^T(y – x)$ = 0 for all feasible y.

Now geometrically, it means this. Here's your feasible set here. These are the level curves of the objective. So this is clearly the optimal point. It's the point in the feasible set with the lowest objective value. You write the negative gradient of the objective function. That would be the normal to the hyperplane here, the local hyperplane of approximately constant objective like this and it says that this hyperplane here actually supports the feasible set at that point. It looks like that.

Optimality criterion for differentiable f_0

- **unconstrained problem**: x is optimal if and only if

$$x \in \mathbf{dom}\, f_0, \qquad \nabla f_0(x) = 0$$

- **equality constrained problem**

$$\text{minimize} \quad f_0(x) \quad \text{subject to} \quad Ax = b$$

x is optimal if and only if there exists a ν such that

$$x \in \mathbf{dom}\, f_0, \qquad Ax = b, \qquad \nabla f_0(x) + A^T\nu = 0$$

- **minimization over nonnegative orthant**

$$\text{minimize} \quad f_0(x) \quad \text{subject to} \quad x \succeq 0$$

x is optimal if and only if

$$x \in \mathbf{dom}\, f_0, \qquad x \succeq 0, \qquad \begin{cases} \nabla f_0(x)_i \geq 0 & x_i = 0 \\ \nabla f_0(x)_i = 0 & x_i > 0 \end{cases}$$

Let's look at an unconstrained problem. What would mean to say if it's unconstrained, then this is just basically for all y in R^n
$\nabla f_0(x)^T (y-x) \geq 0$ for all $y \in R^n \Longrightarrow g^T z \geq 0 \;\; \forall z \in R^n \Longrightarrow z=-g$
What's your conclusion? $g = 0$
Try that z, $g^T z \geq 0 \Longrightarrow z=-g$ and you'll get your conclusion real fast because you get minus norm $g^2 = 0$. g has to be zero. So, you remember picture in the last foil, the condition is that the gradient is zero.

Equality constraint, that – we can also get this and this will recover these things like Lagrange multipliers which you've probably seen earlier and you will see again. So you want to minimize $f_0(x)$ subject to $Ax = b$. Well, x is optimal if and only if –the condition is this: you're feasible and the gradient plus $A^T\nu = 0$.

Optimality criterion for differentiable f_0

- **unconstrained problem**: x is optimal if and only if

$$x \in \mathbf{dom}\, f_0, \qquad \nabla f_0(x) = 0$$

- **equality constrained problem**

$$\text{minimize} \quad f_0(x) \quad \text{subject to} \quad Ax = b$$

x is optimal if and only if there exists a ν such that

$$x \in \mathbf{dom}\, f_0, \qquad Ax = b, \qquad \nabla f_0(x) + A^T \nu = 0$$

- **minimization over nonnegative orthant**

$$\text{minimize} \quad f_0(x) \quad \text{subject to} \quad x \succeq 0$$

x is optimal if and only if

$$x \in \mathbf{dom}\, f_0, \qquad x \succeq 0, \qquad \begin{cases} \nabla f_0(x)_i \geq 0 & x_i = 0 \\ \nabla f_0(x)_i = 0 & x_i > 0 \end{cases}$$

Now to show a quick plausibility argument, not give the whole argument for that, but I'll show that.
$\nabla f_0(x)^T (z-x) \geq 0$ $\forall z$ and $Az=b$
we also know that $Ax = b$, and what we see is that it means $z - x$ is in the null space of A, i.e., $(z-x) \in \mathfrak{N}(A)$.
The gradient has non-negative inner product with everything in the null space of A, $\nabla f_0(x) \in \mathfrak{N}(A)^{\perp}$
It means it's orthogonal to the null space of A.

$\nabla f_0(x) \in R(A^T)$
$\nabla f_0(x) = A^T u$
$\nabla f_0(x) + A^T(-u) = 0$
That means that grad $f_0(x) = A^T$ times some u, and I could write this as grad $f_0(x)$ – I can write plus $A^T(-u) = 0$ for some u. That's the condition.

Optimality criterion for differentiable f_0

- **unconstrained problem**: x is optimal if and only if

$$x \in \mathbf{dom}\, f_0, \qquad \nabla f_0(x) = 0$$

- **equality constrained problem**

$$\text{minimize} \quad f_0(x) \quad \text{subject to} \quad Ax = b$$

x is optimal if and only if there exists a ν such that

$$x \in \mathbf{dom}\, f_0, \qquad Ax = b, \qquad \nabla f_0(x) + A^T\nu = 0$$

- **minimization over nonnegative orthant**

$$\text{minimize} \quad f_0(x) \quad \text{subject to} \quad x \succeq 0$$

x is optimal if and only if

$$x \in \mathbf{dom}\, f_0, \qquad x \succeq 0, \qquad \begin{cases} \nabla f_0(x)_i \geq 0 & x_i = 0 \\ \nabla f_0(x)_i = 0 & x_i > 0 \end{cases}$$

So, what are the conditions under which – when would a point x in the non-negative orthant minimize a convex function?
$\nabla f(x)^T (z-x) \geq 0 \ \forall z \geq 0$. That's our basic condition.
And let's simplify this now.
$\nabla f(x)^T x \leq 0$ and $\nabla f(x) \leq 0$
This has to be positive for any non-negative z here. I can plug in z = 0, and I get the following that
$\nabla f(x)^T x = 0$. That's from z = 0.
$\nabla f(x)_I x_i = 0$. So we can also conclude that this vector is positive.

Equivalent convex problems

two problems are (informally) **equivalent** if the solution of one is readily obtained from the solution of the other, and vice-versa

some common transformations that preserve convexity:

- **eliminating equality constraints**

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

is equivalent to

$$\begin{array}{ll} \text{minimize (over } z\text{)} & f_0(Fz + x_0) \\ \text{subject to} & f_i(Fz + x_0) \leq 0, \quad i = 1, \ldots, m \end{array}$$

where F and x_0 are such that

$$Ax = b \iff x = Fz + x_0 \text{ for some } z$$

Next topic is equivalent convex problems. You say that they're equivalent if the solution of one is obtained from the solution of another. Here's one. You could eliminate equality constraints. So, you minimize $f_0(x)$ subject to $f_i(x)$ = 0 and Ax = b. I can eliminate these, and the way I do that is I find a matrix F and a point z so that Ax = b is equivalent to having the form Fz + x_0 for some z.

Now how do you get F? F is like the null space. The columns of A would be a span for the null. They would span the null space of A. And then x_0 is any particular solution of Ax = b. For example, it could be the least norms solution. So this is from Ax to F and x_0. This is parameterized by z which is now free. If it's a minimal one.

Equivalent convex problems

two problems are (informally) **equivalent** if the solution of one is readily obtained from the solution of the other, and vice-versa

some common transformations that preserve convexity:

- **eliminating equality constraints**

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

is equivalent to

$$
\begin{array}{ll}
\text{minimize (over } z\text{)} & f_0(Fz + x_0) \\
\text{subject to} & f_i(Fz + x_0) \le 0, \quad i = 1, \dots, m
\end{array}
$$

where F and x_0 are such that

$$
Ax = b \quad \Longleftrightarrow \quad x = Fz + x_0 \text{ for some } z
$$

I'll take x = Fz + x_0 and I get this problem here. I don't need this constraint because basically A(Fz + x_0) = b for all z automatically because AF is zero. That's how you constructed F, and Ax_0 is b. That's how you constructed x0. So you don't need this constraint. You drop the constraint, and you get this. Now these are not the same problem. I mean they're totally different. They have dimension in variables. This one has equality constraints. This has none. But they're clearly equivalent, you would solve this problem and then return as x, Fz + x_0.
Now the idea that it has to be a transformation to get A and b, it has to require computation to calculate F and x_0, and indeed it does.

Equivalent convex problems

- **introducing equality constraints**

$$\begin{array}{ll}\text{minimize} & f_0(A_0x + b_0)\\ \text{subject to} & f_i(A_ix + b_i) \le 0, \quad i = 1, \dots, m\end{array}$$

is equivalent to

$$\begin{array}{ll}\text{minimize (over } x,\ y_i) & f_0(y_0)\\ \text{subject to} & f_i(y_i) \le 0, \quad i = 1, \dots, m\\ & y_i = A_ix + b_i, \quad i = 0, 1, \dots, m\end{array}$$

- **introducing slack variables for linear inequalities**

$$\begin{array}{ll}\text{minimize} & f_0(x)\\ \text{subject to} & a_i^Tx \le b_i, \quad i = 1, \dots, m\end{array}$$

is equivalent to

$$\begin{array}{ll}\text{minimize (over } x,\ s) & f_0(x)\\ \text{subject to} & a_i^Tx + s_i = b_i, \quad i = 1, \dots, m\\ & s_i \ge 0, \quad i = 1, \dots m\end{array}$$

Introduce equality constraints, which by the way some people call unelimination. You could do the following. I'll introduce a new variable y_0 here. I'll introduce a new variable y_i here. And I'll have this problem. Minimize over x and now also y_i, so I've added new variables – $f_0(y_0)$ subject to $fi(y_i) = 0$, and then these equality constraints. The problem started with no equality constraints. After this unelimination step, you have added variables and added equality constraints.

Another trick is to introduce slack variables for linear inequalities. Now that works like this. You have $a_i^Tx = b_i$. That's fine. What I'll do is I'll convert that to an equality constraint. I'll convert that to $a_i^Tx + s_i = b_i$. This is the same as saying s_i is the slack. It's $b_i - a_i^Tx$, and then I'll write that s_i are bigger than or equal to zero.

Equivalent convex problems

- **epigraph form**: standard form convex problem is equivalent to

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

- **minimizing over some variables**

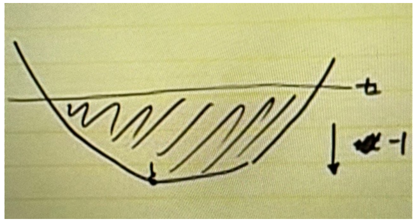
$$\begin{array}{ll} \text{minimize} & f_0(x_1, x_2) \\ \text{subject to} & f_i(x_1) \leq 0, \quad i = 1, \ldots, m \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x_1) \\ \text{subject to} & f_i(x_1) \leq 0, \quad i = 1, \ldots, m \end{array}$$

where $\tilde{f}_0(x_1) = \inf_{x_2} f_0(x_1, x_2)$

You want to minimize f_0. You introduce a new scalar variable t, and you introduce this new constraint, which is $f_0(x) = t$, and you minimize t. Now if you solve this problem, first of all you choose any t other than $f_0(x) = t$. The point because your goal is to minimize t.
It is an equivalent problem to write equals here. That's equivalent to the original problem.
Unfortunately, it's not convex unless f_0 is affine.



Let me just draw this. Here's a function that you minimize like this.

says find the point here that minimizes f_0. This introduces a new t that's up here like that, and it says that's your feasible set or something like that, and then you can in the direction –t. I guess the direction is -1, and you would recover this point here. By the way, this has immediate practical

Equivalent convex problems

- **epigraph form**: standard form convex problem is equivalent to

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

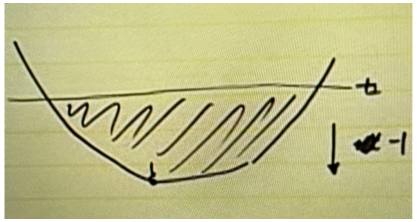
- **minimizing over some variables**

$$
\begin{array}{ll}
\text{minimize} & f_0(x_1, x_2) \\
\text{subject to} & f_i(x_1) \le 0, \quad i = 1, \ldots, m
\end{array}
$$

is equivalent to

$$
\begin{array}{ll}
\text{minimize} & \tilde{f}_0(x_1) \\
\text{subject to} & f_i(x_1) \le 0, \quad i = 1, \ldots, m
\end{array}
$$

where $\tilde{f}_0(x_1) = \inf_{x_2} f_0(x_1, x_2)$



The original problem says find the point here that minimizes f_0.

This introduces a new t that's up here like that, and it says that's your feasible set or something like that, and then you can in the direction –t. The direction is -1, and you would recover this point here. By the way, this has immediate practical application. This says for example that if you write a code to solve convex problems, without any loss of generality, you can just minimize a linear objective.

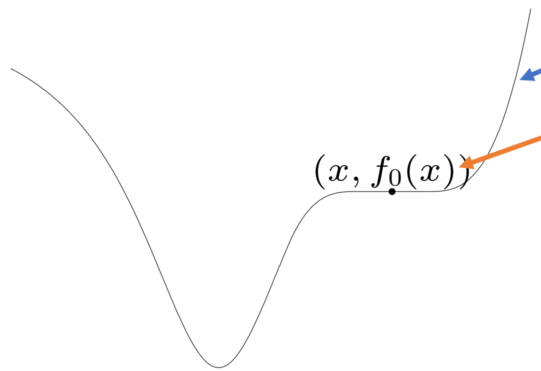
A couple of other equivalences – here's one. You can minimize over some variables, so if you have a problem that says minimize $f_0(x_1,x_2)$ subject to $f_i(x_1) = 0$ for example, I could minimize this thing over x_2, and that would reduce – that would give you this function, f_0 tilde of x_1 .

Quasiconvex optimization

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

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

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



Quasi-convex optimization –here f_0 is quasi-convex. There are lots of practical problems that are quasi-convex. And for example this most basic thing which says a locally optimal is globally optimal is false. And here's a perfect example. I'm making it exactly flat here.
So that's a locally optimal point. And it is obviously not the global optimum here.
And that's a perfectly valid quasi-convex function.

Quasiconvex optimization

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.*,

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

example

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

with p convex, q concave, and $p(x) \ge 0$, $q(x) > 0$ on $\mathbf{dom}\, f_0$

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

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

If this were a convex function of x and t, that would be the proof that this function is convex in x. It's not convex. It's quasi-convex. That means t consider fixed, and is the set of x that satisfies this convex. That's the question.

So if you look at this, well you just multiply through, and you get t= $\frac{p(x)}{q(x)}$ and p(x) = tq(x).
Write that as p(x) – tq(x) = 0. That's convex.

That's concave. t is bigger than or equal to zero without loss of generality because if t is negative, the sublevel set is empty and convex. Then this is convex minus non-negative multiple of concave. This whole thing is convex. That's a convex representation.

Quasiconvex optimization

quasiconvex optimization via convex feasibility problems

$$\phi_t(x) \le 0, \qquad f_i(x) \le 0, \quad i = 1, \dots, m, \qquad Ax = b \qquad (1)$$

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

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)

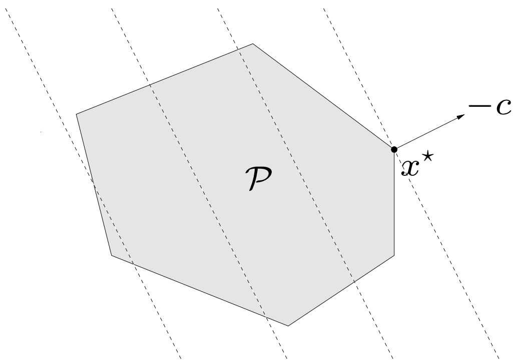
you start with a lower and an upper bound on p*. By the way, you might not have that, in which case sort of a practical version of this would start with t = 1 and query. And if that is infeasible, you would then try t = 2, then four, then eight, then 16 or something like that.

you start with a known upper and lower bound on p*, and you simply query at the midpoint t of your known upper and lower bounds, and you solve the convex feasibility problem. So this is just bisection. And every time you end up with two points, a l and a u. For u, you have an x at most the value $f_0(x)$ = u. And l on the other hand is a value known to be infeasible. There's no x that satisfies $f_0(x)$ = l. Anyway, and then at each step divide s by two, so halves at each iteration. So it takes log steps, and you reduce iteration steps

Linear program (LP)

$$\begin{array}{ll} \text{minimize} & c^T x + d \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

- convex problem with affine objective and constraint functions
- feasible set is a polyhedron



A linear program is almost sort of the simplest optimization problem there is in one sense. All of the functions involved are affine. Everything is affine. The objective is affine. The inequality constraints are affine. And the constraints are affine, again called linear constraints. It depends on whether when you say affine or not – linear, you allow an arbitrary right-hand side. If you allow an arbitrary right-hand side, then of course when someone says linear it means affine.

So the feasible set of course is a polyhedron, so your image of what an LP is should be this. It's a polyhedron, and then the objective is affine, so I mean the level sets of an affine function are hyperplanes, so you're just – here's negative c, and you're told go as far as you can in this direction. And this is a little example makes it totally clear that's the solution

Examples

diet problem: choose quantities $x_1, \ldots, x_n$ of n foods

- one unit of food j costs c_j, contains amount a_{ij} of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

to find cheapest healthy diet,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \succeq b, \quad x \succeq 0 \end{array}$$

piecewise-linear minimization

$$\text{minimize} \quad \max_{i=1,\ldots,m}(a_i^T x + b_i)$$

equivalent to an LP

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

Let's look at some examples. So you're supposed to choose a bunch of foods, so you have n possible food types, and you're going to choose some quantities that you're going to feed some group of people.

One unit of food costs – you have these costs which are c_j, and each one contains a_{ij} of some nutrient. you minimize $c^T x$. Now $c^T x$ here is literally the cost because x_i is the quantity of each type of food. And then when you take the inner product of c, that gives you the actual cost.So that's minimizing the cost subject to Ax = b. A is this matrix here. It's the nutrient matrix, and then b is this minimum daily allowed. So this is a famous problem. x has to be positive. You can't feed people negative amounts or whatever.

Examples

diet problem: choose quantities $x_1, \ldots, x_n$ of n foods

- one unit of food j costs c_j, contains amount a_{ij} of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

to find cheapest healthy diet,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \succeq b, \quad x \succeq 0 \end{array}$$

piecewise-linear minimization

$$\text{minimize} \quad \max_{i=1,\ldots,m}(a_i^T x + b_i)$$

equivalent to an LP

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

Here's one: piecewise-linear minimization. So you want to minimize a piecewise – the maximum of a bunch of affine functions. So you just want to minimize that over x.
It's also quite useful because piecewise-linear functions can be used to approximate any convex function. So just minimize some piecewise-linear function. It's an LP. You just put in epigraph form. So you just write it this way. I'm going to minimize a new variable t subject to $a_i^T x + b_i = t$.

These are affine linear inequalities, or to use the term that most people would use, these are linear inequalities in x and t.

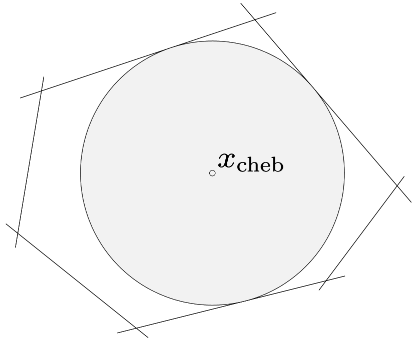
Chebyshev center of a polyhedron

Chebyshev center of

$$\mathcal{P} = \{x \mid a_i^T x \leq b_i,\ i = 1, \ldots, m\}$$

is center of largest inscribed ball

$$\mathcal{B} = \{x_c + u \mid \|u\|_2 \leq r\}$$



- $a_i^T x \leq b_i$ for all $x \in \mathcal{B}$ if and only if

$$\sup\{a_i^T(x_c + u) \mid \|u\|_2 \leq r\} = a_i^T x_c + r\|a_i\|_2 \leq b_i$$

- hence, x_c, r can be determined by solving the LP

$$\begin{array}{ll} \text{maximize} & r \\ \text{subject to} & a_i^T x_c + r\|a_i\|_2 \leq b_i, \quad i = 1, \ldots, m \end{array}$$

We'll look at another one, and it's the Chebyshev center of a polyhedron. We talked a little bit about the idea of yield maximization. So yield maximization was this problem.
Yield maximization was that we had some polyhedron of acceptable values of some parameters. We're going to set a target for manufacturing. So we're going to set our machines to attempt to manufacture these points. In manufacture however, you will add a random perturbation to this point, and so what you'll get is you'll get a PDF that is centered – has mean value at this point. And then we're interested in maximizing the probability that you're in this set. That would be the yield