1 Bounding the value function

Last time, we looked at what happens when we relax or tighten the constraints on a convex program: changing

\[
(P) \quad \begin{cases} 
\min_{x \in S} & f(x) \\
\text{subject to} & g(x) \leq 0
\end{cases}
\]

to

\[
(P(z)) \quad \begin{cases} 
\min_{x \in S} & f(x) \\
\text{subject to} & g(x) \leq z
\end{cases}
\]

The value function \(MP(z)\) gives us (more or less) the optimal value of this perturbed variant of \(P\), in terms of \(z\).

Intuitively (we’ll make this precise later) knowing that \(MP\) is a convex function tells us that it cannot decrease too quickly. So if we make \(P\) an unconstrained program, allowing \(x\) to violate the constraint \(g(x) \leq 0\), then a small violation in the constraint can only yield a small improvement. If we modify the objective function \(f(x)\) to take that into account, we can forget about the constraints at all, and then everything is easy.

That’s the plan. To make it work, we need two technical lemmas:

**Lemma 1.1** (Sensitivity vector lemma). Let \(P\) be a convex program with a point \(x^* \in S\) such that \(g(x^*) < 0\). (This is called the Slater condition.)

Then there is some \(\lambda \in \mathbb{R}^m\), with \(\lambda \geq 0\), such that

\[
MP(z) \geq MP(0) - \lambda \cdot z.
\]

A vector \(\lambda \in \mathbb{R}^m\) with \(\lambda \geq 0\) satisfying this inequality is called a sensitivity vector: it measures the sensitivity of \(P\) to changes in the constraints.

**Lemma 1.2** ("Moving the goalposts" lemma). Let \(P\) be any nonlinear program. For any \(x^{(0)} \in S\),

(a) \(x^{(0)}\) is feasible for \(P(g(x^{(0)}))\), and

(b) \(MP(g(x^{(0)})) \leq f(x^{(0)})\).
The first lemma, when it applies, relates the original problem $P$ to the perturbed problems $P(z)$, which justifies thinking about them to begin with.

The second lemma says that any element of $S$, even if it’s not feasible for $P$, still gives a bound on $MP(z)$ for some $z$, and therefore (by the first lemma) it still gives a bound on $MP(0)$, the minimum value of $P$.

1.1 Proving the sensitivity vector lemma

Proof. The condition that there is some $x^* \in S$ such that $g(x^*) < 0$ is called the Slater condition. It tells us that 0 is an interior point of the domain of $MP$: we can make the constraints of $P$ a little tighter, and $x^*$ will stay feasible.

If 0 is an interior point, then (as we showed in last Friday’s lecture) $MP$ has a subgradient at 0, which is exactly the inequality

$$MP(z) \geq MP(0) - \lambda \cdot z.$$ 

Usually, we’d write that with a +; here, we write it with a −, because then we can show that $\lambda \geq 0$.

To see this, take $z = e(i)$. Then the subgradient inequality becomes

$$MP(e(i)) \geq MP(0) - \lambda_i$$

but on the other hand we have $MP(0) \geq MP(e(i))$: changing the $i^{th}$ constraint from $g_i(x) \leq 0$ to $g_i(x) \leq 1$ can’t make the minimum value larger, only smaller. So $MP(0) \geq MP(0) - \lambda_i$, which means $\lambda_i \geq 0$. 

1.2 Proving the “Moving the goalposts” lemma

Proof. Part (a) is saying “$x$ becomes feasible if we change the constraints to make $x$ feasible”. The program $P(g(x(0)))$ is

$$(P(g(x(0)))) \begin{cases} \text{minimize} & f(x) \\ x \in S & \text{subject to} & g(x) \leq g(x(0)) \end{cases}$$

and setting $x$ to $x(0)$ makes the constraint be $g(x(0)) \leq g(x(0))$, which is true.

We defined $MP(g(x(0)))$ to be the greatest lower bound on $f(x)$ for all $x$ which are feasible for $P(g(x(0)))$. One such $x$ is $x(0)$ itself, so $MP(g(x(0)))$ is a lower bound on $f(x(0))$, which is part (b) of the lemma.

2 The Karush–Kuhn–Tucker theorem, saddle point form

For $x \in S$ and $\lambda \in \mathbb{R}^m$, we define the Lagrangian $L(x, \lambda)$ by

$$L(x, \lambda) = f(x) + \lambda \cdot g(x) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x).$$
We will state the Karush–Kuhn–Tucker theorem slightly differently from the way it is stated in the textbook. The textbook states it for all convex programs satisfying the Slater condition: convex programs with a point satisfying all constraints strictly (with $a <$ in place of $a \leq$).

However, we only need this hypothesis to apply Lemma 1.1 and get a sensitivity vector: a vector $\lambda \in \mathbb{R}^m$, with $\lambda \geq 0$, such that

$$MP(z) \geq MP(0) - \lambda \cdot z,$$

So we’ll skip that step and just ask directly: do we have such a $\lambda$?

**Theorem 2.1** (Karush–Kuhn–Tucker theorem, saddle point form). Let $P$ be any nonlinear program. Suppose that $x^* \in S$ and $\lambda^* \geq 0$. Then $x^*$ is an optimal solution of $P$ and $\lambda^*$ is a sensitivity vector for $P$ if and only if:

1. $L(x^*, \lambda^*) \leq L(x, \lambda^*)$ for all $x \in S$. (Minimality of $x^*$)
2. $L(x^*, \lambda^*) \geq L(x^*, \lambda)$ for all $\lambda \geq 0$. (Maximality of $\lambda^*$)
3. $\lambda^*_i g_i(x^*) = 0$ for $i = 1, 2, \ldots, m$. (Complementary slackness)

### 2.1 Proving that these conditions are necessary

Let $x^*$ be an optimal solution of $P$ and let $\lambda^*$ be a sensitivity vector.

Take any $x \in S$. By Lemma 1.2, $MP(g(x)) \leq f(x)$. Using the sensitivity vector inequality, we get

$$f(x) \leq MP(g(x)) \geq MP(0) - \lambda^* \cdot g(x).$$

But $MP(0)$ is just $f(x^*)$, so we can rewrite this as

$$f(x^*) \leq f(x) + \lambda^* \cdot g(x) = L(x, \lambda^*).$$

This is how the Lagrangian enters the picture. In particular, setting $x^* = x$, we get

$$f(x^*) \leq f(x^*) + \lambda^* \cdot g(x^*),$$

so $\lambda^* \cdot g(x^*) \geq 0$.

But we have $g(x^*) \leq 0$ (since $x^*$ is feasible for $P$) and $\lambda^* \geq 0$ (by assumption). So the dot product $\lambda^* \cdot g(x^*)$ is a sum of nonpositive products $g_i(x^*) \leq \lambda^*_i$. In particular, it’s always $\leq 0$, and the only way it can also be $\geq 0$ is if each of these products is $0$.

This gives us the complementary slackness condition, and says that $\lambda^* \cdot g(x^*) = 0$.

Going back to the inequality $f(x^*) \leq L(x, \lambda^*)$: now that we know that $\lambda^* \cdot g(x^*) = 0$, we can add it to the left-hand side, and get $L(x^*, \lambda^*) \leq L(x, \lambda^*)$. This proves condition 1.

Finally, we have $g(x^*) \cdot \lambda \leq 0$ for any $\lambda \geq 0$, since (once again) this is a sum of nonpositive products. Since $\lambda^* \cdot g(x^*) = 0$, we have

$$g(x^*) \cdot \lambda \leq g(x^*) \cdot \lambda^*$$

and adding $f(x^*)$ to both sides gives $L(x^*, \lambda) \leq L(x^*, \lambda^*)$, proving condition 2.
2.2 Proving that these conditions are sufficient

First, we use the three conditions to prove that \( x^* \) is feasible: that \( g_i(x^*) \leq 0 \) for \( i = 1, 2, \ldots, m \). To do this, take \( \lambda = \lambda^* + e^{(i)} \): we get

\[
L(x^*, \lambda) = f(x^*) + g(x^*) \cdot (\lambda^* + e^{(i)}) = f(x^*) + g x^* \cdot \lambda^* + g_i(x^*) = L(x^*, \lambda^*) + g_i(x^*).
\]

But condition 2 tells us that \( L(x^*, \lambda) \leq L(x^*, \lambda^*) \), so we must have \( g_i(x^*) \leq 0 \).

Next, we prove that \( x^* \) is optimal. Take any feasible \( x \); that is, any \( x \in S \) with \( g(x) \leq 0 \); then

\[
f(x) \geq f(x) + \sum_{i=1}^{m} \lambda^*_i g_i(x) = L(x, \lambda^*) \quad \text{(since each term in the sum is } \leq 0)\]

\[
\geq L(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^{m} \lambda^*_i g_i(x^*) \quad \text{(by condition 1)}\]

\[
= f(x^*) \quad \text{(by condition 3, since } \lambda^*_i g_i(x^*) = 0 \text{ for all } i)\]

This shows that \( f(x^*) \leq f(x) \) for any feasible \( x \), making it optimal.

Finally, we prove that \( \lambda^* \) is a sensitivity vector. To do this, we must show that for all \( z \) in the domain of MP,

\[
MP(z) \geq MP(0) - \lambda^* \cdot z.
\]

Pick a vector \( x \) feasible for \( P(z) \); that is, \( x \in S \), and \( g(x) \leq z \). Then \( \lambda^* \cdot g(x) \leq \lambda^* \cdot z \).

Condition 2 tells us that \( f(x^*) = L(x^*, \lambda^*) \leq L(x, \lambda^*) \), so

\[
f(x^*) \leq f(x) + \lambda^* \cdot g(x) \leq \lambda^* \cdot z
\]

which we can rearrange to get

\[
f(x) \geq f(x^*) - \lambda^* \cdot z = MP(0) - \lambda^* \cdot z.
\]

That is, \( MP(0) - \lambda^* \cdot z \) is a lower bound on \( f(x) \) for any \( x \) feasible for \( P(z) \). In the meantime, \( MP(z) \) is by definition the greatest lower bound on \( f(z) \) for any \( x \) feasible for \( P(z) \). Therefore

\[
MP(z) \geq MP(0) - \lambda^* \cdot z
\]

which shows that \( \lambda^* \) is a sensitivity vector and completes the proof.