Linear inequality constraints (12/1)

A minimization problem with a nonlinear cost function and linear inequality constraints;

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } Ax \geq b \end{aligned} \qquad (*)$$

Note: this covers the case where there are linear equality constraints by our reasoning last lecture.

Assume $A \in \mathbb{R}^{m \times n}$, but no longer require $m < n$. We also don't need any assumptions on the rank of A.

If x^* is a local minimum and $x^* \in \mathcal{X}^\circ = \{x \in \mathbb{R}^n : Ax < b\}$, then usual unconstrained optimization theory applies. On the other hand, if $x^* \in \partial \mathcal{X} = \{x \in \mathbb{R}^n : Ax \leq b\}$, we have $\hat{A}x^* = \hat{b}$ where $\hat{A}$ is the matrix whose rows correspond to active constraints and likewise for $\hat{b}$.

Note: p is a feasible direction at a feasible point x if $Ap \geq 0$.

(Easy to see: $A(x+p) = Ax + Ap \geq b + Ap \geq b$, only if $Ap \geq 0$.)

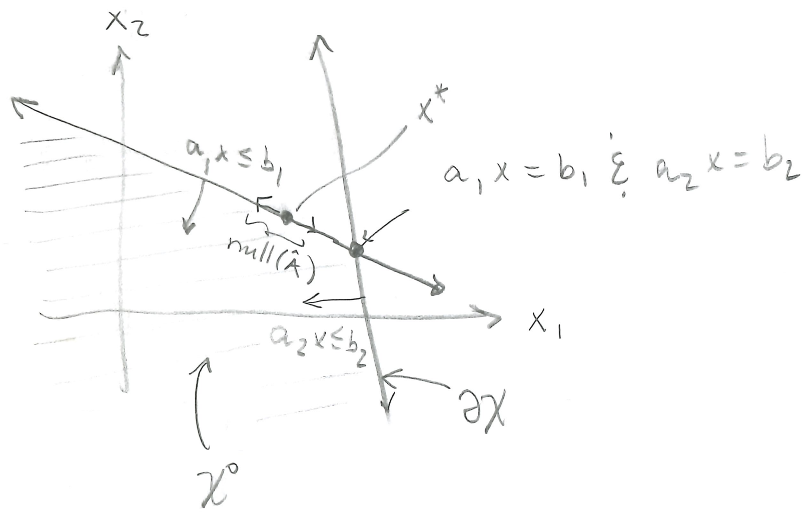
Let's try to develop conditions under which x^* is a constrained local minimum for the case where $x^* \in \partial \mathcal{X}$.

Let's assume $x^* \in \partial\mathcal{X}$ is a constrained local minimum.

Recall what this means: for any feasible direction p at x^*, we have $d_p f(x^*) \geq 0$.

The constraint $\hat{A}x^* = \hat{b}$ defines a linear subspace of feasible directions. How to see this?

E.g.: $n=2$:



If only one of these constraints is active, then the nullspace of $\hat{A}$ has dimension 1. If both constraints are active, null($\hat{A}$) $= 2 = n$, hence null($\hat{A}$) $= \{0\}$.

So: we can conclude that if x^* is a constrained local minimum, then x^* is a local minimum for:

$$(**) \quad \begin{array}{l} \text{minimize } f(x) \\ \text{subject to } x \in \text{null}(\hat{A}(x^*)) \end{array} \iff \begin{array}{l} \text{minimize } f(\bar{x} + Z(x^*)v) \\ \text{subject to } v \in \mathbb{R}^{\text{nullity}(\hat{A}(x^*))} \end{array}$$

Note the absence of the original linear inequality constraints!

Note also that this applies (generalizes) to the case where $x^* \in \mathcal{X}^\circ$, since in this case $\text{null}(\hat{A}(x^*)) = \mathbb{R}^n$.

(This partly depends on how you interpret what $\hat{A}(x^*)$ means in this case...)

Let $\hat{\lambda}$ denote the vector of Lagrange multipliers corresponding to the active constraints encoded in $\hat{A}$.

Claim: If x^* is a constrained local minimum, then $\hat{\lambda}^* \geq 0$.

Pf: Since x^* is a constrained local minimum, we have $\nabla f(x^*)^T p \geq 0$ for any feasible direction p at x^*. Since x^* is a constrained local minimum for $(**)$, our results from last lecture give $\nabla f(x^*) = \hat{A}(x^*)^T \hat{\lambda}^*$. Hence:

$$\hat{\lambda}^{*T} \hat{A}(x^*) p \geq 0.$$

Since $\hat{A}(x^*)$ has full row rank, we can choose p such that $\hat{A}(x^*) p = e_i$ for any i. Why? Full row rank implies the existence of a subset of columns equal in number to the number of rows which are linearly independent. Then:

$$0 \leq \hat{\lambda}^{*T} \hat{A}(x^*) p = \hat{\lambda}^{*T} e_i = \hat{\lambda}^*_i.$$

Since i was arbitrary, this proves the result $\blacksquare$

Now, if we let λ_i be zero if the i^{th} constraint is inactive, we see that the conditions we have just established:

$$\nabla f(x^*) = \hat{A}^T \hat{\lambda}^*, \qquad \hat{\lambda}^* \geq 0$$

are equivalent to:

$$\nabla f(x^*) = A^T \lambda^*, \qquad \lambda^* \geq 0.$$

To force $\lambda_i = 0$ for an inactive constraint, we use the familiar complementary slackness condition:

$$\lambda_i (b_i - a_i^T x) = 0 \quad \forall i.$$

In summary:

<u>Thm</u>: (Necessary conditions for linear inequality constraints)

If x^* is a constrained local minimizer of $(*)$, then:

$$\nabla f(x^*) = A^T \lambda^* \qquad \text{[stationarity]}$$
$$\lambda^* \geq 0 \qquad \text{[dual feasibility]}$$
$$(b - Ax^*)^T \lambda^* = 0 \qquad \text{[complementary slackness]}$$
$$Ax^* \geq b \qquad \text{[primal feasibility]}$$

If $\text{range}(Z) = \text{null}(A)$, then the second-order necessary condition is:

$Z^T \nabla^2 f(x^*) Z$ is positive definite.

Sufficient conditions are more complicated.

Example: Solve:

$$\text{minimize} \quad x^3 + y^2$$
$$\text{subject to} \quad -1 \leq x \leq 0$$

Consider the point $(x,y) = (0,0)$, where $\nabla f(0,0) = (0,0)$.

Note that then $\hat{A}^T \nabla f = 0 = \hat{\lambda}$. So first-order necessary conditions hold. Furthermore, since $Z = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, we have:

$$Z^T \nabla^2 f(0,0) Z = \frac{\partial^2 f}{\partial y^2} = 2 > 0.$$

So, satisfies obvious choice for 2^{nd} order sufficient conditions...

But, it is clear that $f(-\varepsilon, 0) < f(0,0)$ for any feasible $(-\varepsilon, 0)$, where $\varepsilon > 0$!

Problem here is degenerate constraint, where an active constraint has a zero Lagrange multiplier.

If either $\lambda_i = 0$ or $(Ax - b)_i = 0$, we say strict complementarity holds.

Thm (sufficient conditions for linear inequality constraints):

If x^* satisfies:

$Ax^* \geq b$ [primal feasibility]

$\nabla f(x^*) = A^T \lambda^*$ [stationarity]

$\lambda^* \geq 0$ [dual feasibility]

$\lambda_i^* = 0$ xor $(b - Ax^*)_i = 0$ [strict complementarity]

$Z^T \nabla^2 f(x^*) Z$ is positive definite

then x^* is a constrained local minimum.

Pf: Let p be a feasible direction at x^*. If $\hat{A}p = 0$, $d_p f(x^*) > 0$ since x^* is a strict local minimum for the unconstrained problem on null $\hat{A}$. Otherwise, if $\hat{A}p \geq 0$ with some components positive, p points into the interior of $\mathcal{X}$.

Then $p^T \nabla f(x^*) = p^T \hat{A}^T \hat{\lambda}^* > 0$; since th