4 Constrained Minimization

Feasible directions, tangent cone. First order conditions (polyhedral sets, convex sets). Convex programming intro.

Let C be a polyhedral set and suppose $C = \{x\colon Dx \le d\}$, for some $D \in \mathbf{R}^{m\times n}$. For a point $\bar{x} \in C$, we let $\mathcal{A}(\bar{x}) = \{i\colon D_{i\cdot}\bar{x} = d_i\}$ denote the set of *active* constraints at $\bar{x}$. A similar convention is used for $\mathcal{I}(\bar{x})$ regarding inactive constraints.

Proposition 1 *Let $K = \{z\colon D_{\mathcal{A}\cdot}z \le 0\}$. Then $K \supset C - \bar{x}$. Furthermore, there is a neighborhood U if the origin such that $U \cap (C - \bar{x}) = U \cap K$. Finally, $K = T_C(\bar{x})$.*

Proof Let $c \in C$. Then $D_{\mathcal{A}\cdot}c \le d_{\mathcal{A}}$; therefore $D_{\mathcal{A}\cdot}(c - \bar{x}) \le 0$ (since $D_{\mathcal{A}\cdot}\bar{x} = d_{\mathcal{A}}$), and therefore $c - \bar{x} \in K$; hence $C - \bar{x} \subset K$. For the second assertion, note that $U \cap K \supset U \cap (C - \bar{x})$ for evey U; therefore we prove the inclusion "$\subset$". As $D_{\mathcal{I}\cdot}\bar{x} < d_{\mathcal{I}}$, there is a neighborhood U of the origin such that for each $z \in U$, $D_{\mathcal{I}\cdot}(\bar{x} + z) \le d_{\mathcal{I}}$. Let $z \in U \cap K$. Then $D_{\mathcal{I}\cdot}(\bar{x} + z) \le d_{\mathcal{I}}$ and $D_{\mathcal{A}\cdot}(\bar{x} + z) \le d_{\mathcal{A}}$, so $\bar{x} + z \in C$ and hence $z \in U \cap (C - \bar{x})$.

Suppose $w \in T_C(\bar{x})$. Then there is some sequence $x^\nu \to \bar{x}$ in C and $\tau^\nu \downarrow 0$ such that $w^\nu = (x^\nu - \bar{x})/\tau^\nu$ satisfies $w^\nu \to w$. It follows that for large ν, $x^\nu - \bar{x} \in U \cap (C - \bar{x})$ and hence in $U \cap K$. Thus $\tau^\nu w^\nu \in K$ and hence by closure and cone properties of K so is w.

If $w \in K$, then $\mu w \in U \cap K$ for small but positive μ, hence $\tau w \in C - \bar{x}$ for all $0 < \tau \le \mu$. Just take a decreasing sequence of such τ to define x^ν and $w^\nu = w$ with the required properties. $\square$

Corollary 2 *Let R and S be polyhedral convex sets in $\mathbf{R}^n$. If $x \in R \cap S$ then $T_{R\cap S}(x) = T_R(s) \cap T_S(x)$.*

Proof Let $R = \{x\colon Dx \le d\}$ and $S = \{x\colon Gx \le g\}$. We have

$$R \cap S = \left\{x\colon \begin{bmatrix} D \\ G \end{bmatrix} x \le \begin{bmatrix} d \\ g \end{bmatrix}\right\}$$

so for appropriately defined $\mathcal{A}_d$ and $\mathcal{A}_g$ it follows

$$\begin{aligned} T_{R\cap S}(\bar{x}) &= \left\{w\colon \begin{bmatrix} D_{\mathcal{A}_d\cdot} \\ G_{\mathcal{A}_g\cdot} \end{bmatrix} w \le 0\right\} = \{w\colon D_{\mathcal{A}_d\cdot}w \le 0\} \cap \left\{w\colon G_{\mathcal{A}_g\cdot}w \le 0\right\} \\ &= T_R(\bar{x}) \cap T_S(\bar{x}). \end{aligned}$$

$\square$

Theorem 6 *(local optimality conditions on a polyhedral set). Consider the problem of minimizing* f_0 *over a polyhedral set* C*, with* f_0 *of class* $\mathcal{C}^2$*. Let* $\bar{x} \in C$*.*

(a) (necessary). *If* $\bar{x}$ *is a locally optimal solution, then*

$$\begin{gathered} \nabla f_0(\bar{x}) \cdot w \geq 0 \text{ for every } w \in T_C(\bar{x}), \\ w \cdot \nabla^2 f_0(\bar{x}) w \geq 0 \text{ for every } w \in T_C(\bar{x}) \text{ satisfying } \nabla f_0(\bar{x}) \cdot w = 0. \end{gathered}$$

(b) (sufficient). *If* $\bar{x}$ *has the property that*

$$\begin{gathered} \nabla f_0(\bar{x}) \cdot w \geq 0 \text{ for every } w \in T_C(\bar{x}), \\ w \cdot \nabla^2 f_0(\bar{x}) w > 0 \text{ for every } w \in T_C(\bar{x}) \text{ satisfying } \nabla f_0(\bar{x}) \cdot w = 0, w \neq 0, \end{gathered}$$

then $\bar{x}$ *is a locally optimal solution. Moreover, in these circumstances the local optimality of* $\bar{x}$ *is strict, in the sense that there exists a* $\delta > 0$ *such that* $f_0(x) > f_0(\bar{x})$ *for all points* $x \in C$ *with* $0 < \|x - \bar{x}\| \leq \delta$*.*

Proof To set the stage, we invoke the polyhedral nature of C to get the existence of $\rho > 0$ such that the points $x \in C$ with $0 < \|x - \bar{x}\| \leq \rho$ are the points expressible as $\bar{x} + \tau w$ for some vector $w \in T_C(\bar{x})$ with $\|w\| = 1$ and scalar $\tau \in (0, \rho]$. Then too, for any $\delta \in (0, \rho)$, the points $x \in C$ with $0 < \|x - \bar{x}\| \leq \delta$ are the points expressible as $\bar{x} + \tau w$ for some $w \in T_C(\bar{x})$ with $\|w\| = 1$ and some $\tau \in (0, \delta]$. Next we use the twice differentiability of f_0 to get second-order estimates around $\bar{x}$ in this notation: for any $\epsilon > 0$ there is a $\delta > 0$ such that

$$\left| f_0(\bar{x} + \tau w) - f_0(\bar{x}) - \tau \nabla f_0(\bar{x}) \cdot w - \frac{\tau^2}{2} w \cdot \nabla^2 f_0(\bar{x}) w \right| \leq \epsilon \tau^2 \quad \text{for all } \tau \in [0, \delta] \text{ when } \|w\| = 1.$$

In (a), the local optimality of $\bar{x}$ gives in this setting the existence of $\bar{\delta} > 0$ such that $f_0(\bar{x} + \tau w) - f_0(\bar{x}) \geq 0$ for $\tau \in [0, \bar{\delta}]$ when $w \in T_C(\bar{x})$, $\|w\| = 1$. For such w and any $\epsilon > 0$ we then have through second-order expansion the existence of $\delta > 0$ such that

$$\tau \nabla f_0(\bar{x}) \cdot w + \frac{\tau^2}{2} \left[w \cdot \nabla^2 f_0(\bar{x}) w + 2\epsilon \right] \geq 0 \text{ for all } \tau \in [0, \delta].$$

This condition implies that $\nabla f_0(\bar{x}) \cdot w \geq 0$, and if actually $\nabla f_0(\bar{x}) \cdot w = 0$ then also that $w \cdot \nabla^2 f_0(\bar{x}) w + 2\epsilon \geq 0$. Since ϵ can be chosen arbitrarily, it must be true in the latter case that $w \cdot \nabla^2 f_0(\bar{x})w \geq 0$. Thus, the claim in (a) is valid for all $w \in T_C(\bar{x})$ with $\|w\| = 1$. It is also valid then for positive multiples of such vectors w, and hence for all $w \in T_C(\bar{x})$.

In (b), the desired conclusion corresponds to the existence of $\delta > 0$ such that $f_0(\bar{x}+\tau w) - f_0(\bar{x}) > 0$ for $\tau \in (0, \delta]$ when $w \in T_C(\bar{x})$, $\|w\| = 1$. Through the second-order expansion it suffices to demonstrate the existence of $\epsilon > 0$ and $\delta' > 0$ such that

$$\tau \nabla f_0(\bar{x}) \cdot w + \frac{\tau^2}{2}\left[w \cdot \nabla^2 f_0(\bar{x})w - 2\epsilon\right] > 0$$
$$\text{for all } \tau \in (0, \delta'] \text{ when } w \in T_C(\bar{x}), \|w\| = 1.$$

Pursuing an argument by contradiction, let's suppose that such ϵ and δ' don't exist. Then, for any sequence $\epsilon^\nu \searrow 0$ there must be sequences $\tau^\nu \searrow 0$ and $w^\nu \in T_C(\bar{x})$ with $\|w^\nu\| = 1$ and

$$\tau^\nu \nabla f_0(\bar{x}) \cdot w^\nu + \frac{(\tau^\nu)^2}{2}\left[w^\nu \cdot \nabla^2 f_0(\bar{x})w^\nu - 2\epsilon^\nu\right] \leq 0.$$

Because the sequence of vectors w^ν is bounded, it has a cluster point w; there is a subsequence $w^{\nu_\kappa} \to w$ as $\kappa \to \infty$. Then $\|w\| = 1$ (because the norm is a continuous function), and $w \in T_C(\bar{x})$ (because the tangent cone is a closed set). Rewriting our inequality as

$$w^{\nu_\kappa} \cdot \nabla^2 f_0(\bar{x})w^{\nu_\kappa} - 2\epsilon^{\nu_\kappa} \leq -2\nabla f_0(\bar{x}) \cdot w^{\nu_\kappa}/\tau^{\nu_\kappa},$$

where $\nabla f_0(\bar{x}) \cdot w^{\nu_\kappa} \geq 0$ under the assumption of (b), we see when $\kappa \to \infty$ with

$$\nabla f_0(\bar{x}) \cdot w^{\nu_\kappa} \to \nabla f_0(\bar{x}) \cdot w, w^{\nu_\kappa} \cdot \nabla^2 f_0(\bar{x})w^{\nu_\kappa} - 2\epsilon^{\nu_\kappa}) \to w \cdot \nabla^2 f_0(\bar{x})w,$$

that $w \cdot \nabla^2 f_0(\bar{x})w \leq 0$, yet also $\nabla f_0(\bar{x}) \cdot w = 0$ (for if $\nabla f_0(\bar{x}) \cdot w > 0$ the right side of the inequality would go to $-\infty$). This mix of properties of w is impossible under the assumption of (b). The contradiction finishes the proof.
□