THE UNIVERSITY OF CHICAGO

Section 12:FUNDAMENTALS OF CONSTRAINED OPTIMIZATION

Mihai Anitescu, Stat 310

12.1 Introduction. Notations

- Problem Formulation

$$\min_{x \in \mathbb{R}^n} f(x) \text{ subject to x } \in \Omega$$

- Feasible set

$$\Omega = \left\{ x \mid c_i(x) = 0, i \in \mathcal{E}; \quad c_i(x) \geq 0, i \in \mathcal{I} \right\}$$

- Compact formulation

$$\min_{x \in \Omega} f(x)$$

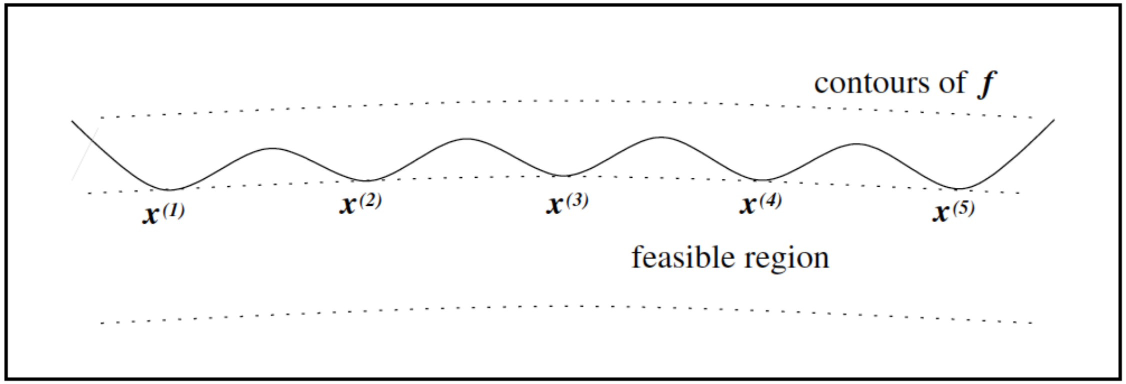
- We REALLY want differentiability

Local and Global Solutions

- Constraints make make the problem simpler since the search space is smaller.
- But it can also make things more complicated.

$$\min(x_2 + 100)^2 + 0.01x_1^2 \text{ subject to } x_2 \geq cos(x_1)$$

- Unconstrained problem has one minimum, constrained problem has MANY minima.



Types of Solutions

- Similar as the unconstrained case, except that we now restrict it to a neighborhood of the solution.
- Recall, we aim only for local solutions.

A vector x^* is a *local solution* of the problem (12.3) if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that $f(x) \geq f(x^*)$ for $x \in \mathcal{N} \cap \Omega$.

A point x^* is an *isolated local solution* if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that x^* is the only local solution in $\mathcal{N} \cap \Omega$.

A vector x^* is a *strict local solution* (also called a *strong local solution*) if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that $f(x) > f(x^*)$ for all $x \in \mathcal{N} \cap \Omega$ with $x \neq x^*$.

Smoothness

- It is ESSENTIAL that the problem be formulated with smooth constraints and objective function (since we will take derivatives).
- Rephrasing a constraint:

$$\max\left\{ f_1(x), f_2(x) \right\} \le a \Leftrightarrow f_1(x) \le a \wedge f_2(x) \le a$$

$$\|x\|_1 = |x_1| + |x_2| \le 1 \Leftrightarrow \max\left\{ -x_1, x_1 \right\} + \max\left\{ -x_2, x_2 \right\} \le 1 \Leftrightarrow$$

$$-x_1 - x_2 \le 1, \quad x_1 - x_2 \le 1, \quad -x_1 + x_2 \le 1, \quad x_1 + x_2 \le 1$$

$$\min f(x);\, f(x) = \max\left\{ x^2, x \right\};\, \begin{cases} \min & t \\ \text{subject to} & \max\left\{ x^2, x \right\} \le t \end{cases}$$

$$\Leftrightarrow \begin{cases} \min & t \\ \text{subject to} & x^2 \le t,\, x \le t \end{cases}$$

- Single equality constraint

$$\min x_1 + x_2 \quad \text{subject to} \quad x_1^2 + x_2^2 - 2 = 0$$

- Single inequality constraint

$$\min x_1 + x_2 \qquad \text{s.t.} \qquad 2 - x_1^2 - x_2^2 \geq 0,$$

- Two inequality constraints

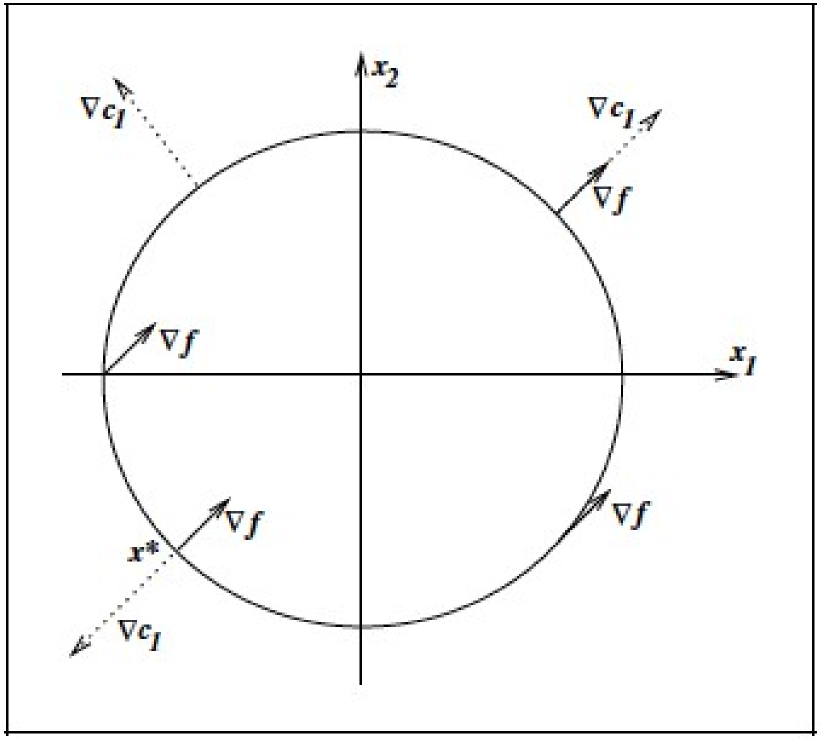
$$\min x_1 + x_2 \qquad \text{s.t.} \qquad 2 - x_1^2 - x_2^2 \geq 0, \quad x_2 \geq 0,$$

Single Equality Constraint Example

- Problem:

$$\min f(x_1,x_2) := x_1 + x_2 \quad \text{s.t.} \quad g(x_1,x_2) := x_1^2 + x_2^2 - 2 = 0$$

- Imagine that the vector [-0.5,-0.5] is “gravity” and the feasible point can slide on the circle.
- The “minimum height” will be reached on the lower left, at point [-1,-1].
- At that point the gradient of the objective and the constraint are collinear.



$$\nabla f(x^*) = [1,1]^T; \quad \nabla g(x^*) = [-2,-2]^T \Rightarrow \nabla f(x^*) + \frac{1}{2}\nabla g(x^*) = 0$$

Define $\mathcal{L}(x,\lambda) = f(x) - \lambda g(x) \Rightarrow \nabla_x \mathcal{L}(x^*,\lambda^*) = 0$

Single Inequality Constraint Example

- Problem:

$$\min f(x_1,x_2) := x_1 + x_2 \quad \text{s.t.} \quad g(x_1,x_2) := -x_1^2 - x_2^2 + 2 > 0$$

- The solution is the same and satisfies.

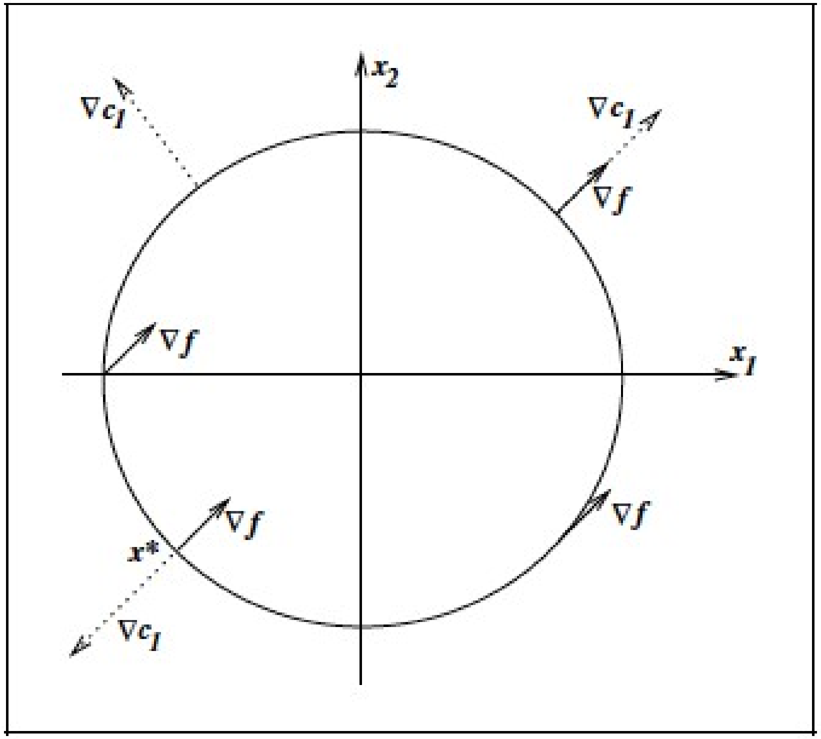
$$\nabla f(x^*) = \begin{pmatrix}1\\1\end{pmatrix}; \nabla g(x^*) = \begin{pmatrix}2\\2\end{pmatrix} \Rightarrow \nabla f(x^*) - \frac{1}{2}\nabla g(x^*) = 0$$

Define $\mathcal{L}(x,\lambda) = f(x) - \lambda g(x) \Rightarrow \nabla_x \mathcal{L}(x^*,\lambda^*) = 0, \lambda^* = \frac{1}{2}$

- But wait, the point [1,1] satisfies:

$$\nabla f(x^*) = \begin{pmatrix}1\\1\end{pmatrix}; \quad \nabla g(x^*) = \begin{pmatrix}2\\2\end{pmatrix} \Rightarrow \nabla f(x^*) + \frac{1}{2}\nabla g(x^*) = 0$$

$$\nabla_x \mathcal{L}(x^*,\lambda^*) = 0, \lambda^* = -0.5$$



- Is it a minimum? Obviously not:

$$x(t) = \begin{pmatrix}1\\1\end{pmatrix} + t\begin{pmatrix}-1\\-1\end{pmatrix} \to f(x(t)) = 2 - 2t; g(x(t)) = 4t - 2t^2 \Rightarrow f(x(t)) < 0, g(x(t)) > 0, t \in (0, 0.5)$$

Single Inequality Constraint Example (Continued)

$$x(t) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + t \begin{pmatrix} -1 \\ -1 \end{pmatrix} \rightarrow f(x(t)) = 2 - 2t; g(x(t)) = 4t - 2t^2 \Rightarrow f(x(t)) < 0, g(x(t)) > 0, t \in (0, 0.5)$$

- There is an entire strictly feasible arc on which the function decreases!
- So the point [1,1]^T cannot be a minimum even as it satisfies

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$$

- What is different? The sign of the multiplier \lambda^* needs to be >=0, which is true in the first case but not the second.
- Also, it seems clear that if the solution would be inside of the circle, then from optimality conditions, we would have \lambda^*=0.
- Suggests the condition:

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \;\; \lambda^* \geq 0, \;\; g(x^*) \geq 0, \;\; \lambda^* g(x^*) = 0$$

No first-order descent

- To make sure we are at a minimum, at least the following should happen.
- If we have a feasible arc the gradient function should not decrease from a minimum point.
- In other words, in our case, we should have that the system of inequalities: $\nabla g(x^*)^T d < 0, \ \nabla f(x^*)^T d < 0$ has no solution.

12.3 Tangent cone. The optimality conditions our example

- This is an instance of the Karush-Kuhn-Tucker (KKT)

Conditions for problems of inequality constraints.

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \lambda^* \geq 0, \quad c(x^*) \geq 0,\ c(x^*)^T \lambda^* = 0$$

- Here c is a vector of inequality constraints.
- This is equivalent to :

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \quad \lambda_i^* \geq 0, c_i(x^*) \geq 0,\ c_i(x^*)^T \lambda_i^* = 0, i \in \mathcal{I}$$

Inequality Constraints: Active Set

$$\min_{x\in\mathbb{R}^n} f(x) \text{ subject to } \begin{cases} c_i(x) = 0 & i\in\mathcal{E} \\ c_i(x) \geq 0 & i\in\mathcal{I} \end{cases}$$

- One of the key differences with equality constraints.
- Definition at a feasible point x.

$$x \in \Omega(x) \quad \mathcal{A}(x) = \mathcal{E} \cup \{ i \in \mathcal{I};\, c_i(x) = 0 \}$$

Tangent and linearized cone

- For our Optimization problem:
- Tangent Cone at x (can prove it is a cone)

$$T_\Omega(x) = \left\{ d \,\middle|\, \exists \{z_k\} \in \Omega, z_k \to x, \exists \{t_k\} \in \mathbb{R}_+, t_k \to 0, \lim_{k\to\infty} \frac{z_k - x}{t_k} = d \right\}$$

- Linearized feasible direction set

$$\mathcal{F}(x) = \left\{ d \,\middle|\, d^T \nabla c_i(x) = 0, i \in \mathcal{E}; d^T \nabla c_i(x) \geq 0, i \in \mathcal{A}(x) \cap \mathcal{I} \right\} \Rightarrow T_\Omega(x) \subset \mathcal{F}(x)$$

- It is immediate that the tangent directions are feasible, so the inclusion.
- Under what conditions does the reverse happen and thus:

$$T_\Omega(x) = \mathcal{F}(x)\,?$$

- We say then that the constraint qualification holds

Example

- Consider our equality constrained problem.

$$\min x_1 + x_2 \quad \text{subject to} \quad x_1^2 + x_2^2 - 2 = 0$$

- Look at the point (1,1). Unsurprisingly the tangent cone is parallel to the tangent line x_1+x_2=2, and is thus d_1+d_2=0.
- Its equation is the same as the equation of the linearized feasible directions: 2x_1d_1+2x_2d_2=0.
- So in this case the two sets are equal.

Example (2)

- Consider now the same problem, but with the equality represented differently, by squaring it:

$$\min x_1 + x_2 \quad \text{subject to} \quad \left(x_1^2 + x_2^2 - 2\right)^2 = 0$$

- The tangent cone is dependent of the feasible set and not of the representation, so it does not change, it is still d_1+d_2=0
- The feasible directions cone, however, now satisfies

$$4(x_1^2 + x_2^2 - 2)x_1 d_1 + 4(x_1^2 + x_2^2 - 2)x_2 d_2 = 0$$

$$\equiv 0 \cdot d_1 + 0 \cdot d_2 = 0 \Rightarrow \mathcal{F} = \mathbb{R}^2 \neq T_\Omega$$

- So the equivalence condition does not depend only on the geometry, it depends also on the algebraic representation of the constraints.

The Linear Independence Constraint Qualification

Definition 12.4 (LICQ).

Given the point x and the active set $\mathcal{A}(x)$ defined in Definition 12.1, we say that the linear independence constraint qualification (LICQ) *holds if the set of active constraint gradients $\{\nabla c_i(x), i \in \mathcal{A}(x)\}$ is linearly independent.*

- The significance of LICQ:

Lemma 12.2.

Let x^* be a feasible point. The following two statements are true.

(i) $T_\Omega(x^*) \subset \mathcal{F}(x^*)$.

(ii) *If the LICQ condition is satisfied at x^*, then $\mathcal{F}(x^*) = T_\Omega(x^*)$.*

$$\Longrightarrow T_\Omega(x) = \mathcal{F}(x)$$

Sketch of the LICQ is a constraint qualification proof

- How do we prove equality of the cones ?
- If LICQ holds, then, we can use the implicit function theorem.

$$d \in \mathcal{F}(x) \Rightarrow c_{\mathcal{A}(x)}\left(\tilde{x}(t)\right) = t\nabla c_{\mathcal{A}(x)} d \Rightarrow \exists \tau > 0, \forall 0 < t < \tau;$$

$$c_{\bar{\mathcal{A}}(x)}\left(\tilde{x}(t)\right) > 0; c_{\mathcal{A}(x)\cap\mathcal{I}}\left(\tilde{x}(t)\right) \geq 0; c_{\mathcal{E}}\left(\tilde{x}(t)\right) = 0 \Rightarrow \tilde{x}(t) \in \Omega \Rightarrow d \in T_{\Omega}(x)$$

Another useful feature (when it holds) : strict complementarity

- It is a notion that makes the problem look "almost" like an equality.

Definition 12.5 (Strict Complementarity).

Given a local solution x^* of (12.1) and a vector λ^* satisfying (12.34), we say that the strict complementarity condition *holds if* exactly one *of λ_i^* and $c_i(x^*)$ is zero for each index $i \in \mathcal{I}$. In other words, we have that $\lambda_i^* > 0$ for each $i \in \mathcal{I} \cap \mathcal{A}(x^*)$.*

12.4 First-order Optimality Conditions.

- The Lagrangian:

$$\mathcal{L}(x) = f(x) - \sum_{i \in \mathrm{E} \cup \mathcal{A}} \lambda_i c_i(x)$$

First-Order Optimality Condition Theorem

Suppose that x^* is a local solution of (12.1), that the functions f and c_i in (12.1) are continuously differentiable, and that the LICQ holds at x^*. Then there is a Lagrange multiplier vector λ^*, with components λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at (x^*, λ^*)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{12.34a}$$

$$c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E}, \tag{12.34b}$$

$$c_i(x^*) \geq 0, \quad \text{for all } i \in \mathcal{I}, \tag{12.34c}$$

$$\lambda_i^* \geq 0, \quad \text{for all } i \in \mathcal{I}, \tag{12.34d}$$

$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}. \tag{12.34e}$$

Equivalent Form:

$$\nabla f\left(x^*\right) - \lambda^T{}_{\mathcal{A}\left(x^*\right)} \nabla c_{\mathcal{A}\left(x^*\right)}\left(x^*\right) = 0 \Rightarrow \text{Multipliers are unique !!}$$

First, a tangent cone necessary condition

- Assuming only first-order continuous differentiability, we have:

Theorem 12.3.
If x^* is a local solution of (12.1), then we have

$$\nabla f(x^*)^T d \geq 0, \quad \text{for all } d \in T_\Omega(x^*).$$

- Proof: by Taylor Expansion over feasible directions. For the sequences defining that d is an element of the tangent cone, we get:

$$\begin{aligned} f(z_k) &= f(x^*) + (z_k - x^*)^T \nabla f(x^*) + o(\|z_k - x^*\|) \\ &= f(x^*) + t_k d^T \nabla f(x^*) + o(t_k), \end{aligned}$$

- And conclusion follows from optimality of x^*, since $z_k - x^* = t_k d + o(t_k)$

Second, the Farkas Lemma

- Consider a cone K defined as follows:

$$K = \{By + Cw \mid y \geq 0\},$$

- This cone is closed, though that is not trivial to prove (though easy when columns of B and C are independent)
- Consider the alternatives:

$$g \in K \quad (1) \quad \text{AND} \quad \exists d \in \mathbb{R}^n,\ g^T d < 0, \qquad B^T d \geq 0, \qquad C^T d = 0. \ (2)$$

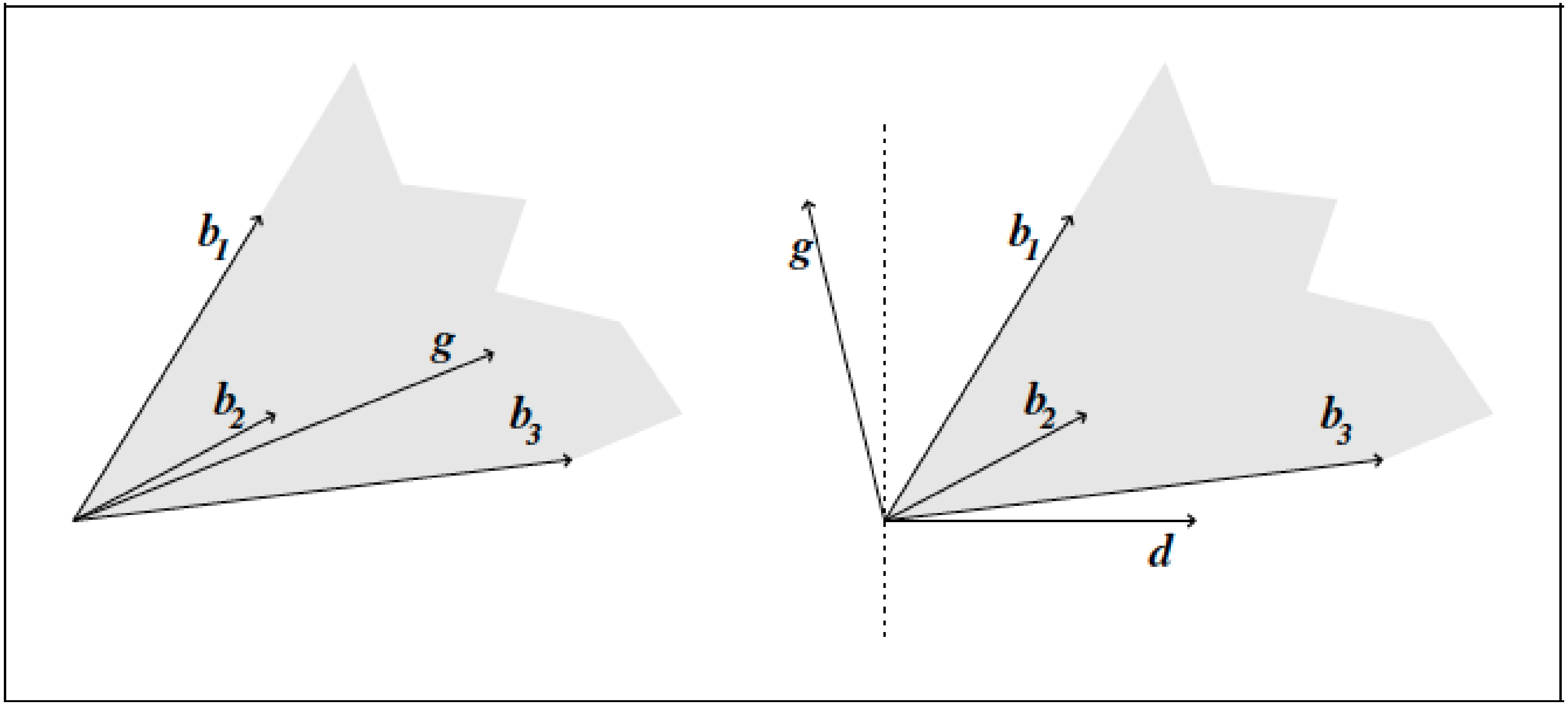
- Farkas Lemma:

Lemma 12.4 (Farkas).

Let the cone K be defined as in (12.45). Given any vector $g \in \mathbb{R}^n$, we have either that $g \in K$ or that there exists $d \in \mathbb{R}^n$ satisfying (12.46), but not both.

The even deeper meaning: separating hyperplanes

- "Either a point g is in a convex cone C, or otherwise, it is on the other side of a hyperplane from the cone C; the hyperplane can be chosen to go through the origin".
- Picture:



Proof of KKT conditions.

- From LICQ, the tangent cone equals the feasible directions.
- From Lemma 12.3, we must then have

$$\nabla c_{\mathcal{I}\cap\mathcal{A}}(x)^T d \geq 0, \quad \nabla c_{\mathcal{E}}^T d = 0 \Rightarrow \nabla f(x)^T d \geq 0$$

- From Farkas' Lemma, it means that the second alternative does not hold so the first one must hold for the convex cone:

$$\nabla f(x) \in \{\nabla c_{\mathcal{I}\cap\mathcal{A}}(x) y + \nabla c_{\mathcal{E}}^T w, y \geq 0\} \Rightarrow$$

$$\nabla f(x) = \sum_{i\in\mathcal{A}} \nabla c_i(x) \lambda_i^*, \quad \lambda^*{}_{\mathcal{I}\cap\mathcal{A}} \geq 0$$

- Equivalent to the FO optimality conditions.

12.5 Second-order conditions. Critical Cone

- The subset of the tangent space, where the objective function does not vary to first-order.
- The book definition.

$$\mathcal{C}(x^*, \lambda^*) = \{w \in \mathcal{F}(x^*) \mid \nabla c_i(x^*)^T w = 0, \text{ all } i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0\}.$$

- An even simpler equivalent definition.

$$\mathcal{C}\left(x^*, \lambda^*\right) = \left\{ w \in T_\Omega\left(x^*\right) \middle| \nabla f\left(x^*\right)^T w = 0 \right\}$$

Rephrasing of the Critical Cone

- By investigating the definition

$$w \in \mathcal{C}\left(x^*, \lambda^*\right) \Leftrightarrow \begin{cases} \nabla c_i\left(x^*\right)^T w = 0 & i \in \mathcal{E} \\ \nabla c_i\left(x^*\right)^T w = 0 & i \in \mathcal{A}\left(x^*\right) \bigcap \mathcal{I} \quad \lambda_i^* > 0 \\ \nabla c_i\left(x^*\right)^T w \geq 0 & i \in \mathcal{A}\left(x^*\right) \bigcap \mathcal{I} \quad \lambda_i^* = 0 \end{cases}$$

- In the case where strict complementarity holds, the critical cone has a MUCH simpler expression (they are a linear subspace)

$$w \in \mathcal{C}\left(x^*, \lambda^*\right) \Leftrightarrow \nabla c_i\left(x^*\right) w = 0 \; \forall \, i \in \mathcal{A}\left(x^*\right)$$

Statement of the Second-Order Conditions

Theorem 12.5 (Second-Order Necessary Conditions).

Suppose that x^* is a local solution of (12.1) and that the LICQ condition is satisfied. Let λ^* be the Lagrange multiplier vector for which the KKT conditions (12.34) are satisfied. Then

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0, \quad \text{for all } w \in \mathcal{C}(x^*, \lambda^*). \tag{12.57}$$

- How to prove this? In the case of Strict Complementarity the critical cone is the same as the problem constrained with equalities on active index.

Statement of second-order sufficient conditions

Theorem 12.6 (Second-Order Sufficient Conditions).

Suppose that for some feasible point $x^* \in \mathbb{R}^n$ there is a Lagrange multiplier vector λ^* such that the KKT conditions (12.34) are satisfied. Suppose also that

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w > 0, \quad \text{for all } w \in \mathcal{C}(x^*, \lambda^*),\ w \neq 0. \tag{12.65}$$

Then x^* is a strict local solution for (12.1).

- Proof (HWK for LICQ+SC)

Projected Hessians and Second Order Conditions

- When strict complementarity holds, the second order conditions both necessary and sufficient are much more easily expressed (and proved).
- The null space of the constraints is precisely the critical cone:

$$\mathcal{C}(x^*, \lambda^*) = \{Zu \mid u \in \mathbb{R}^{|\mathcal{A}(x^*)|}\}.$$

- The projected Hessian of the Lagrangian is: $Z^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z$
- Second Order Optimality conditions become:

$$u^T Z^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z u \geq 0 \quad \text{for all } u,$$

or, more succinctly,

$$Z^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z \quad \text{is positive semidefinite.}$$