Lecture 5: Primal-dual optimality conditions

Overview

- Want to establish that $\boldsymbol{x}^*$ local minimum of f over S implies that a well-defined condition holds that we can easily check
- This is possible when constraints are linear, since the set of feasible directions then can be stated simply
- With non-linear constraints things become more complicated
- *Constraint qualifications* (CQ) are needed to make sure that the *well-defined* condition is a necessary condition for local optimality (rule out strange cases)
- Under convexity, the condition turns out to also always (under no CQ requirement) be sufficient for global optimality
- Called the *Karush–Kuhn–Tucker* conditions
- Karush: master's student at Univ. of Chicago, 1939
  Tucker/Kuhn: prof./Ph.D. student at Princeton Univ., 1951

- Of course, a globally optimal solution must then satisfy the KKT conditions. But it is *not* practical to search for all KKT points and pick the best. Its use is for checking that an algorithm has found the right solution and as the basis for the design of algorithms
- Compare checking for every x with $f'(x) = 0$ in $\mathbb{R}$!
- The user has all the responsibility!
- Logic: $\boldsymbol{x}^*$ local optimal AND a CQ holds $\implies$ KKT holds at $\boldsymbol{x}^*$
- Equivalent: KKT does not hold at $\boldsymbol{x}^*$ $\implies$ $\boldsymbol{x}^*$ is not a local optimal OR no CQ holds
- CQ important because if it holds then if we are not at a KKT point then we know we are not locally optimal!

Cautions needed!

- Costly errors can be made if one ignores that KKT conditions are *necessary*, but not always *sufficient*
- US Air Force's B-2 Stealth bomber program: Reaganism, 1980s
- Design variables: various dimensions, distribution of volume between wing and fuselage, flying speed, thrust, fuel consumption, drag, lift, air density, etc
- Objective: maximum range on full tank
- Model from the 1940s which had produced B-29, B-52, etc
- Solution to the KKT conditions found; specified design variable values that put almost all of the total volume in the wing, leading to the *flying wing design* for the B-2 bomber
- Billions of dollars later, found the design solution works, but its range too low in comparison with other bomber designs

- Review carried out. The model is correct!
- But ... The model was a nonconvex NLP; the review revealed a second solution to the KKT system
- Much less wing volume! Looks like an airplane! Maximizes range!
- In other words, the design implemented was the aerodynamically *worst* possible choice of configuration, leading to a very costly error
- Still flies. Why? Happens that it has good properties wrt. radar protection (stealth) ...

Nice photos, I

Nice photos, II

Overview, cont'd

- The condition must not only be easy to check, it should also state something useful
- It is easy to state some condition: *If* $\boldsymbol{x}^*$ *is a local minimum of* f *over* S *then it is also feasible*
- Completely useless, since it is satisfied for every feasible point
- That is what we end up with if we want something that is applicable to every problem. We need to get rid of some weird problems, and that is a main reason for introducing the CQs
- We begin by studying an abstract problem and provide a *geometric optimality condition*
- Next, we state the corresponding result for an explicit representation of S in terms of constraints. This is the *Fritz John* condition

- Introducing a CQ we then obtain the *Karush–Kuhn–Tucker* conditions
- There is more than one CQ, some more useful than others in particular cases
- *Linear independence of the equality constraints* is the classic one from the Lagrange multiplier rule. We extend it and show others

Geometric optimality conditions

Problem:

$$\begin{aligned} &\text{minimize } f(\boldsymbol{x}), \\ &\text{subject to } \boldsymbol{x} \in S, \end{aligned} \tag{1}$$

$S \subset \mathbb{R}^n$ nonempty, closed; $f : \mathbb{R}^n \to \mathbb{R}$ in C^1

- Idea: at a local minimum $\boldsymbol{x}^*$ of f over S it is impossible to draw a curve from $\boldsymbol{x}^*$ such that it is feasible and f decreases along it
- Cannot work with f itself; descent is measured in terms of directional derivatives. Linearize f
- We must also "linearize" S. Reason: the cone of feasible directions may be too small to be useful; also, it is difficult to state it explicitly. We replace the cone of feasible directions with the *tangent cone* to S at $\boldsymbol{x}^*$

- The *cone of feasible directions* for S at $\boldsymbol{x} \in \mathbb{R}^n$ is

$$R_S(\boldsymbol{x}) := \{\, \boldsymbol{p} \in \mathbb{R}^n \mid \exists\, \widetilde{\delta} > 0 \text{ such that } \boldsymbol{x} + \delta \boldsymbol{p} \in S, 0 \leq \delta \leq \widetilde{\delta} \,\}$$

- The *tangent cone* for S at $\boldsymbol{x} \in \mathbb{R}^n$ is

$$\begin{aligned} T_S(\boldsymbol{x}) := \{\, \boldsymbol{p} \in \mathbb{R}^n \mid \exists\, \{\boldsymbol{x}_k\} \subset S, \{\lambda_k\} \subset (0, \infty) : & \lim_{k \to \infty} \boldsymbol{x}_k = \boldsymbol{x}, \\ & \lim_{k \to \infty} \lambda_k (\boldsymbol{x}_k - \boldsymbol{x}) = \boldsymbol{p} \,\} \end{aligned}$$

- $T_S(\boldsymbol{x})$ is closed; the set of tangents to sequences $\{\boldsymbol{x}_k\} \subset S$
- *It holds that* $\operatorname{cl} R_S(\boldsymbol{x}) \subset T_S(\boldsymbol{x})$ *for every* $\boldsymbol{x} \in \mathbb{R}^n$
- Suppose that for functions $g_i \in C^1$, $i = 1, \ldots, m$:

$$S := \{\, \boldsymbol{x} \in \mathbb{R}^n \mid g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \,\}$$

- Two further cones:

$$\mathring{G}(\boldsymbol{x}) := \{\, \boldsymbol{p} \in \mathbb{R}^n \mid \nabla g_i(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{p} < 0, i \in \mathcal{I}(\boldsymbol{x}) \,\},$$

and

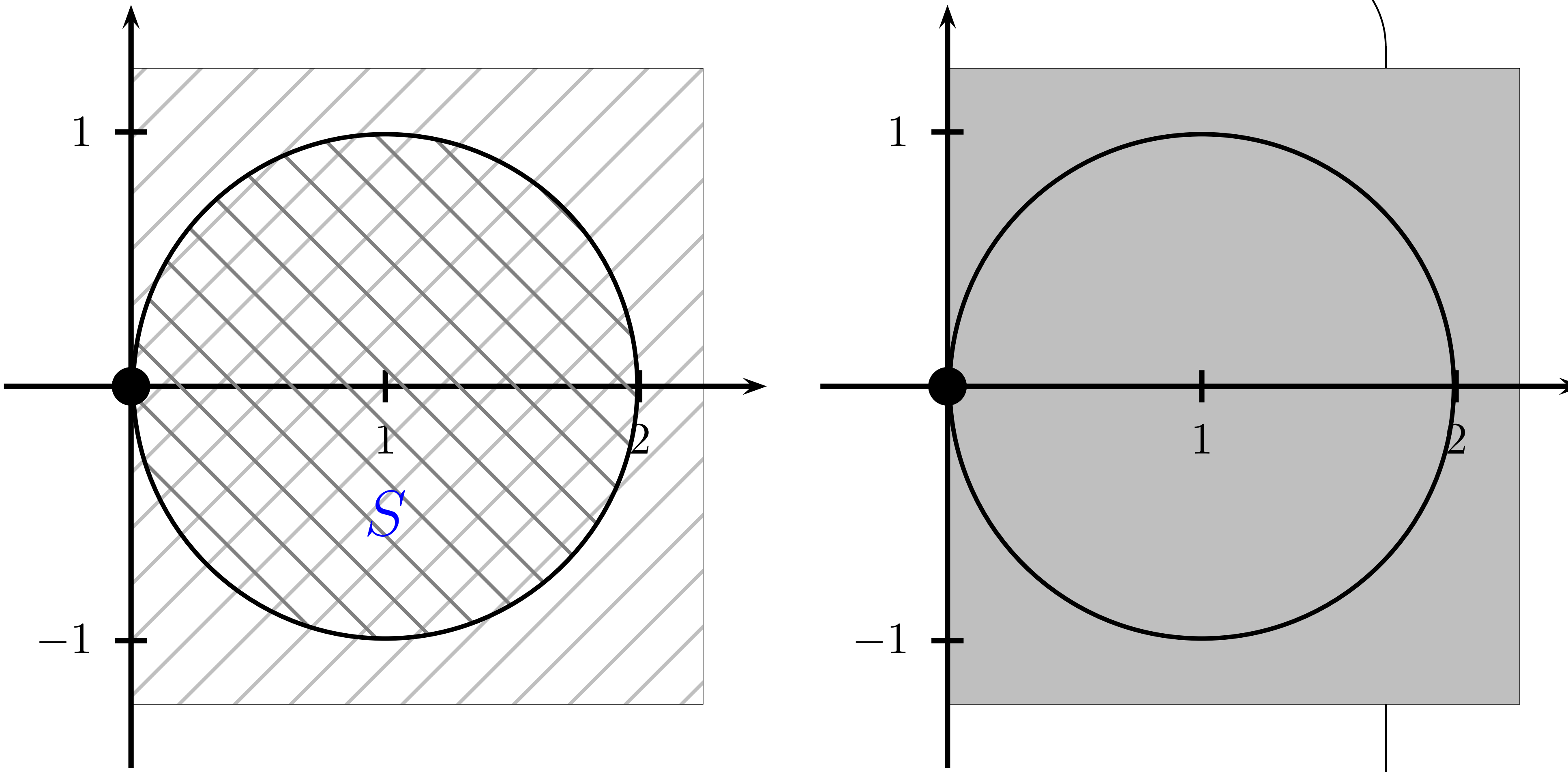
$$G(\boldsymbol{x}) := \{\, \boldsymbol{p} \in \mathbb{R}^n \mid \nabla g_i(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{p} \leq 0, i \in \mathcal{I}(\boldsymbol{x}) \,\}$$

- *For every* $\boldsymbol{x} \in \mathbb{R}^n$ *it holds that* $\overset{\circ}{G}(\boldsymbol{x}) \subset R_S(\boldsymbol{x})$, and $T_S(\boldsymbol{x}) \subset G(\boldsymbol{x})$
- So, for every $\boldsymbol{x} \in \mathbb{R}^n$,

$$\overset{\circ}{G}(\boldsymbol{x}) \subset R_S(\boldsymbol{x}) \subset \operatorname{cl} R_S(\boldsymbol{x}) \subset T_S(\boldsymbol{x}) \subset G(\boldsymbol{x})$$

Four examples, I

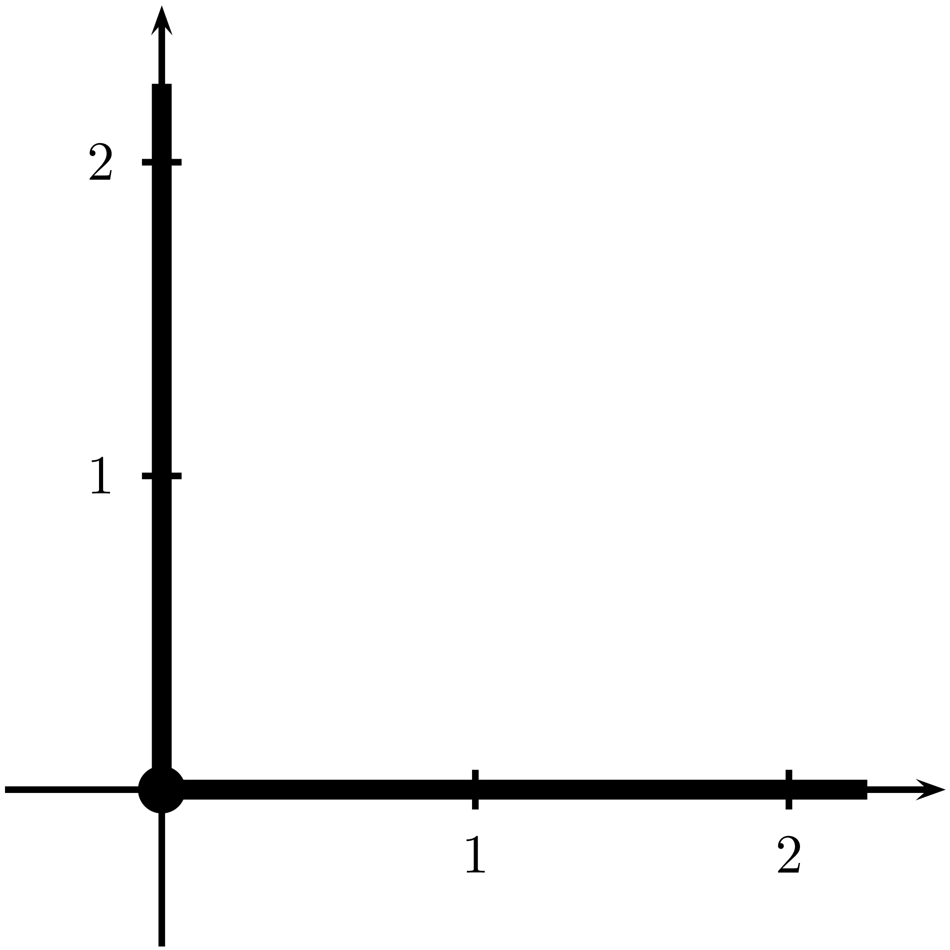
- $S = \{\, \boldsymbol{x} \in \mathbb{R}^2 \mid -x_1 \leq 0, (x_1 - 1)^2 + x_2^2 \leq 1 \,\}$
- $R_S(\mathbf{0}^2) = \{\, \boldsymbol{p} \in \mathbb{R}^2 \mid p_1 > 0 \,\}$
- $T_S(\mathbf{0}^2) = \{\, \boldsymbol{p} \in \mathbb{R}^2 \mid p_1 \geq 0 \,\}$
- $T_S(\mathbf{0}^2) = \operatorname{cl} R_S(\mathbf{0}^2)$



S $\qquad\qquad$ $T_S(\mathbf{0}^2)$

Four examples, II

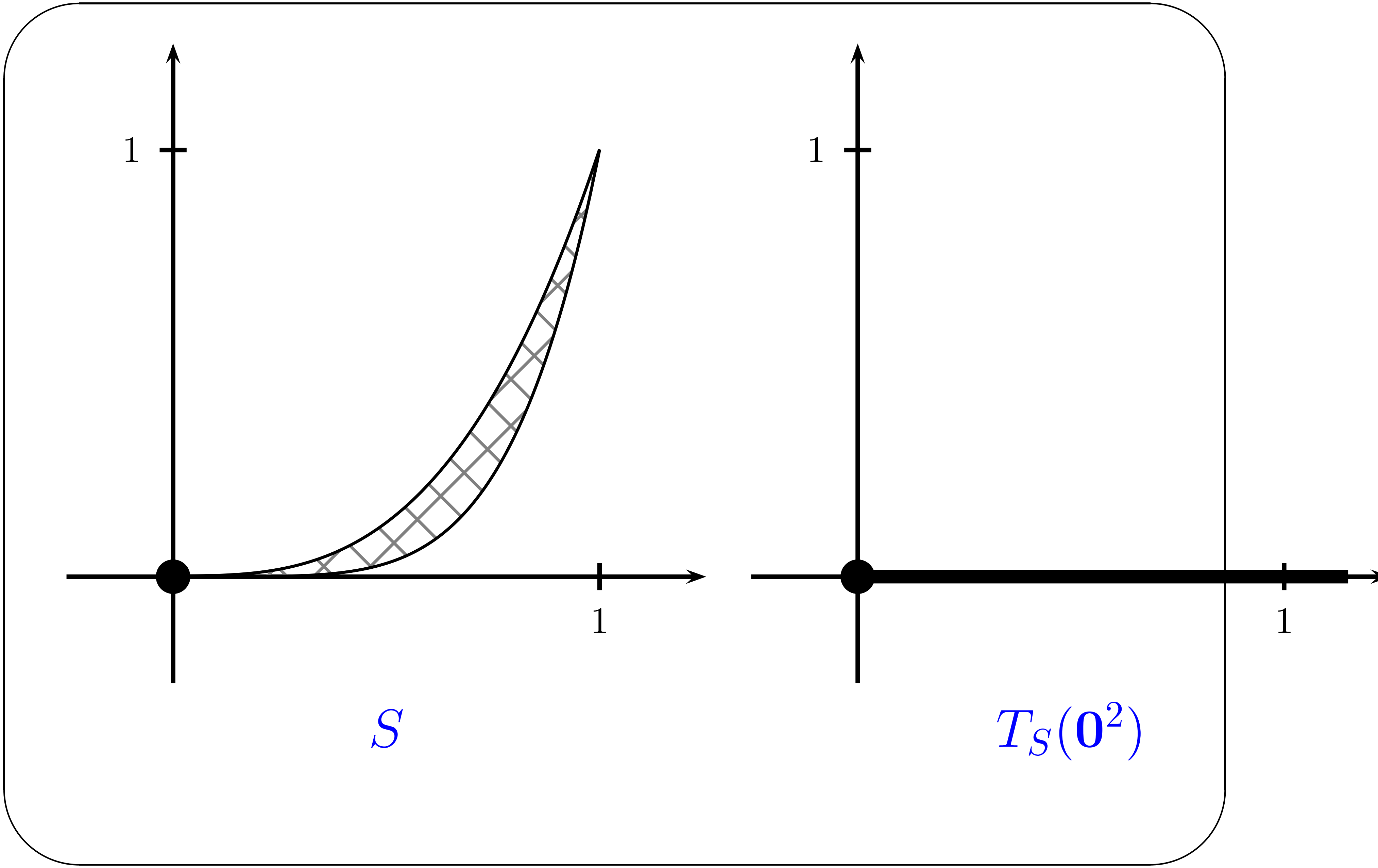
- $S = \{\, \boldsymbol{x} \in \mathbb{R}^2 \mid -x_1 \leq 0, -x_2 \leq 0, x_1 x_2 \leq 0 \,\}$
- $R_S(\mathbf{0}^2) = T_S(\mathbf{0}^2) = S$



$$S = T_S(\mathbf{0}^2)$$

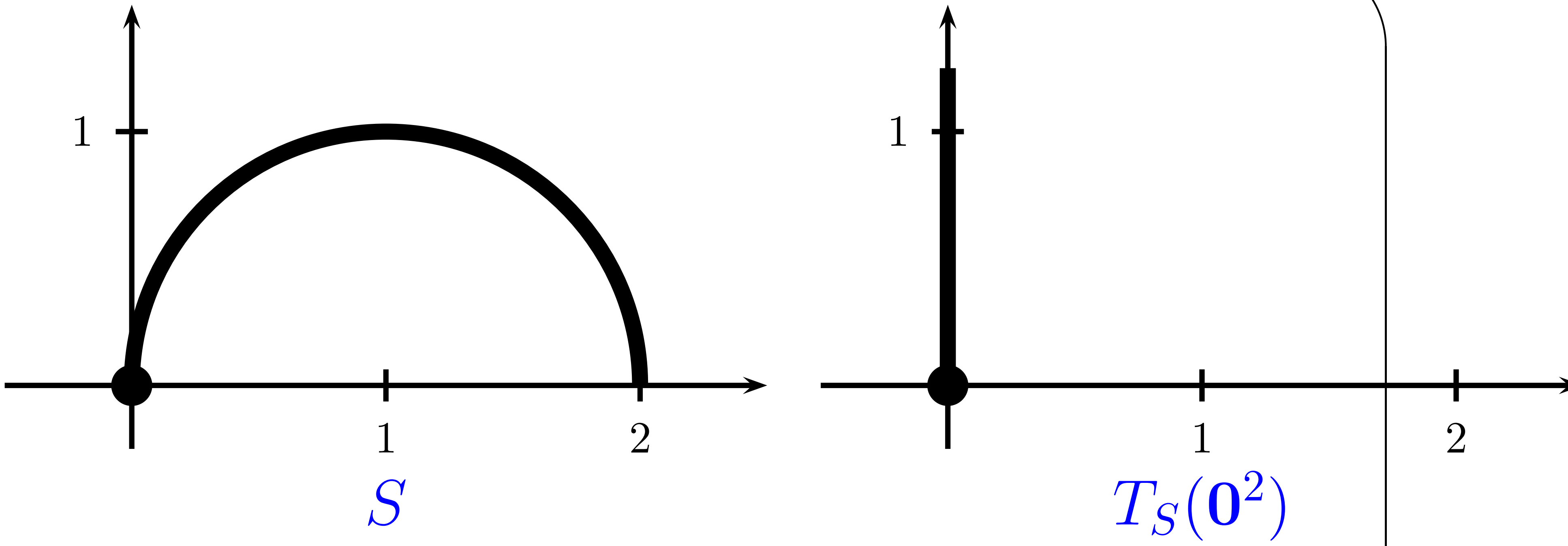
Four examples, III

- $S = \{\, \boldsymbol{x} \in \mathbb{R}^2 \mid -x_1^3 + x_2 \leq 0, x_1^5 - x_2 \leq 0, -x_2 \leq 0 \,\}$
- $R_S(\mathbf{0}^2) = \emptyset$
- $T_S(\mathbf{0}^2) = \{\, \boldsymbol{p} \in \mathbb{R}^2 \mid p_1 \geq 0, p_2 = 0 \,\}$


1
1
1
1
S
$T_S(\mathbf{0}^2)$

Four examples, IV

- $S = \{\, \boldsymbol{x} \in \mathbb{R}^2 \mid -x_2 \leq 0, (x_1 - 1)^2 + x_2^2 = 1 \,\}$
- $R_S(\mathbf{0}^2) = \emptyset$
- $T_S(\mathbf{0}^2) = \{\, \boldsymbol{p} \in \mathbb{R}^2 \mid p_1 = 0, p_2 \geq 0 \,\}$


1
1
2
S
1
1
2
$T_S(\mathbf{0}^2)$

A geometric necessary optimality condition

- $\mathring{F}(\boldsymbol{x}^*) := \{\, \boldsymbol{p} \in \mathbb{R}^n \mid \nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{p} < 0 \,\}$
- *Consider the problem* (1). *If* $\boldsymbol{x}^* \in S$ *is a local minimum of* f *over* S *then* $\mathring{F}(\boldsymbol{x}^*) \cap T_S(\boldsymbol{x}^*) = \emptyset$
- This is an elegant criterion for checking whether a given point is a candidate for a local minimum. There is a catch though:
- The set $T_S(\boldsymbol{x}^*)$ is nearly impossible to compute in general!
- We will compute other cones that we hope will approximate $T_S(\boldsymbol{x}^*)$ well enough
- Specifically, we will use the cone $\mathring{G}(\boldsymbol{x})$

Example problem

- Consider the differentiable (linear) function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(\boldsymbol{x}) = x_1$
- Then, $\nabla f = (1, 0)^{\mathrm{T}}$, and $\mathring{F}(\mathbf{0}^2) = \{\, \boldsymbol{p} \in \mathbb{R}^2 \mid p_1 < 0 \,\}$
- $\boldsymbol{x}^* = \mathbf{0}^2$ is a local (in fact, even global) minimum in problem (1) with S given by either one of Examples I–IV
- Easy to check that the geometric necessary optimality condition $\mathring{F}(\mathbf{0}^2) \cap T_S(\mathbf{0}^2) = \emptyset$ is satisfied in all four examples (no surprise, in view of the above geometric theorem)

The Fritz John conditions

- *If* $\boldsymbol{x}^* \in S$ *is a local minimum of* f *over* S *then there exist multipliers* $\mu_0 \in \mathbb{R}$, $\boldsymbol{\mu} \in \mathbb{R}^m$ *such that*

$$\mu_0 \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}^*) = \mathbf{0}^n, \tag{2a}$$

$$\mu_i g_i(\boldsymbol{x}^*) = 0, \quad i = 1, \ldots, m, \tag{2b}$$

$$\mu_0, \mu_i \geq 0, \quad i = 1, \ldots, m, \tag{2c}$$

$$(\mu_0, \boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}} \neq \mathbf{0}^{m+1} \tag{2d}$$

- Proof via the geometric necessary conditions and Farkas' Lemma
- What's bad about the Fritz John conditions? It may be possible to fulfill (2) at every feasible point by setting $\mu_0 = 0$! Then, f plays no role, which is bad. We will develop conditions (constraint qualifications) which ensure that $\mu_0 > 0$

Comments

- The vector $\boldsymbol{\mu}$ is a vector of *Lagrange multipliers*. Each of them is associated with a constraint, and will be shown to be a measure of the sensitivity of the solution to changes in the constraints
- Conditions (2a), (2c) are known as the *dual feasibility* conditions
- Condition (2b) is the *complementarity condition*. States that for inactive constraints $i \notin \mathcal{I}(\boldsymbol{x}^*)$, $\mu_i = 0$ must hold
- Will take a closer look at the Examples I–IV, but wait until the KKT conditions have been developed
- We do this by introducing conditions that bring either $\overset{\circ}{G}(\boldsymbol{x})$ or $G(\boldsymbol{x})$ to be tight enough approximations of $T_S(\boldsymbol{x})$

The Karush–Kuhn–Tucker conditions

- *Abadie's CQ:* At $\boldsymbol{x} \in S$ Abadie's constraint qualification holds if $G(\boldsymbol{x}) = T_S(\boldsymbol{x})$
- Satisfied by Example I and IV
- *Assume that at $\boldsymbol{x}^* \in S$ Abadie's CQ holds. If $\boldsymbol{x}^* \in S$ is a local minimum of f over S then there exists $\boldsymbol{\mu} \in \mathbb{R}^m$ such that*

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}^*) = \mathbf{0}^n, \tag{3a}$$

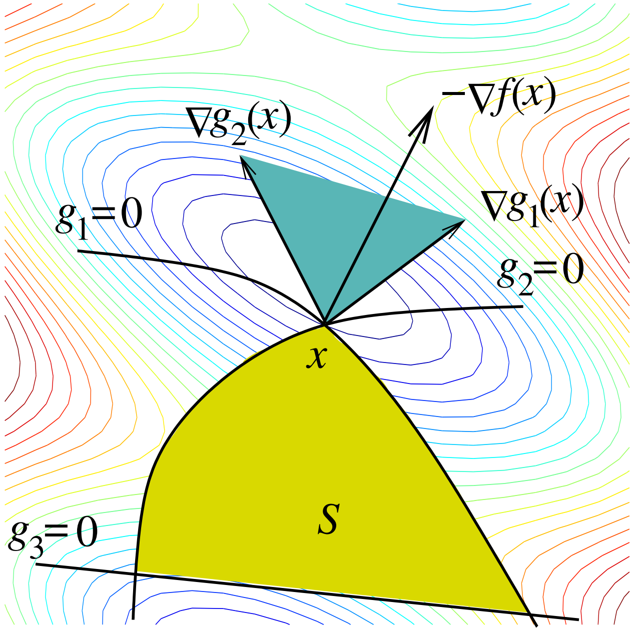
$$\mu_i g_i(\boldsymbol{x}^*) = 0, \quad i = 1, \ldots, m, \tag{3b}$$

$$\boldsymbol{\mu} \geq \mathbf{0}^m \tag{3c}$$

- Proof by first noting that $\overset{\circ}{F}(\boldsymbol{x}^*) \cap T_S(\boldsymbol{x}^*) = \emptyset$, which due to our CQ implies that $\overset{\circ}{F}(\boldsymbol{x}^*) \cap G(\boldsymbol{x}^*) = \emptyset$. Rest of the proof by Farkas' Lemma. [*Note:* case of $m = 0$!]

Comments

- The statement in (3a) is that $\boldsymbol{x}^*$ is a stationary point to the Lagrangian function $\boldsymbol{x} \mapsto f(\boldsymbol{x}) + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^m \mu_i g_i(\boldsymbol{x})$
- The condition (3) is that $-\nabla f(\boldsymbol{x}^*) \in N_S(\boldsymbol{x}^*)$ holds. The normal cone $N_S(\boldsymbol{x}^*)$ is spanned by the normals of the active constraints



Example I

- Abadie's CQ is fulfilled, therefore the KKT-system is solvable

  Indeed, the system

$$\begin{cases} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 & -2 \\ 0 & 0 \end{pmatrix} \boldsymbol{\mu} = \mathbf{0}^2, \\ \boldsymbol{\mu} \geq \mathbf{0}^2, \end{cases}$$

  possesses solutions $\boldsymbol{\mu} = (\mu_1, 2^{-1}(1 - \mu_1))^{\mathrm{T}}$ for every $0 \leq \mu_1 \leq 1$. Therefore, there are infinitely many multipliers, that all belong to a bounded set

- Case of a non-unique *dual* solution $\boldsymbol{\mu}$

Equality constraints

Additional constraints $h_j(\boldsymbol{x}) = 0$, $j = 1, \ldots, \ell$

- KKT system:

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{\ell} \lambda_j \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}^n, \tag{4a}$$

$$\mu_i g_i(\boldsymbol{x}^*) = 0, \quad i = 1, \ldots, m, \tag{4b}$$

$$\boldsymbol{\mu} \geq \mathbf{0}^m \tag{4c}$$

- $\mu_i \geq 0$ for the $\leq$-constraints; λ_j is sign free for $=$-constraints
- Interpretation: The condition (4) is a force equilibrium condition
- $-\nabla f(\boldsymbol{x}^*)$ is a force to violate the active constraints
- The remaining terms equal this force. $\mu_i \geq 0$ must hold (force towards feasibility). λ_j? Cannot determine before-hand in which direction the surface must move

Other constraint qualifications

- *Slater CQ—convex sets with interior points*: The feasible set is convex, and there exists a feasible point such that every inequality constraint is satisfied strictly
- *Linear independence CQ*: The gradients of all the active constraints are linearly independent
- *Linear constraints CQ*: All the constraints are affine/linear
- *Mangasarian–Fromowitz CQ*: The gradients of all the functions h_j are linearly independent, and the set $\mathring{G}(\boldsymbol{x}) \cap H(\boldsymbol{x})$ is nonempty, where

$$H(\boldsymbol{x}) := \{\, \boldsymbol{p} \in \mathbb{R}^n \mid \nabla h_i(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{p} = 0, \quad i = 1, \ldots, \ell \,\}$$

- Some CQs are stronger than others: LICQ $\Longrightarrow$ MFCQ $\Longrightarrow$ Abadie; Slater $\Longrightarrow$ MFCQ; linear constraints $\Longrightarrow$ Abadie

Convexity implies sufficiency

- *Assume the problem* (1) *is convex, that is, f as well as g_i, $i = 1, \ldots, m$, are convex, and h_j, $j = 1, \ldots, \ell$, are affine; also, all functions are in C^1. Assume further that for $\boldsymbol{x}^* \in S$ the KKT conditions* (4) *are satisfied. Then, $\boldsymbol{x}^*$ is* a globally *optimal solution to the problem* (1)
- *Proof.*

- Check interesting applications in the book on the characterization of eigenvalues and eigenvectors!