Lecture 3: Constrained Optimization

Kevin Carlberg

Stanford University

July 31, 2009

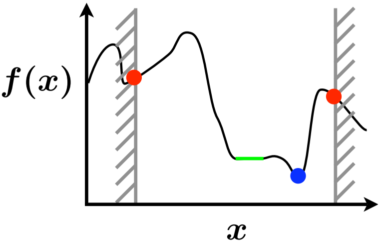
Constrained optimization

- This lecture considers constrained optimization

$$
\begin{aligned}
\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & c_i(x) = 0, \quad i = 1, \ldots, n_e \\
& d_j(x) \geq 0, \quad j = 1, \ldots, n_i
\end{aligned}
\tag{1}
$$

- Equality constraint functions: $c_i : \mathbb{R}^n \to \mathbb{R}$
- Inequality constraint functions: $d_j : \mathbb{R}^n \to \mathbb{R}$
- Feasible set:
  $\Omega = \{x \mid c_i(x) = 0,\ d_j(x) \geq 0, i = 1, \ldots, n_e,\ j = 1, \ldots, n_i\}$
- We continue to assume all functions are twice-continuously differentiable

What is a solution?



- Global minimum: A point $x^* \in \Omega$ satisfying $f(x^*) \le f(x)$ $\forall x \in \Omega$
- Strong local minimum: A neighborhood $\mathcal{N}$ of $x^* \in \Omega$ exists such that $f(x^*) < f(x)$ $\forall x \in \mathcal{N} \cap \Omega$.
- Weak local minima A neighborhood $\mathcal{N}$ of $x^* \in \Omega$ exists such that $f(x^*) \le f(x)$ $\forall x \in \mathcal{N} \cap \Omega$.

Convexity

As with the unconstrained case, conditions hold where any local minimum is the global minimum:

- $f(x)$ convex
- $c_i(x)$ affine ($c_i(x) = A_i x + b_i$) for $i = 1, \ldots, n_e$
- $d_j(x)$ convex for $j = 1, \ldots, n_i$

Active set

- The active set at a feasible point $x \in \Omega$ consists of the equality constraints and the inequality constraints for which $d_j(x) = 0$

$$\mathcal{A}(x) = \{c_i\}_{i=1}^{n_i} \cup \{d_j \mid d_j(x) = 0\}$$

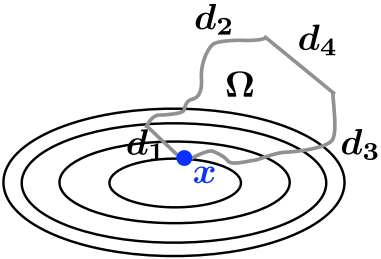


Figure: $\mathcal{A}(x) = \{d_1,\ d_3\}$

Formulation of first-order conditions

Words

to first-order, the function *cannot decrease* by moving in feasible directions

$\downarrow$

Geometric description

description using the geometry of the feasible set

$\downarrow$

Algebraic description

description using the equations of the active constraints

- The algebraic description is required to actually solve problems (use equations!)

First-order conditions for unconstrained problems

- **Geometric description**: a weak local minimum is a point x^* with a neighborhood $\mathcal{N}$ such that $f(x^*) \leq f(x)$ $\boxed{\forall x \in \mathcal{N}}$
- **Algebraic description**:

$$f(x^*) \leq f(x^* + p),\ \forall p \in \mathbb{R}^n \text{ “small”} \tag{2}$$

  - For $f(x^*)$ twice-continuously differentiable, Taylor's theorem is $f(x^*+p) = f(x^*) + \nabla f(x^*)^T p + \frac{1}{2} p^T \nabla^2 f(x^* + tp) p,\ \ t \in (0, 1)$
  - Ignoring the $O(\|p\|^2)$ term, (2) becomes

$$0 \leq f(x^* + p) - f(x^*) \approx \nabla f(x^*)^T p,\ \forall p \in \mathbb{R}^n$$

  - Since $p_1^T \nabla f(x^*) > 0$ implies that $p_2^T \nabla f(x^*) < 0$ with $p_2 = -p_1$, we know that strict equality must hold

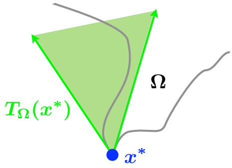
$\rightarrow$ This reduces to the first-order necessary condition: $\nabla f(x^*)^T p = 0\ \forall p \in \mathbb{R}^n \Rightarrow \boxed{\nabla f(x^*) = 0}$ (*stationary point*)

First-order conditions for constrained problems

- **Geometric description**: A weak local minimum is a point x^* with a neighborhood $\mathcal{N}$ such that $f(x^*) \leq f(x)\ \forall x \in \boxed{\mathcal{N} \cap \Omega}$

Definition (Tangent cone $T_\Omega(x^*)$)

The set of all tangents to Ω at x^*.
(set of *geometrically feasible directions*, the limit of $\mathcal{N} \cap \Omega - x^*$)



- Using the tangent cone, we can begin to formulate the first-order conditions algebraically

First-order conditions for constrained problems

- **Geometric description** (continued)
  - The limit of $f(x^*) \leq f(x)$, $\forall x \in \mathcal{N} \cap \Omega$ is

$$f(x^*) \leq f(x^* + p), \ \forall p \in T_\Omega(x^*) \text{ ``small''}$$

  - Using Taylor's theorem and ignoring high-order terms, this condition is

$$0 \leq f(x^* + p) - f(x^*) \approx \nabla f(x^*)^T p, \ \forall p \in T_\Omega(x^*)$$

$$\boxed{\nabla f(x^*)^T p \geq 0, \ \forall p \in T_\Omega(x^*)} \qquad (3)$$

$\longrightarrow$ To first-order, the objective function cannot decrease in any feasible direction

Constraint qualifications

- (3) is *not purely algebraic* since $T_\Omega(x^*)$ is geometric
- We require an *algebraic description* of the tangent cone in terms of the constraint equations

Definition (Set of linearized feasible directions $\mathcal{F}(x)$)

Given a feasible point x and the active constraint set $\mathcal{A}(x)$,

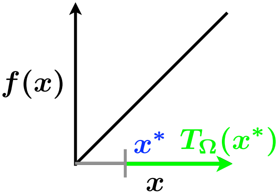
$$\mathcal{F}(x) = \left\{ p \mid p \text{ satisfies} \begin{cases} \nabla c_i(x)^T p = 0 & \forall i \\ \nabla d_j(x)^T p \geq 0 & \forall d_j \in \mathcal{A}(x) \end{cases} \right\}$$

- The set of linearized feasible directions is the best algebraic description available, but in general $T_\Omega(x) \subset \mathcal{F}(x)$
- *Constraint qualifications* are sufficient for $T_\Omega(x) = \mathcal{F}(x)$

Example

- Consider the following problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) = x \\ \text{subject to} \quad & d_1(x) = x - 3 \geq 0 \end{aligned}$$



- Since $d_1'(x^*) = 1$, $pd_1'(x^*) \geq 0$ for any $p \geq 0$, and we have

$$\mathcal{F}(x^*) = p, \quad \forall p \geq 0$$

- Thus, $\mathcal{F}(x^*) = T_\Omega(x^*)$ $\checkmark$

Example

- Consider the mathematically equivalent reformulation

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) = x \\ \text{subject to} \quad & d_1(x) = (x-3)^3 \geq 0 \end{aligned}$$

- The solution $x^* = 3$ and (geometric) tangent cone $T_\Omega(x^*)$ are unchanged
- However, $d_1'(x^*) = 3(3-3)^2 = 0$ and $pd_1'(x^*) \geq 0$ for any $p \in \mathbb{R}$ (positive or negative), and we have

$$\mathcal{F}(x^*) = p, \quad \forall p \in \mathbb{R} \quad \text{X}$$

- Thus, $T_\Omega(x^*) \subset \mathcal{F}(x^*)$, and directions in $\mathcal{F}(x^*)$ *may actually be infeasible!*

Constraint qualifications (sufficient for $T_\Omega(x^*) = \mathcal{F}(x^*)$)

- Types
  - **Linear independence constraint qualification (LICQ)**: the set of active constraint gradients at the solution $\{\nabla c_i(x^*)\}_{i=1}^{n_i} \cup \{\nabla d_j(x^*) \mid d_j(x^*) \in \mathcal{A}(^*x)\}$ is linearly independent
  - **Linear constraints**: all active constraints are linear functions
- None of these hold for the last example
- We proceed by assuming these conditions hold $(\mathcal{F}(x) = T_\Omega(x)) \Rightarrow$ the algebraic expression $\mathcal{F}(x)$ can be used to describe geometrically feasible directions at x

Algebraic description

- When constraint qualifications are satisfied, $\mathcal{F}(x) = T_\Omega(x)$ and (3) is
$$\nabla f(x^*)^T p \geq 0, \ \forall p \in \mathcal{F}(x^*) \tag{4}$$
- What form $\nabla f(x^*)$ ensures that (4) holds?
- **Equality constraints**: if we set $\nabla f(x^*) = \sum_{i=1}^{n_e} \gamma_i \nabla c_i(x^*)$, then
$$\nabla f(x^*)^T p = \sum_{i=1}^{n_e} \gamma_i \left(\nabla c_i(x^*)^T p\right) = 0, \quad \forall p \in \mathcal{F}(x^*) \quad \checkmark$$
- **Inequality constraints**: if we set $\nabla f(x^*) = \sum_{j=1}^{n_i} \lambda_j \nabla d_j(x^*)$ with $\lambda_j \geq 0$, then
$$\nabla f(x^*)^T p = \sum_{j=1}^{n_i} \lambda_j \left(\nabla d_j(x^*)^T p\right) \geq 0, \quad \forall p \in \mathcal{F}(x^*) \quad \checkmark$$

Theorem (First-order necessary KKT conditions for local solutions)

If x^* is a weak local solution of (1), constraint qualifications hold

$$\nabla f(x^*) - \sum_{i=1}^{n_e} \gamma_i \nabla c_i(x^*) - \sum_{j=1}^{n_i} \lambda_j \nabla d_j(x^*) = 0$$

$$\lambda_j \geq 0, \quad j = 1, \ldots, n_i$$

$$c_i(x^*) = 0, \quad i = 1, \ldots, n_e$$

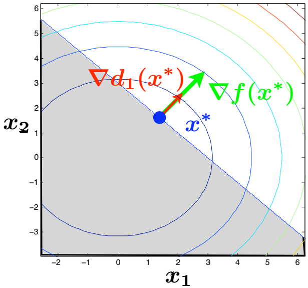
$$d_j(x^*) \geq 0, \quad j = 1, \ldots, n_i$$

$$\lambda_j d_j(x^*) = 0, \quad j = 1, \ldots, n_i$$

- Stationarity, Dual feasibility, Primal feasibility ($x^* \in \Omega$), Complementarity conditions, Lagrange multipliers γ_i, λ_j

Intuition for stationarity

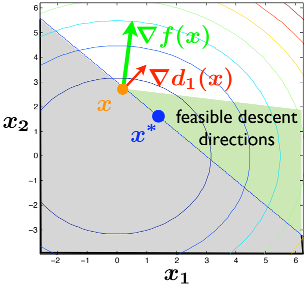
$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) = x_1^2 + x_2^2$$
$$\text{subject to} \quad d_1(x) = x_1 + x_2 - 3 \geq 0$$



- The solution is $x^* = (1.5, 1.5)$

Intuition for stationarity (continued)

- The KKT conditions say $\nabla f(x^*) = \lambda_1 \nabla d_1(x^*)$ with $\lambda_1 \geq 0$
- Here, $\nabla f(x^*) = [3,3]^T$, while $\nabla d_1(x^*) = [1.5, 1.5]^T$, so these conditions are indeed verified with $\lambda_1 = 2 \geq 0$
- This is obvious from the figure: if $\nabla f(x^*)$ and $\nabla d_1(x^*)$ were "misaligned", there would be some feasible descent directions!



- This gives us some intuition for stationarity and dual feasibility

Lagrangian

Definition (Lagrangian)

The Lagrangian for (1) is

$$\mathcal{L}(x,\gamma,\lambda) = f(x) - \sum_{i=1}^{n_e} \gamma_i c_i(x) - \sum_{j=1}^{n_i} \lambda_j d_j(x)$$

- Stationarity in the sense of KKT is equivalent to stationarity of the Lagrangian with respect to x:

$$\mathcal{L}_x(x,\gamma,\lambda) = \nabla f(x) - \sum_{i=1}^{n_e} \gamma_i \nabla c_i(x) - \sum_{j=1}^{n_i} \lambda_j \nabla d_j(x)$$

- KKT stationarity $\Leftrightarrow \mathcal{L}_x(x^*,\gamma,\lambda) = 0$

Lagrange multipliers

- Lagrange multipliers γ_i and λ_j arise in constrained minimization problems
- They tell us something about the *sensitivity* of $f(x^*)$ to the presence of their constraints. γ_i and λ_j indicate how hard f is "pushing" or "pulling" the solution against c_i and d_j.
- If we perturb the right-hand side of the i^{th} equality constraint so that $c_i(x) \geq -\epsilon\|\nabla c_i(x^*)\|$, then

$$\frac{df(x^*(\epsilon))}{d\epsilon} = -\gamma_i\|\nabla c_i(x^*)\|.$$

- If the j^{th} inequality is perturbed so $d_j(x) \geq -\epsilon\|\nabla d_j(x^*)\|$,

$$\frac{df(x^*(\epsilon))}{d\epsilon} = -\lambda_j\|\nabla d_i(x^*)\|.$$

Constraint classification

Definition (Strongly active constraint)

A constraint is strongly active at if it belongs to $\mathcal{A}(x^*)$ and it has:

- a strictly positive Lagrange multiplier for inequality constraints ($\lambda_j > 0$)
- a strictly non-zero Lagrange multiplier for equality constraints ($\gamma_i > 0$)

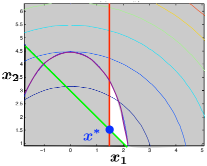
Definition (Weakly active constraint)

A constraint is weakly active at if it belongs to $\mathcal{A}(x^*)$ and it has a zero-valued Lagrange multiplier ($\gamma_i = 0$ or $\lambda_j = 0$)

Constraint classification (continued)

- Weakly active and inactive constraints "do not participate"

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) = x_1^2 + x_2^2 \\ \text{subject to} \quad & d_1(x) = x_1 + x_2 - 3 \geq 0 \text{ (strongly active)} \\ & d_2(x) = x_1 - 1.5 \geq 0 \text{ (weakly active)} \\ & d_3(x) = -x_1^2 - 4x_2^2 + 5 \geq 0 \text{ (inactive)} \end{aligned}$$



- The solution is unchanged if d_2 and d_3 are removed, so $\lambda_2 = \lambda_3 = 0$

Intuition for complementarity

- We just saw that non-participating constraints have zero Lagrange multipliers
- The complementarity conditions are

$$\lambda_j d_j(x^*) = 0, \quad j = 1, \ldots, n_i$$

- This means that each inequality constraint must be either:
  1. Inactive (non-participating): $d_j(x^*) > 0$, $\lambda_j = 0$,
  2. Strongly active (participating): $d_j(x^*) = 0$ and $\lambda_j > 0$, *or*
  3. Weakly active (active but non-participating): $d_j(x^*) = 0$ and $\lambda_j = 0$
- **Strict complementarity**: either case 1 or 2 is true for all constraints (no constraints are weakly active)

Second-order optimality conditions

- Second-order conditions for constrained optimization play a "tiebreaking" role: determine whether "undecided" directions for which $p^T\nabla f(x^*) = 0$ will increase or decrease f.
- We call these ambiguous directions the "critical cone"

Definition (Critical cone $\mathcal{C}(x^*, \gamma)$)

Directions that "adhere" to strongly active constraints and equality constraints

$$\mathcal{C}(x^*, \gamma) = \{w \in \mathcal{F}(x^*) \mid \nabla d_j(x^*)^T w = 0, \forall\, j \in \mathcal{A}(x^*) \text{ with } \lambda_j > 0\}$$

Note that $\lambda_j > 0$ implies the constraint will remain active even when small changes are made to the objective function!

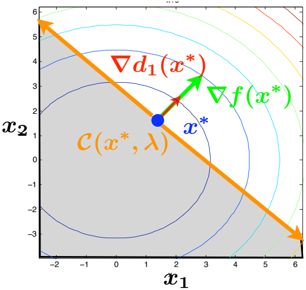
Critical cone

- For the problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) = x_1^2 + x_2^2$$

$$\text{subject to} \quad d_1(x) = x_1 + x_2 - 3 \geq 0$$

the critical cone is $\mathcal{C}(x^*, \gamma) = \alpha(-1, 1),\ \forall \alpha \in \mathbb{R}$



Second-order conditions for unconstrained problems

- Recall, second-order conditions for unconstrained problems

Theorem (Necessary conditions for a weak local minimum)

A1. $\nabla f(x^*) = 0$ *(stationary point)*
A2. $\nabla^2 f(x^*)$ *is positive semi-definite* $(p^T \nabla^2 f(x^*) p \geq 0$ *for all* $p \neq 0)$

Theorem (Sufficient conditions for a strong local minimum)

B1. $\nabla f(x^*) = 0$ *(stationary point)*
B2. $\nabla^2 f(x^*) > 0$ *is positive definite* $(p^T \nabla^2 f(x^*) p > 0$ *for all* $p \neq 0)$.

Second-order conditions for constrained problems

- We make an analogous statement for constrained problems, but limit the directions p to the critical cone $\mathcal{C}(x^*, \gamma)$

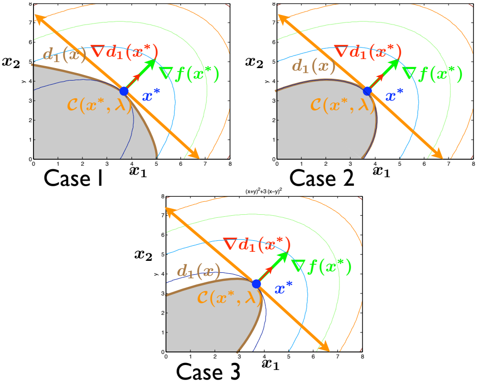
Theorem (Necessary conditions for a weak local minimum)

D1. KKT conditions hold
D2. $p^T \nabla^2 \mathcal{L}(x^*, \gamma) p \geq 0$ *for all* $p \in \mathcal{C}(x^*, \gamma)$

Theorem (Sufficient conditions for a strong local minimum)

E1. KKT conditions hold
E2. $p^T \nabla^2 \mathcal{L}(x^*, \gamma) p > 0$ *for all* $p \in \mathcal{C}(x^*, \gamma)$.

Intuition for second-order conditions



- **Case 1**: E1 and E2 are satisfied (sufficient conditions hold)
- **Case 2**: D1 and D2 are satisfied (necessary conditions hold)
- **Case 3**: D1 holds, D2 does not (necessary conditions **failed**)

Next

- We now know how to correctly formulate constrained optimization problems and how to verify whether a given point x could be a solution (necessary conditions) or is certainly a solution (sufficient conditions)
- Next, we learn algorithms that are use to *compute* solutions to these problems

Constrained optimization algorithms

- Linear programming (LP)
  - Simplex method: created by Dantzig in 1947. Birth of the modern era in optimization
  - Interior-point methods
- Nonlinear programming (NLP)
  - Penalty methods
  - Sequential quadratic programming methods
  - Interior-point methods
- Almost all these methods rely strongly on line-search and trust region methodologies for unconstrained optimization

Penalty methods

- Penalty methods combine the objective function and constraints

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) \quad \text{s.t. } c_i(x) = 0, \quad i = 1, \ldots, n_i$$

$$\downarrow$$

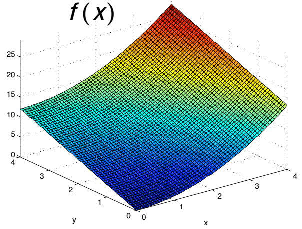
$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) + \frac{\mu}{2}\sum_{i=1}^{n_i} c_i^2(x)$$

- A sequence of *unconstrained* problems is then solved for μ increasing

Penalty methods example

- Original problem:

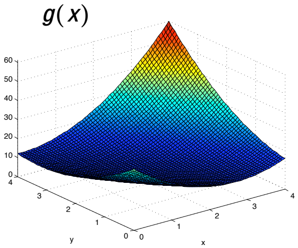
$$\underset{x\in\mathbb{R}^2}{\text{minimize}} \quad f(x) = x_1^2 + 3x_2, \quad \text{s.t. } x_1 + x_2 - 4 = 0$$



Penalty methods example

- Penalty formulation:

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad g(x) = x_1^2 + 3x_2 + \frac{\mu}{2}(x_1 + x_2 - 4)^2$$



- A valley is created along the constraint $x_1 + x_2 - 4 = 0$

Sequential quadratic programming

- Perhaps the most effective algorithm
- Solve a QP subproblem at each iterate

$$
\begin{aligned}
\underset{p}{\text{minimize}} \quad & \frac{1}{2} p^T \nabla^2_{xx} \mathcal{L}(x_k, \lambda_k) p + \nabla f(x_k)^T p \\
\text{subject to} \quad & \nabla c_i(x_k) T p + c_i(x_k) = 0, \quad i = 1, \ldots, n_e \\
& \nabla d_j(x_k)^T p + d_j(x_k) \geq 0, \quad j = 1, \ldots, n_i
\end{aligned}
$$

- When $n_i = 0$, this is equivalent to Newton's method on the KKT conditions
- When $n_i > 0$, this corresponds to an "active set" method, where we keep track of the set of active constraints $\mathcal{A}(x_k)$ at each iteration

Interior-point methods

- These methods are also known as "barrier methods," because they build a barrier at the inequality constraint boundary

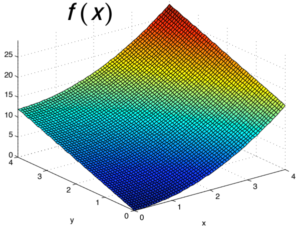
$$
\begin{aligned}
\underset{p}{\text{minimize}} \quad & f(x) - \mu \sum_{i=1}^{m} \log s_i \\
\text{subject to} \quad & c_i(x) = 0, \quad i = 1, \ldots, n_e \\
& d_j(x) - s_i = 0, \quad j = 1, \ldots, n_i
\end{aligned}
$$

- **Slack variables**: s_i, indicates distance from constraint boundary
- Solve a sequence of problems with μ decreasing

Interior-point methods example

- Original problem:

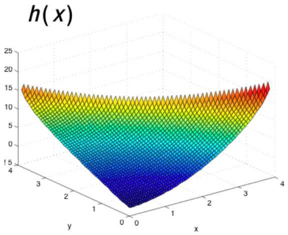
$$\underset{x\in\mathbb{R}^2}{\text{minimize}} \quad f(x) = x_1^2 + 3x_2, \quad \text{s.t.} \ -x_1 - x_2 + 4 \geq 0$$



Interior-point methods example

- Interior-point formulation:

$$\underset{x\in\mathbb{R}^2}{\text{minimize}} \quad h(x) = x_1^2 + 3x_2 - \log\left(-x_1 - x_2 + 4\right)$$



- A barrier is created along the boundary of the inequality constraint $x_1 + x_2 - 4 = 0$

Summary

- We now now something about:
  - Modeling and classifying unconstrained and constrained optimization problems
  - Identifying local minima (necessary and sufficient conditions)
  - Solving the problem using numerical optimization algorithms
- We next consider the case of PDE-constrained optimization, which enables us to use to tools learned earlier (finite elements) in optimal design and control settings, for example