Numerical Optimisation: Constraint Optimisation

Department of Computer Sc
Centre for Medical Image Com
Centre for Inverse Proble
University College London

Lecture 12

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

- $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$: objective function, assume smooth
- n

-

Feasible set Ω is a set of all points satisfying the co

$$\Omega = \{x \in \mathcal{D} : \ c_i(x) = 0, \ i \in$$

Optimal value: $x^\star = \inf_{x \in \Omega} f(x)$

- $x^\star = \infty$ if (COP) is infeasible i.e. $\Omega = \emptyset$
- $x^\star = -\infty$ if (COP) is unbounded below

Smooth constraints can describe regions with *kinks*.

Example: 1-norm:

$$\|x\|_1 = |x_1| + |x_2| \le 1$$

can be

x_1 [illegible] $\le 1.$

Example: pointwise max

$$\min f(x) = \max(x \ , x)$$

can be reformulated as

$$\min t, \quad \text{s.t.} \quad t \ge x, \quad t \ge x^2.$$

A point $x^\star \in \Omega$ is a **global minimiser** if

$$f(x^\star) \leq f(x), \ \forall x \in \Omega.$$

A point $x^\star \in \Omega$ is a **local minimiser** if

A point $x^\star \in \Omega$ is a **strict (or strong) local minimiser** if

$$\exists \mathcal{N}(x^\star) : f(x^\star) < f(x), \ \forall x \in$$

A point $x^\star \in \Omega$ is an **isolated local minimiser** if

$$\exists \mathcal{N}(x^\star) : x^\star \text{ is the only local minimiser in } \mathcal{N}(x^\star) \cap \Omega.$$

Feasibility problem: Find x such that all constraints are satisfied at x.

Acti

cons for

whic

$$\mathcal{A}(x) = \mathcal{E} \cup \{ i \in \mathcal{I} :$$

At a feasible point $x \in \Omega$, the inequality co
be *active* if $c_i(x) = 0$ and *inactive* if the s
$c_i(x) > 0$.

Single equality constraint

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

Feasibility: (Taylor expansion of c_1)

$c_1(x)^{\mathrm{T}} s$

Decrease direction: (Taylor expa

$0 > f$

The only situation that such s does not exist is if for some scalar λ_1

$$\nabla f(x) = \lambda_1 \nabla c_1(x).$$

Single inequality constraint

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

Feasibility: (Taylor expansion of c_1)

$c_1(x)^{\mathrm{T}} s$

expa

$0 > f$

Case: x inside the circle, i.e. $c_1(x) > 0$

$$s = -\alpha \nabla f(x)$$

Single inequality constraint

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

Feasibility: (Taylor expan-
sion of c_1)

$c_1(x)^{\mathrm{T}} s$

expa

$0 \geq f$

Case: x on the boundary of the circle, i.e. $c_1(x) = 0$

$$\nabla f(x)^{\mathrm{T}} s < 0, \quad \nabla c_1(x)^{\mathrm{T}} s \geq 0$$

Empty only if $\quad \nabla f(x) = \lambda_1 \nabla c_1(x) \quad$ for some $\lambda_1 \geq 0$.

Given the point x in the active set $\mathcal{A}(x)$, the **linear independent constraint qualification (LICQ)** holds if the set of of active constraint gradients $c_i(x)$, i (x) is linearly independent.

Note t

gradi

Example: LICQ is not satisfied if we define the equali

$c_1(x_2^2 + x_2^2 - 2)^2 = 0$ (same feasibility regio

There are other constraint qualifications e.g. Sl

convex problems.

Lagrangian function

$$\mathcal{L}(x,\lambda) = f(x) - \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i c_i(x).$$

Let $x^\star$ be a local solution of (COP) and f and c_i be continuously diffe **grange mult** follo $(x^\star, \lambda$)

$$\nabla_x \mathcal{L}(x^\star, \lambda^\star) = 0 \quad (1a)$$

$$c_i(x^\star) = 0, \quad (1b)$$

$$c_i(x^\star) \geq 0, \quad \forall i \in \mathcal{I}, \quad (1c)$$

$$\lambda^\star \geq 0, \quad \forall i \in \mathcal{I}, \quad (1d)$$

$$\lambda_i^\star c_i(x^\star) = 0, \quad \forall i \in \mathcal{E}\cup\mathcal{I}. \quad (1e)$$

The *complementarity condition* (2)(e) can be made stronger.

Given $x^\star$ a local solution of (COP) and a vector $\lambda^\star$ satisfying the KKT conditions (2), the **strict complementarity condition** holds if exac

word

Strict complementarity makes it easier for the algorithms to identify the active set and converge quickly to the so

For a given solution $x^\star$ of (COP), there ma * which satisfy the KKT condition (2). However, if LICQ holds, the optimal $\lambda^\star$ is unique.

For convenience we change (and refine) our notation for the constraint optimisation problem. The following slides are based on Boyd (Convex Optimization I).

Let p^* be the optimal value of the **primal problem**

i

The **Lagrangian** $\mathcal{L} : \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \to$

$$\mathcal{L}(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1} \nu_i h_i(x)$$

- λ_i are Lagrange multipliers associated with $f_i(x) \leq 0$
- ν_i are Lagrange multipliers associated with $h_i(x) = 0$

Lagrange dual function: $g: \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$g(\lambda,\nu) = \inf_{x\in\mathcal{D}} \mathcal{L}(x,\lambda,\nu) \qquad \text{(LD)}$$

m p

g: is co

Lower bound property: If $\lambda \geq 0$, then

Proof: For any feasible $\tilde{x}$ and $\lambda \geq 0$ we

$$f(\tilde{x}) \geq \mathcal{L}(\tilde{x},\lambda,\nu) \geq \inf_{x\in\mathcal{D}} \mathcal{L}(x,\lambda,\nu) = g(\lambda,\nu).$$

Minimising over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda,\nu)$.

Convex problem in standard form

$$\min_{x \in \mathcal{D} \subset \mathbb{R}^n} \quad f(x)$$

- f is convex and $\mathcal{D}$ is convex
- f_i are convex
- h_i are affine i.e. $a_i^T x = b_i$

Feasibility set Ω of a convex problem is a convex set.

$$\min_{x \in \mathbb{R}^n} \quad x^{\mathrm{T}} x$$

$$\text{subject to} \quad Ax = b$$

- Lagrangian: $\mathcal{L}(x,\nu) = x^{\mathrm{T}} x + \nu^{\mathrm{T}}(Ax - b)$
- Dual function:

- [illegible] radient

$$\nabla_x \mathcal{L}(x,\nu) = 2x + A^{\mathrm{T}}\nu = 0$$

- Plug $x_{\min}$ into g

$$g(\nu) = \mathcal{L}(x_{\min},\nu) = -\frac{1}{4}\nu A^{\mathrm{T}} A \nu - b^{\mathrm{T}}\nu.$$

g is a concave function of ν.

Lower bound property: $p^\star \geq -1/4\nu A^{\mathrm{T}} A\nu - b^{\mathrm{T}}\nu$ for all ν.

$$\min_{x\in\mathbb{R}^n} \quad c^{\mathrm{T}}x$$
$$\text{subject to} \quad Ax = b, \quad x \geq 0$$

- Lagrangian:
  $(x,\nu) = c^{\mathrm{T}}x + \nu^{\mathrm{T}}(Ax \quad b) \quad \lambda^{\mathrm{T}}x = \quad b^{\mathrm{T}}\nu + (c + A^{\mathrm{T}}\nu - \lambda)^{\mathrm{T}}x$

-
  $\lambda)^{\mathrm{T}}x)$
  $x\in\mathbb{R}^n$ $x\in\mathbb{R}^n$

- $\mathcal{L}(x,\nu)$ is affine in x, hence

  $$g(\lambda,\nu) = \inf_{x\in\mathbb{R}^n} \mathcal{L}(x,\nu) = \begin{cases} - & \\ -\infty, & \text{otherwise} \end{cases}$$
  g is linear on affine domain $\{(\lambda,\nu) : A^{\mathrm{T}}\nu - \lambda + c = 0\}$, hence concave.

Lower bound property: $p^\star \geq -b^{\mathrm{T}}\nu$ if $A^{\mathrm{T}}\nu + c \geq 0$.

$$\min_{x \in \mathbb{R}^n} \quad \|x\|$$
$$\text{subject to} \quad Ax = b$$

- Lagrangian:

$$\mathcal{L}(x, \nu) = \|x\| + \nu^{\mathrm{T}}(b - Ax) = \|x\| + b^{\mathrm{T}}\nu - \nu^{\mathrm{T}}Ax$$

- ere

$x = 0.$

If $\|y\|_\star > 1$, choose $x = tu$, u : $\|$

$$\|x\| - \nu^{\mathrm{T}}x = t(\|u\| - \|y\|$$

- Dual function:

$$g(\nu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \nu) = \begin{cases} b^{\mathrm{T}}\nu, & \|A^{\mathrm{T}}\nu\|_\star \leq 1 \\ -\infty, & \text{otherwise} \end{cases}$$

Lower bound property: $p^\star \geq b^{\mathrm{T}}\nu$ if $\|A^{\mathrm{T}}\nu\|_\star \leq 1$.

The **conjugate** of function f is

$$f^\star(y) = \sup_{x \in \mathcal{D}} (y^{\mathrm{T}} x - f(x))$$

The conjugate $f^\star$ is convex (even if f is not)

Figure: Boyd, Convex Optimization I

$$\min_{x\in\mathbb{R}^n} \quad f(x)$$

$$\text{subject to} \quad Ax \le b, \quad Cx = d$$

- Lagrangian:

$$\ldots {}^{\mathrm{T}}\nu.$$

- Dual function:

$$\begin{aligned} g(\lambda,\nu) &= \inf_{x\in\mathcal{D}} \left(f(x) + (A^{\mathrm{T}}\lambda + \ldots\right. \\ &= -\sup_{x\in\mathcal{D}} \left(-f(x) - (A^{\mathrm{T}}\lambda + C^{\mathrm{T}}\nu)^{\mathrm{T}}x - b^{\mathrm{T}}\lambda - d^{\mathrm{T}}\nu\right) \\ &= -f^\star(-A^{\mathrm{T}}\lambda - C^{\mathrm{T}}\nu) - b^{\mathrm{T}}\lambda - d^{\mathrm{T}}\nu \end{aligned}$$

Lagrange dual problem

$$\begin{aligned} \text{max} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & \lambda \quad 0 \end{aligned}$$

- dual
- is a convex optimization problem, we deno
  with d^*

- λ, ν are dual feasible if $\lambda \geq 0$, $(\lambda,$
- often simplified by making implicit constraint $(\lambda, \nu) \in \operatorname{dom} g$, explicit

Weak duality: $d^\star \leq p^\star$

- always holds (for convex and nonconvex problems)
-

Strong duality: $d^\star = p^\star$

- does not hold in general
- holds for convex problems under

Slater's constraint qualification

Strong duality holds for a convex problem

$$\begin{aligned} \min_{x \in \mathcal{D}} \quad & f(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{aligned}$$

if it is stri

$$i$$

- also guarantees that the dual optimum is att
  $p^\star > -\infty$)
- can be sharpened: e.g. can replace int$\mathcal{D}$ with relint$\mathcal{D}$ (interior of the affine hull); linear inequalities do not need to hold with strict inequality, ...
- other constraint qualifications exist e.g. LICQ

Primal problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^{\mathrm{T}} x \\ \text{subject to} \quad & Ax \leq b \end{aligned}$$

Dual function

$$g(\lambda) = \begin{cases} -b^{\mathrm{T}}\lambda & A^{\mathrm{T}}\lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Dual problem

$$\begin{aligned} \max \quad & -b^{\mathrm{T}}\lambda \\ \text{subject to} \quad & A^{\mathrm{T}}\lambda + c = 0, \quad \lambda \geq 0 \end{aligned}$$

- From Slater's condition: $p^\star = d^\star$ if $\exists \tilde{x} : \ A\tilde{x} < b$
- In fact, $p^\star = d^\star$ except when primal and dual are infeasible

Example: Quadratic program

Primal problem (assume P symmetric positive definite)

$$\min_{x\in\mathbb{R}^n} \quad x^{\mathrm{T}}Px$$

$$\text{subject to} \quad Ax \leq b$$

Dual function

$$g(\quad - \quad b^{\mathrm{T}}\lambda$$

Dual problem

$$\max \quad -\frac{1}{4}\lambda^{\mathrm{T}}AP$$

$$\text{subject to} \quad \lambda \geq 0$$

- From Slater's condition: $p^\star = d^\star$ if $\exists \tilde{x}: \; A\tilde{x} < b$
- In fact, $p^\star = d^\star$ always

Example: nonconvex problem with strong duality

Primal problem

$$\begin{array}{ll} \min_{x\in\mathbb{R}^n} & x^{\mathrm{T}}Ax + 2b^{\mathrm{T}}x \\ \text{subject to} & x^{\mathrm{T}}x \leq 1 \end{array}$$

$A \not\succeq$

Dual f

$$g(\lambda) = \inf_{x\in\mathbb{R}^n} \left(x^{\mathrm{T}}(A+\lambda I)\right)$$

- unbounded below if $A+\lambda I \not\succeq 0$ or if $A+\lambda I \succeq 0$ and $b \notin \mathcal{R}(A+\lambda I)$
- otherwise minimised by $x = -(A+\lambda I)^{\dagger}b$: $g(\lambda) = -b^{\mathrm{T}}(A+\lambda I)^{\dagger}b - \lambda$

Dual problem

$$
\begin{aligned}
\max \quad & -b^{\mathrm{T}}(A+\lambda I)^{\dagger} b - \lambda \\
\text{subject to} \quad & A + \lambda I \succeq 0 \\
& b \quad (A+\lambda I)
\end{aligned}
$$

and e

$$
\text{subject to} \quad \begin{bmatrix} A+\lambda & b \\ b^{\mathrm{T}} & \end{bmatrix}
$$

Strong duality although primal problem is not convex (not easy to show).

KKT conditions revisited

Karush-Kuhn-Tucker conditions are satisfied at $x^\star, \nu^\star, \lambda^\star$ i.e.

$$\nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i \nabla h_i(x^\star) = 0, \quad (2a)$$

$\lambda_i^\star f_i(x^\star) = 0, \quad i = 1, \dots, m$ [com

Necessary condition: If strong duality hol
optimal, then they must satisfy KKT conditions.

For *any* problem for which strong duality holds, KKT are necessary conditions.

Sufficient condition: If $x^\star, \nu^\star, \lambda^\star$ satisfy KKT conditions and the problem is convex, then $x^\star, \nu^\star, \lambda^\star$ are primal and dual optimal:

- from complementary slackness: $f(x^\star) = f(x^\star) + \sum_{i=1}^{m} \underbrace{\lambda_i f_i(x^\star)}_{=0} + \sum_{i=1}^{p} \underbrace{\nu_i h_i(x^\star)}_{=0} = \mathcal{L}(x^\star, \lambda^\star, \nu^\star)$
- $\star \quad \star \quad \star \quad \star$

Thus i

If Slater's condition is satisfied:
$x^\star$ is optimal if and only if there exists

conditions

- recall that Slater implies strong duality, and that the dual optimum is attained
- generalises optimality condition $\nabla f(x) = 0$ for unconstrained problem