

Numerical Optimisation:
Constraint Optimisation

Assignment Project Exam Help

Marta M. Betcke

<https://powcoder.com>
m.betcke@ucl.ac.uk,
f.rullan@cs.ucl.ac.uk

Add WeChat
Department of Computer Science,
Centre for Medical Image Computing,
Centre for Inverse Problems
University College London

Lecture 12

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

Assignment Project Exam Help

- $f: \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$: objective function, assume smooth
- $c_i: \mathbb{R}^n \rightarrow \mathbb{R}$: constraint function, assume smooth
 - $i \in \mathcal{E}$ equality constraints,
 - $i \in \mathcal{I}$ inequality constraints.
- $x \in \mathbb{R}^n$: optimisation variable

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

$$\Omega = \{x \in \mathcal{D} : c_i(x) = 0, i \in \mathcal{E}; c_i \geq 0, i \in \mathcal{I}\}.$$

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

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

Examples: smooth constraints

Smooth constraints can describe regions with *kinks*.

Example: 1-norm:

Assignment Project Exam Help

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

can be described as

$$x_1 + x_2 \leq 1, \quad x_1 - x_2 \leq 1, \quad -x_1 + x_2 \leq 1, \quad -x_1 - x_2 \leq 1.$$

Example: pointwise max

Add WeChat powcoder

$$\min f(x) = \max(x^2, x)$$

can be reformulated as

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

Types of minimisers of constraint problems

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

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

Assignment Project Exam Help

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

$$\exists \mathcal{N}(x^*) : f(x^*) \leq f(x), \forall x \in \mathcal{N}(x^*) \cap \Omega$$

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

$$\exists \mathcal{N}(x^*) : f(x^*) < f(x), \forall x \in \mathcal{N}(x^*) \cap \Omega, x \neq x^*$$

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

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

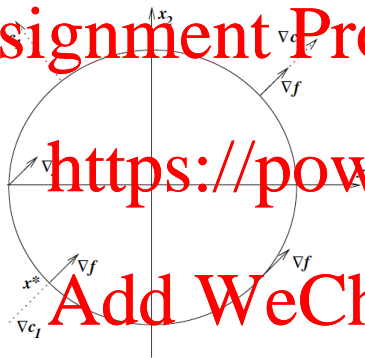
Feasibility problem: Find x such that all constraints are satisfied at x .

Active set $\mathcal{A}(x)$ at any feasible $x \in \Omega$ consists of the equality constraint indices set \mathcal{E} and the inequality constraints $i \in \mathcal{I}$ for which $c_i(x) = 0$.

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} : c_i(x) = 0\}.$$

At a feasible point $x \in \Omega$, the inequality constraint $i \in \mathcal{I}$ is said to be *active* if $c_i(x) = 0$ and *inactive* if the strict inequality holds $c_i(x) > 0$.

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



Feasibility: (Taylor expansion of c_1)

$$0 = c_1(x+s) \approx \underbrace{c_1(x)}_{=0} + \nabla c_1(x)^T s$$

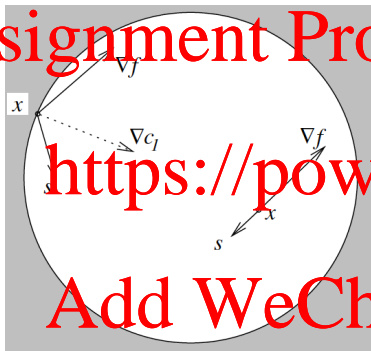
Decrease direction: (Taylor expansion of f)

$$0 > f(x+s) - f(x) \approx \nabla f(x)^T s$$

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

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

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



Feasibility: (Taylor expansion of c_1)

$$0 \leq c_1(x+s) \approx c_1(x) + \nabla c_1(x)^T s$$

Decrease direction: (Taylor expansion of f)

$$0 > f(x+s) - f(x) \approx \nabla f(x)^T s$$

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

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

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

Feasibility: (Taylor expansion of c_1)

$$0 \leq c_1(x+s) \approx \underbrace{c_1(x)}_{=0} + \nabla c_1(x)^T s$$

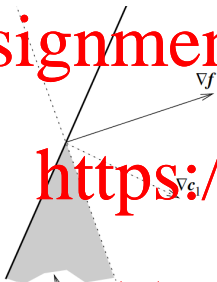
Decrease direction: (Taylor expansion of f)

$$0 \geq f(x+s) - f(x) \approx \nabla f(x)^T s$$

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

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

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



Any d in this cone is a good search direction, to first order.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Linear independent constraint qualification (LICQ)

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

Note that for LICQ to be satisfied, none of the active constraint gradients can be 0.

Example: LICQ is not satisfied if we define the equality constraint $c_1(x_2^2 + x_2^2 - 2)^2 = 0$ (same feasibility region, different constraint)

There are other constraint qualifications e.g. Slater's conditions for convex problems.

Theorem: 1st order necessary conditions

Lagrangian function

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

Let x^* be a local solution of (COP) and f and c_i be continuously differentiable and LICQ hold at x^* . Then there exists a **Lagrange multiplier** λ^* with components λ_i^* , $i \in \mathcal{E} \cup \mathcal{I}$ such that the following **Karush-Kuhn-Tucker conditions** are satisfied at (x^*, λ^*)

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

$$c_i(x^*) = 0, \quad \forall i \in \mathcal{E}, \quad (1b)$$

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

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

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

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

Given x^* a local solution of (COP) and a vector λ^* satisfying the KKT conditions (2), the **strict complementarity condition** holds if exactly one of λ_i^* and $c_i(x^*)$ is zero for each $i \in \mathcal{I}$. In other words, $\lambda_i^* > 0$ for each $i \in \mathcal{I} \cap \mathcal{A}(x^*)$.

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

For a given solution x^* of (COP), there may be many vectors λ^* which satisfy the KKT condition (2). However, if LICQ holds, the optimal λ^* is unique.

Lagrangian: primal problem

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

Assignment Project Exam Help

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

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

subject to $f_i(x) \leq 0, \quad i = 1, \dots, m,$

$$h_i(x) = 0, \quad i = 1, \dots, p$$

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

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

- 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: $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$

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

$$= \inf_{x \in \mathcal{D}} \left(f(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right).$$

g : is concave, can be $-\infty$ for some λ, ν .

Lower bound property: If $\lambda \geq 0$, then $g(\lambda, \nu) \leq p^*$.

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

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

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

Convex problem in standard form

Assignment Project Exam Help

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

subject to $f_i(x) \leq 0, \quad i = 1, \dots, m,$

$h_l(x) = 0, \quad l = 1, \dots, p$
<https://powcoder.com>

- f is convex and \mathcal{D} is convex
- f_i are convex
- h_l are affine i.e. $a_l^T x = b_l$

Add WeChat powcoder

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

Example: least norm solution of linear equations

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

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

Assignment Project Exam Help

- Lagrangian $\mathcal{L}(x, \nu) = x^T x + \nu^T (Ax - b)$
- Dual function:
$$g(\nu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \nu) = \inf_{x \in \mathbb{R}^n} (x^T x + \nu^T (Ax - b))$$
- $\mathcal{L}(x, \nu)$ is strictly convex in x , to minimise over x set gradient equal zero:

$$\nabla_x \mathcal{L}(x, \nu) = 2x + A^T \nu = 0 \quad \Rightarrow \quad x_{\min} = -1/2 A^T \nu$$

- Plug x_{\min} into g

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

g is a concave function of ν .

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

Example: standard form LP

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

Assignment Project Exam Help

- Lagrangian:

$$\mathcal{L}(x, \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x = -b^T \nu + (c + A^T \nu - \lambda)^T x$$

- Dual function:

<https://powcoder.com>

$$g(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \nu) = \inf_{x \in \mathbb{R}^n} (-b^T \nu + (c + A^T \nu - \lambda)^T x)$$

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

$$g(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \nu) = \begin{cases} -b^T \nu, & A^T \nu - \lambda + c = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

g is linear on affine domain $\{(\lambda, \nu) : A^T \nu - \lambda + c = 0\}$,
hence concave.

Lower bound property: $p^* \geq -b^T \nu$ if $A^T \nu + c \geq 0$.

Example: equality constraint norm minimisation

$$\min_{x \in \mathbb{R}^n} \|x\|$$

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

Assignment Project Exam Help

- Lagrangian

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

- $\inf_{x \in \mathbb{R}^n} (\|x\| - y^T x) = 0$ if $\|y\|_* \leq 1$, $-\infty$ otherwise, where $\|y\|_* = \sup_{\|u\| \leq 1} u^T y$ is dual norm of $\|\cdot\|$.
If $\|y\|_* \leq 1$, then $\|x\| - y^T x \geq 0, \forall x$, with equality if $x = 0$.
If $\|y\|_* > 1$, choose $x = tu, u: \|u\| \leq 1, u^T y = \|y\|_* > 1$

Add WeChat powcoder

- Dual function:

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

Lower bound property: $p^* \geq b^T \nu$ if $\|A^T \nu\|_* \leq 1$.

Conjugate function

The **conjugate** of function f is

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

The conjugate f^* is convex (even if f is not)

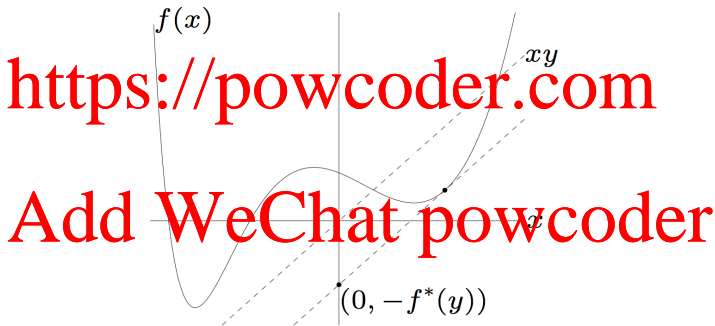


Figure: Boyd, Convex Optimization I

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

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

Assignment Project Exam Help

- Lagrangian:

$$\begin{aligned}\mathcal{L}(x, \lambda, \nu) &= f(x) + \lambda^T (Ax - b) + \nu^T (Cx - d) \\ &= f(x) + (A^T \lambda + C^T \nu)^T x - b^T \lambda - d^T \nu.\end{aligned}$$

- Dual function:

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} (f(x) + (A^T \lambda + C^T \nu)^T x - b^T \lambda - d^T \nu)$$

$$= - \sup_{x \in \mathcal{D}} (-f(x) - (A^T \lambda + C^T \nu)^T x - b^T \lambda - d^T \nu)$$

$$= -f^*(-A^T \lambda - C^T \nu) - b^T \lambda - d^T \nu$$

Lagrange dual problem

Assignment Project Exam Help

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

- finds the best lower bound on p^* obtained from Lagrange dual function
- is a convex optimization problem, we denote its optimal value with d^*
- λ, ν are dual feasible if $\lambda \geq 0, (\lambda, \nu) \in \text{dom } g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \text{dom } g$, explicit

Assignment Project Exam Help

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

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

<https://powcoder.com>

Strong duality: $d^* = p^*$

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

Add WeChat powcoder

Slater's constraint qualification

Strong duality holds for a convex problem

$$\min_{x \in \mathcal{D}} f(x)$$

$$\text{subject to } f_i(x) \leq 0, \quad i = 1, \dots, m$$

$$Ax = b,$$

if it is strictly feasible i.e.

$$\exists x \in \text{int}\mathcal{D} : f_i(x) < 0, \quad i = 1, \dots, m, \quad Ax = b$$

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

Example: inequality form LP

Primal problem

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

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

Dual function

$$g(\lambda) = \inf_{x \in \mathbb{R}^n} ((c + A^T \lambda)^T x - b^T \lambda) = \begin{cases} -b^T \lambda, & A^T \lambda + c = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Dual problem

$$\max_{\lambda} -b^T \lambda$$

$$\text{subject to } A^T \lambda + c = 0, \quad \lambda \geq 0$$

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

Example: Quadratic program

Primal problem (assume P symmetric positive definite)

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

subject to $Ax \leq b$

Dual function

$$g(\lambda) = \inf_{x \in \mathbb{R}^n} (x^T P x + \lambda^T (Ax - b)) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

Dual problem

$$\max \quad -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

subject to $\lambda \geq 0$

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

Example: nonconvex problem with strong duality

Primal problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & x^T A x + 2b^T x \\ \text{subject to} \quad & x^T x \leq 1 \end{aligned}$$

$A \not\succeq 0$ is not positive definite.

Dual function

$$g(\lambda) = \inf_{x \in \mathbb{R}^n} (x^T (A + \lambda I) x + 2b^T x - \lambda)$$

- 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 $x = -(A + \lambda I)^\dagger b$:
$$g(\lambda) = -b^T (A + \lambda I)^\dagger b - \lambda$$

Dual problem

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

and equivalent semidefinite program:

<https://powcoder.com>

$$\max \quad -t - \lambda$$

$$\text{subject to} \quad \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0$$

Add WeChat powcoder

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

Karush-Kuhn-Tucker conditions are satisfied at x^*, ν^*, λ^* i.e.

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

$$h_i(x^*) = 0, \quad i = 1, \dots, p \quad [\text{primary constraints}] \quad (2b)$$

$$f_i(x^*) \leq 0, \quad i = 1, \dots, m \quad [\text{primary constraints}] \quad (2c)$$

$$\lambda_i^* \geq 0, \quad i = 1, \dots, m \quad [\text{dual constraints}] \quad (2d)$$

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \dots, m \quad [\text{complementary slackness}] \quad (2e)$$

Necessary condition: If strong duality holds and x, ν, λ are optimal, then they must satisfy KKT conditions.

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

KKT conditions for convex problem

Sufficient condition: If x^*, ν^*, λ^* satisfy KKT conditions and the problem is convex, then x^*, ν^*, λ^* are primal and dual optimal:

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

- $g(\lambda^*, \nu^*) = \inf_x \mathcal{L}(x, \lambda^*, \nu^*)$ and from the 1st order necessary condition and convexity of f we have that the minimum is attained at x^* , hence $g(\lambda^*, \nu^*) = \mathcal{L}(x^*, \lambda^*, \nu^*)$

Thus it follows that $f(x^*) = g(\lambda^*, \nu^*)$.

If Slater's condition is satisfied:

x^* is optimal if and only if there exists λ^*, ν^* that satisfy KKT conditions

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