Lecture 9
Decomposition and Distributed Optimization

Mikael Johansson
S3/Automatic Control
mikaelj@s3.kth.se

Anders Forsgren
Math/OptSyst
andersf@kth.se

Aims

After this lecture, you should be able to

- bound optimal value by solving (convex but maybe non-smooth) dual
- compute subgradients and use the subgradient method
- use dual decomposition for efficient large-scale optimization
- derive distributed optimization schemes using decomposition techniques

Disposition

- Lagrange duality review
- Subgradients and non-differentiable optimization
- Dual optimization
- Distributed optimization
- Primal decomposition

The Lagrangian

Optimization problem in standard form

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \le 0 & i = 1, \dots, m \\ & h_i(x) = 0 & i = 1, \dots, p \end{array}$$

variable $x \in \mathbb{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$.

Lagrangian: $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \mapsto \mathbb{R}$ with domain $\mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

Weighted sum of objective and constraint functions,

- λ_i is *Lagrange multiplier* (or *dual variable)* associated with $f_i(x) \le 0$
- ν_i is *Lagrange multiplier* associated with $h_i(x)=0$

Lagrange dual: concave and lower bounding

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

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

Fact: g is concave (can be $-\infty$ for some λ, ν)

Fact (lower bound property): if $\lambda \succeq 0$, then $g(\lambda,\nu) \le p^\star$

Dual problem: best lower bound

Lagrange dual problem:

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

Finds *best lower bound* $d^\star$ on $p^\star$ obtained from Lagrange dual function

Observations:

- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$ and $(\lambda, \nu) \in$ **dom** g
- often simplified by making the constraint $(\lambda, \nu) \in$ **dom** g explicit

Duality: structure of solution

If strong duality holds, optimal ($x^\star$, $\lambda^\star$, $\nu^\star$) must satisfy KKT conditions:

1. Primal constraints:
   $f_i(x^\star) \le 0,\ i = 1,\dots,m, \quad h_i(x^\star) = 0,\ i = 1,\dots,p$
2. Dual constraints:
   $\lambda^\star \succeq 0$
3. Complementary slackness:
   $\lambda_i^\star f_i(x^\star) = 0,\ i = 1,\dots,m$
4. Gradient of Lagrangian w.r.t. x vanishes
   $$\nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star)$$

Duality: sensitivity analysis

Nominal and perturbed problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0 \quad i = 1,\dots,m \end{array} \qquad \begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le u_i \quad i = 1,\dots,m \end{array}$$

$p^\star(u)$ is optimal value as function of u

What can we say about $p^\star(u)$ by solving the nominal primal and its dual?

By weak duality,

$$p^\star(u) \ge p^\star(0) - \sum_{i=1}^{m} \lambda_i^\star u_i$$

If $p^\star(u)$ differentiable

$$\lambda_i = -\frac{\partial p^\star(0)}{\partial u_i}$$

λ_i is the *sensitivity* of $p^\star$ with respect to ith constraint.

Duality: efficient optimization

Today's lecture: duality sometimes enables efficient optimization

Disposition

- Lagrange duality review
- Subgradients and non-differentiable optimization
- Dual optimization
- Distributed optimization
- Primal decomposition

Subgradients

A *subgradient* h of a convex function f at x is any vector that satisfies

$$f(y) \ge f(x) + h^T(x)(y - x) \qquad \text{for all } y$$

Subgradients
- gives affine global underestimator of f
- if f is convex and differentiable, $\nabla f(x)$ is a subgradient of f at x

The set of subgradients at x is called the subdifferetial, denoted ∂ f(x)

Quiz: determine ∂ f(x) for f(x) = |x|,

Subgradient method

For convex function f, subgradient method uses iteration

$$x(t+1) = x(t) - \alpha(t) h(x(t))$$

where
- x(t) is t^{th} iterate, $g^{(k)}$ *any* subgradient, a(t) >0 is step-length

Guaranteed to converge if (for example) h bounded and α satisfies

$$\sum_t \alpha^2(t) < \infty, \qquad \sum_{t=1}^{\infty} \alpha(t) = \infty$$

Proof:

Common choice:

$$\alpha(t) = \frac{a}{t+b} \quad \text{for some parameters } a > 0,\ b \ge 0$$

Extension: projected subgradient

When solving the dual problem, we need to enforce $\lambda \succeq 0$

Can be done by projected subgradient method

$$x(t+1) = [x(t) - \alpha(t) h(x(t))]^{+}$$

where $[\cdot]^{+}$ denotes projection on the positive orthant

Same steplength rules and convergence results as simple method

Example: piecewise linear minimization

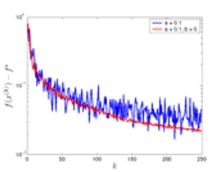
Consider the problem

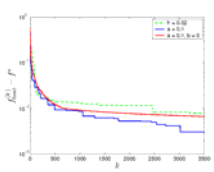
$$\text{minimize} \quad f(x) = \max_{k=1,\dots,K} \left\{ a_k^T x + b_k \right\}$$

Quiz: determine the associated subgradient update

Example: PWL minimization cont'd

Convergence of subgradient method for different stepsize rules



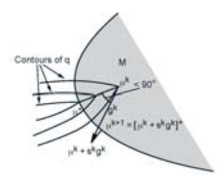


Practical observation: convergence is slow (in the limit)

Subgradient method

Not a descent method

- function value iterates may increase
- distance to optimal set guaranteed to decrease (if stepsize small enough)



Subgradients and supgradients

Similarly, a supgradient of a function g at x is any vector h such that

$$g(y) \leq g(x) + h^T(x)(y - x) \qquad \text{for all } y$$

(supgradients give global overestimators)

For concave maximization, supgradient method reads

$$x^{(k+1)} = \left[x^{(k)} + \alpha^{(k)} h^{(k)} \right]^{+}$$

Many texts do not make a difference between sup and subgradients.

Example: supgradient of dual function

Let x_λ be a minimizer for the Lagrangian, *i.e.*,

$$x_\lambda = \arg\min_{x \in X} L(x, \lambda) = \arg\min_{x \in X} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right\}$$

then $(f_1(x_\lambda) \dots f_m(x_\lambda))$ is a supgradient of the dual function at λ

$$g(\bar{\lambda}) \leq g(\lambda) + \sum_{i=1}^{m} (\bar{\lambda}_i - \lambda_i) f_i(x_\lambda) \quad \text{for all } \bar{\lambda} \in \mathbb{R}^m$$

Proof:

Observation: subgradient for free when evaluating dual function!

When is the dual differentiable?

Fact: If f_0 is strictly convex, $f_1, \ldots, f_m$ linear, then g is differentiable (much weaker formulations exist)

Disposition

- Lagrange duality review
- Subgradients and non-differentiable optimization
- Dual optimization
- Distributed optimization
- Primal decomposition

Basic idea

Optimization problem in many variables

$$\begin{array}{lll} \text{minimize} & \sum_{k=1}^{n} f_{0k}(x_k) & \\ \text{subject to} & \sum_{k=1}^{n} f_{ik}(x_k) \leq 0 & i = 1, \ldots, m \\ & x_k \in X_k & i = 1, \ldots, n \end{array}$$

with few coupling constraints (i.e., n is large, m is small)

Introducing dual variables for coupling constraints only

$$\begin{aligned} g(\lambda) &= \inf_x \sum_k f_{0k}(x_k) + \sum_i \lambda_i \sum_k f_{ik}(x_k) \\ &= \sum_{k=1}^{n} \inf_{x \in X_k} f_{0k}(x_k) + \sum_i \lambda_i f_{ki}(x_k) \end{aligned}$$

Dual function separable – easy to evaluate if X_k has simple structure

Solving the dual

Dual problem can be solved with projected supgradient method

$$\lambda_i(t+1) = \left[\lambda_i(t) + \alpha(t) \sum_k f_{ik}(x_k^\star(t))\right]^+$$

At each iteration, $g(\lambda)$ underestimates optimal objective

Convergence in the limit

Example: resource allocation

$$\begin{array}{ll} \text{minimize} & -\sum_{k=1}^{n} \log(x_k + \alpha_k) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x \leq 1 \end{array}$$

Introduce Lagrange multiplier λ for coupling constraint $\mathbf{1}^T x \leq 1$

Dual function

$$g(\lambda) = \sum_k \left(\inf_{x_k \geq 0} -\log(x_k + \alpha_k) + \lambda x_k \right) - \lambda$$

Inner problems solved by

$$x_k^\star = \max\{0, 1/\nu - \alpha_k\}$$

Subgradient method: $\lambda(t+1) = [\lambda + \alpha(t)(\sum x_k - 1)]^+$

- Increase multipliers when resource overutilized, decrease if underutilized

Pricing or tax interpretation

Operation of enterprise subject to resource constraints

$$\begin{array}{ll} \text{minimize} & \sum_i f_i(x_i) \\ \text{subject to} & \sum_i x_i \leq x_{\text{tot}} \end{array}$$

$f_i(x_i)$ cost of operating system i using x_i amount of resources

Dual function

$$g(\lambda) = \sum_i \inf_{x_i} f_i(x_i) + \lambda x_i$$

Interpret λ as resource price: systems operate to minimize total cost

Subgradient method:
- increase resource prices when resource is overutilized,
- decrease resource prices when resource is underutilized

Problems and challenges

Slow convergence of subgradient method

- cutting-plane methods faster, but computationally more intensive

Primal iterates $x^\star(t)$ not necessarily feasible

- coupling constraints not enforced in dual formulation
- need structure, or heuristic, to produce (suboptimal) primal solutions

Disposition

- Lagrange duality review
- Subgradients and non-differentiable optimization
- Dual optimization
- Distributed optimization
- Primal decomposition

Distributed optimization

In some cases, decomposition schemes reveal solutions that rely on

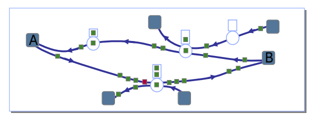
- distributed optimization (in local variables)
- and coordination mechanisms (dual variable updates/prices)

A powerful methodology for finding decentralized solutions!

We will simply exemplify this on a model of Internet congestion control.

Internet congestion control

Users contend for limited network resources



Congestion control: helps share resources fairly and efficiently

- When router buffer fills up, packets are marked (or dropped)
- Users react to congestion signals by reducing transmission rate

Interplay between control mechanisms in routers and end nodes

Internet Congestion Control

Current Internet

- source algorithm is some Transmission Control Protocol (TCP)
- link algorithm is some Active Queue Management (AQM) scheme

Two types of fundamental studies

- equilibrium properties (fairness, throughput, ...) – via cvx. opt.
- dynamical properties (stability, convergence) – control theory

Our focus: equilibrium properties using a dual decomposition approach

Notation

Network with L links of finite capacities c_l, l=1, ..., L

Capacity shared by P source-destination pairs sending at rate s_p

Sending rates must satsify capacity constraints

$$\sum_p r_{lp} s_p \leq c_l \quad \text{where } r_{lp} = \begin{cases} 1 & \text{if source } s_p \text{ sends data across link } l \\ 0 & \text{otherwise} \end{cases}$$

Capacity constraints on vector form

$$r_l^T s \leq c_l \qquad l = 1, \ldots, L$$

The utility maximization problem

Optimal network operation from solving utility maximization problem

$$\begin{array}{lll} \text{maximize} & \sum_{p=1}^{P} u_p(s_p) & \\ \text{subject to} & r_l^T s \le c_l & l = 1, \dots, L \\ & s_p \ge 0 & p = 1, \dots, P \end{array}$$

where u_p is a strictly concave *utility function*

Key observation: the dual problem admits decentralized solution

- links update congestion measures (Lagrange multipliers)
- sources react to congestion signals, update source rates

Gives insight into current (and future) congestion control schemes!

A dual approach

Lagrangian of utility maximization problem

$$L(x, \lambda) = \left\{ \sum_{p=1}^{P} u_p(s_p) - \sum_{l=1}^{L} \lambda_l (r_l^T s - c_l) \mid s_p \ge 0,\ p = 1, \dots, P \right\}$$

Dual function

$$g(\lambda) = \max_{s \succeq 0} \sum_{p=1}^{P} u_p(s_p) - \sum_{l=1}^{L} \lambda_l (r_l^T s - c_l) = \cdots =$$

$$= \sum_{p=1}^{P} \max_{s_p \ge 0} u_p(s_p) - \Big(\sum_{l \in \mathcal{L}(p)} \lambda_l \Big) s_p + \sum_{l=1}^{L} \lambda_l c_l = \sum_{p=1}^{P} B_p(q_p) + \sum_{l=1}^{L} \lambda_l c_l$$

Observation: sources can deduce optimal rate from $q_p = \sum_{l \in \mathcal{L}(p)} \lambda_l$

Equilibrium rates and utilities

Sources maximize utilities minus "resource cost"

$$\max_{s_p \ge 0} u_p(s_p) - s_p \sum_{l \in \mathcal{L}(p)} \lambda_l = \max_{s_p \ge 0} u_p(s_p) - q_p s_p$$

Equilibrium rates from first-order conditions

$$u_p'(s_p^\star) - q_p = 0 \quad \Rightarrow s_p^\star = \max \left\{ 0, \left(u_p' \right)^{-1} (q_p) \right\}$$

Can also be used to deduce utility function from equilibrium $(s_p^\star, q_p^\star)$

Solving the dual problem

Apply (sub-)gradient method to compute optimal multipliers

$$\lambda_l^{(k+1)} = \left[\lambda_l^{(k)} + \alpha^{(k)} \Big(\sum_{p \in \mathcal{P}(l)} s_p - c_l \Big) \right]^+$$

Observation: links update multipliers based on local excess capacity

Common to use constant step-length

A synchronous gradient projection

Link algorithm (carried out at each link at each time k=1, ..., K)

1. Compute the requested aggregate data rate $\sum_{p \in \mathcal{P}(l)} s_p$
2. Update its price $\lambda_l^{(k+1)} = [\lambda_l^{(k)} + \alpha^{(k)} (\sum_{p \in \mathcal{P}(l)} s_p - c_l)]^+$
3. Communicate new price to all sources p that use link l

Source algorithm (executed at each source)

1. Receive sum of link prices along its path $q_p = \sum_{l \in \mathcal{L}(p)} \lambda_l$
2. Choose new transmission rate $s_p^{(t+1)} = s^\star(q_l^{(t)})$
3. Communicate new rate to all links in path

Primal and dual decomposition

Consider the problem

$$\begin{array}{ll} \text{minimize} & f_x(x) + f_y(y) \\ \text{subject to} & g_x(x) \le r_x \quad g_y(y) \le r_y \\ & r_x + r_y \le r_{\text{tot}} \end{array}$$

Dual decomposition: relax resource constraint, solve subproblems

$$\begin{array}{ll} \text{minimize} & f_x(x) + \lambda r_x \\ \text{subject to} & g_x(x) \le r_x \end{array} \qquad \begin{array}{ll} \text{minimize} & f_y(y) + \lambda r_y \\ \text{subject to} & g_y(y) \le r_y \end{array}$$

Coordinate by subgradient method (that finds optimal resource price)

The primal approach

Re-write problem as

$$\text{minimize} \quad p_x(r) + p_y(r)$$

where

$$p_x(r_x) = \inf \{ f_x(x) \mid g_x(x) \le r \}$$
$$p_y(r_x) = \inf \{ f_y(y) \mid g_y(y) \le r_{\text{tot}} - r \}$$

Recall: $p_x(r_x)$ is convex, a subgradient is given by $-\lambda_x$

Coordinator updates resource allocation (rather than prices)

- all iterates are feasible

Primal decomposition

A perfect research paper presentation topic!

Reference: "Notes on Decomposition Methods", Boyd, Xiao and Mutapcic

Summary

- subgradients and the subgradient method
- solving the primal via the dual
- dual decomposition for efficient optimization
- distributed optimization example: duality model of TCP