TMA521/MMA511
Large-Scale Optimization
Lecture 6
Lagrangian duality, part IV: Primal recovery in the non-strictly convex case
Extension to the case of mixed binary linear optimization: Core problems

Ann-Brith Strömberg

2018-01-26

A convex optimization problem[1]

- Assumptions
  - The functions $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $g_i : \mathbb{R}^n \mapsto \mathbb{R}$, $i \in \mathcal{I} := \{1, \dots, m\}$, are convex and (possibly) nonsmooth (i.e., nondifferentiable)
  - $X \subset \mathbb{R}^n$ is convex and compact
  - The set X is simple $\iff$ easily solved subproblems
  - Nonempty feasible set: $\{\, \mathbf{x} \in X \,|\, g_i(\mathbf{x}) \leq 0,\ i \in \mathcal{I} \,\} \neq \emptyset$

$\Rightarrow$ A convex optimization problem

$$\begin{aligned} f^* := \min \ & f(\mathbf{x}), && \text{(1a)} \\ \text{subject to} \ & g_i(\mathbf{x}) \leq 0, \quad i \in \mathcal{I}, && \text{(1b)} \\ & \mathbf{x} \in X && \text{(1c)} \end{aligned}$$

- Solution set: $X^* := \operatorname{argmin} \{\, f(\mathbf{x}) \,|\, g_i(\mathbf{x}) \leq 0,\ i \in \mathcal{I};\ \mathbf{x} \in X \,\}$

[1] Larsson, Patriksson, Strömberg (1999): *Ergodic, primal convergence in dual subgradient schemes for convex programming*, Math. Program. 86:283–312

Lagrangian dual problem

- Relax the constraints (1b) $\Rightarrow$ Lagrange function:

$$\mathcal{L}(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x}), \qquad (\mathbf{x}, \mathbf{u}) \in \mathbb{R}^n \times \mathbb{R}^m$$

- $\mathbf{g}(\mathbf{x}) = [g_i(\mathbf{x})]_{i \in \mathcal{I}}$, $\mathbf{x} \in \mathbb{R}^n$; $\mathbf{u} = [u_i]_{i \in \mathcal{I}}$
- For any $\mathbf{u} \in \mathbb{R}^m_+$, $\mathcal{L}(\cdot, \mathbf{u})$ is convex on $\mathbb{R}^n$.
- Concave dual objective function:

$$\theta(\mathbf{u}) := \min_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x}) \right\}, \qquad \mathbf{u} \in \mathbb{R}^m \tag{2}$$

- Solution set to the subproblem at $\mathbf{u} \in \mathbb{R}^m$:

$$X(\mathbf{u}) := \left\{ \mathbf{x} \in X \,\middle|\, f(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x}) \leq \theta(\mathbf{u}) \right\} \neq \emptyset \tag{3}$$

- The Lagrange dual to the program (1):

$$\theta^* := \max_{\mathbf{u} \geq \mathbf{0}} \quad \theta(\mathbf{u}) \tag{4}$$

- Solution set to the dual: $U^* := \operatorname{argmax} \left\{ \theta(\mathbf{u}) \,\middle|\, \mathbf{u} \geq \mathbf{0} \right\}$

Subdifferential of the Lagrangian dual function

Weak duality

$\theta(\mathbf{u}) \leq f(\mathbf{x})$ holds whenever $\mathbf{u} \geq \mathbf{0}$, $\mathbf{x} \in X$, and $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$

- The *subdifferential* of the concave function θ at $\mathbf{u} \in \mathbb{R}^m$:

  $$\partial\theta(\mathbf{u}) = \left\{ \gamma \in \mathbb{R}^m \,\middle|\, \theta(\mathbf{v}) \leq \theta(\mathbf{u}) + \gamma^T(\mathbf{v} - \mathbf{u}), \quad \mathbf{v} \in \mathbb{R}^m \right\}$$

  the elements of which are called *subgradients*

Proposition: Subdifferential of the dual objective function

For each $\mathbf{u} \in \mathbb{R}^m$,

$$\partial\theta(\mathbf{u}) = \{ \mathbf{g}(\mathbf{x}) \,|\, \mathbf{x} \in X(\mathbf{u}) \}$$

θ is differentiable at $\mathbf{u} \iff$ each g_i is constant on $X(\mathbf{u})$, in which case $\nabla\theta(\mathbf{u}) = \mathbf{g}(\mathbf{x})$ for any $\mathbf{x} \in X(\mathbf{u})$. [a]

[a] Proposition 1 in Larsson, Patriksson, Strömberg (1999)

Optimality conditions

- The *normal cone* to the set $\mathbb{R}^m_+$ at $\mathbf{u} \in \mathbb{R}^m_+$:

$$N_{\mathbb{R}^m_+}(\mathbf{u}) = \left\{ \boldsymbol{\nu} \in \mathbb{R}^m_- \,\middle|\, \nu_i u_i = 0,\; i \in \mathcal{I} \right\}$$

Proposition: Optimality conditions for the Lagrangian dual

$$\mathbf{u} \in U^* \iff \exists\, \gamma \in \partial\theta(\mathbf{u}) \text{ such that } \gamma \leq \mathbf{0} \text{ and } \mathbf{u}^T\gamma = 0$$

Equivalently: $\partial\theta(\mathbf{u}) \cap N_{\mathbb{R}^m_+}(\mathbf{u}) \neq \emptyset$

Assumption: Slater constraint qualification

$$\{\mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) < \mathbf{0}\} \neq \emptyset \tag{5}$$

Proposition: Primal–dual optimality conditions

Suppose that Assumption (5) holds and let $\mathbf{u} \in U^*$.

$$\mathbf{x} \in X^* \iff \mathbf{x} \in X(\mathbf{u}),\ \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \text{ and } \mathbf{u}^T\mathbf{g}(\mathbf{x}) = 0.$$ [a]

[a] Proposition 3 in Larsson, Patriksson, Strömberg (1999)

Optimal primal set

- Under Assumption (5), the solution set to the primal program (1) may be expressed as

$$X^* = \left\{ \mathbf{x} \in X(\mathbf{u}) \;\middle|\; \mathbf{g}(\mathbf{x}) \leq \mathbf{0},\ \mathbf{u}^T \mathbf{g}(\mathbf{x}) = 0 \right\}, \tag{6}$$

  irrespective of the choice $\mathbf{u} \in U^*$
- The primal–dual optimality conditions may be expressed as

$$\mathbf{x} \in X^* \text{ and } \mathbf{u} \in U^* \quad \Longleftrightarrow \quad \mathbf{g}(\mathbf{x}) \in \partial\theta(\mathbf{u}) \cap N_{\mathbb{R}^m_+}(\mathbf{u}) \tag{7}$$

- At a dual solution $\mathbf{u} \in U^*$, the subproblem solution set $X(\mathbf{u})$ is typically *not* a singleton

⇒ The dual objective function is nonsmooth (nondifferentiable) on the optimal set U^*

⇒ A subgradient that can be used to verify the optimality of such a solution is not directly available.

Subgradient optimization applied to the Lagrange dual (4)

- Choose a starting solution $\mathbf{u}^0 \geq \mathbf{0}$ and compute iterates $\mathbf{u}^t$ according to the formula[2]

$$\mathbf{u}^{t+\frac{1}{2}} = \mathbf{u}^t + \alpha_t \mathbf{g}(\mathbf{x}^t), \quad \mathbf{u}^{t+1} = \left[\mathbf{u}^{t+\frac{1}{2}}\right]_+, \quad t = 0, 1, \ldots \quad (8)$$

- $\mathbf{x}^t \in X(\mathbf{u}^t)$; a solution to the subproblem (3) at $\mathbf{u}^t$

$\Rightarrow$ $\mathbf{g}(\mathbf{x}^t) \in \partial\theta(\mathbf{u}^t)$; a subgradient of θ at $\mathbf{u}^t$

- α_t is the step length chosen at iteration t
- $[\cdot]_+ := ([\cdot]_+)_{i \in \mathcal{I}}$ denotes the Euclidean projection onto $\mathbb{R}^m_+$ (i.e., the component-wise projection onto $\mathbb{R}_+$)

[2] The method (9) in Larsson, Patriksson, Strömberg (1999)

Convergence for divergent series steplengths

Proposition

Apply the method (8) to the program (4), with the step lengths α_t fulfilling the *divergent series conditions*

$$\alpha_t > 0,\ \forall t, \quad \{\alpha_t\} \rightarrow 0, \quad \lim_{t\to\infty} \sum_{s=0}^{t-1} \alpha_s = \infty, \tag{9a}$$

and

$$\lim_{t\to\infty} \sum_{s=0}^{t-1} \alpha_s^2 < \infty. \tag{9b}$$

Then, $\{\mathbf{u}^t\} \rightarrow \mathbf{u}^\infty \in U^*$ and $\{\theta(\mathbf{u}^t)\} \rightarrow \theta^*$. [a]

[a] Proposition 4 in Larsson, Patriksson, Strömberg (1999)

Ergodic sequence of subproblem solutions

- A sequence $\{A_t\}$ of cumulative step lengths:

$$A_t = \sum_{s=0}^{t-1} \alpha_s, \qquad t = 1, 2, \dots . \tag{10}$$

- Ergodic sequence $\{\overline{\mathbf{x}}^t\}$ of subproblem solutions, $\mathbf{x}^s$, computed by the method (8)–(9) applied to (4), is defined as the weighted average

$$\overline{\mathbf{x}}^t = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \mathbf{x}^s \tag{11}$$

- Each $\overline{\mathbf{x}}^t$ is a *convex combination* of the subproblem solutions found up to iteration t

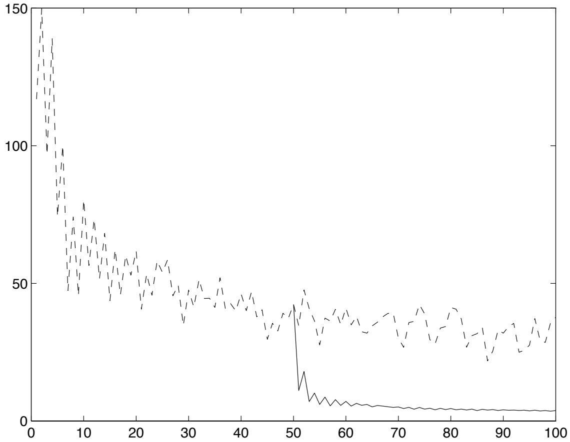
Theorem: $\overline{\mathbf{x}}^t$ converges to the solution set

Suppose that Assumption (5) holds. Apply the method (8)–(9) to the program (4). Then, $\{\mathtt{dist}\,(\overline{\mathbf{x}}^t, X^*)\} \to 0$. [a]

[a] Theorem 1 in Larsson, Patriksson, Strömberg (1999)

Convergence to the solution set

- A measure of the distances from $\bar{\mathbf{x}}^t$ (solid line) and $\mathbf{x}^t$ (dashed line) to the set X^*
- Iterations t on the horizontal axis



Improved convergence results[3]

- An ergodic primal sequence defined as

$$\overline{\mathbf{x}}^t := \sum_{s=0}^{t-1} \mu_s^t \mathbf{x}^s, \quad t = 1, 2, \ldots \tag{12a}$$

where

$$\sum_{s=0}^{t-1} \mu_s^t = 1, \quad \mu_s^t \geq 0, \ s = 0, \ldots, t-1 \tag{12b}$$

Definition

$$\gamma_s^t := \mu_s^t / \alpha_s, \quad s = 0, \ldots, t-1, \ t = 1, 2, \ldots$$

and

$$\Delta\gamma_{\max}^t := \max_{s \in \{1,\ldots,t-1\}} \{\gamma_s^t - \gamma_{s-1}^t\}$$

[3] Gustavsson, Patriksson, Strömberg (2015): *Primal convergence from dual subgradient methods for convex optimization*, Math. Program. 150(2):365–390

Assumptions on the relations between the steplengths and the convexity weights

A1: $\gamma_s^t \geq \gamma_{s-1}^t$, $s = 1, \dots, t-1$, $t = 2, 3, \dots$

A2: $\Delta\gamma_{\max}^t \to 0$ as $t \to \infty$, and

A3: $\gamma_0^t \to 0$ as $t \to \infty$ and, for some $\Gamma > 0$, $\gamma_{t-1}^t \leq \Gamma$ for all t

Theorem: optimality of $\overline{\mathbf{x}}^t$ in the limit

Apply the method (8) to the program (4) with step lengths α_t such that $\{\mathbf{u}^t\} \to \mathbf{u}^\infty$, and generate the sequence $\{\overline{\mathbf{x}}^t\}$ as in (12). Under the assumption (5), if the steplengths α_t and the convexity weights μ_s^t fulfill A1–A3, then it holds that [a]

$$\mathbf{u}^\infty \in U^* \quad \text{and} \quad \text{dist}\,(\overline{\mathbf{x}}^t, X^*) \to 0$$

[a] Theorem 1 in Gustavsson, Patriksson, Strömberg (2015)

Special cases

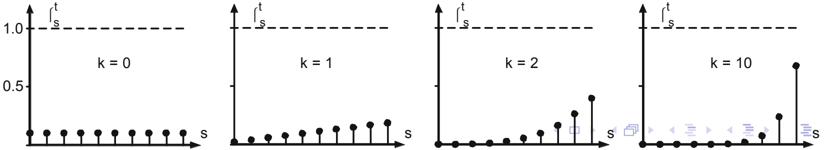
Definition: The s^k-rule

Let $k \geq 0$. The s^k-rule creates an ergodic sequence by choosing convexity weights according to [a]

$$\mu_s^t = \frac{(s+1)^k}{\sum_{\ell=0}^{t-1}(\ell+1)^k}, \quad s = 0, \ldots, t-1, \; t = 1, 2, \ldots$$

[a] Definition 1 in Gustavsson, Patriksson, Strömberg (2015)

- The effect (for $k > 0$) is that later subproblem solutions receive larger weights $\Rightarrow$ faster convergence to the primal solution set



Performance of the different rules

τ on the horizontal axis

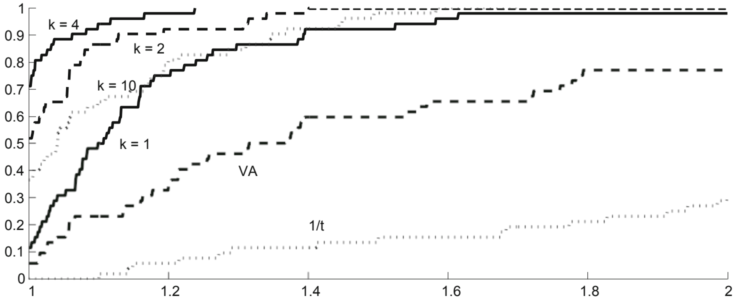


Fig. 2 Performance profiles of the methods for the 56 test instances (28 instances, each with the BPR congestion function (21) and the Kleinrock delay function (22), respectively). The graphs illustrate the proportion of the instances which each of the methods solved within τ times the number of iterations required by the method which solved each corresponding instance within the least number of iterations

Extension to mixed binary linear optimization problems

- A methodology for utilizing the ergodic sequences of subproblem solutions in the context of mixed binary linear optimization is described in the master's thesis:

  Recovery of primal solutions from dual subgradient methods for mixed binary linear programming; a branch-and-bound approach, by Pauline Aldenvik and Mirjam Schierscher, University of Gothenburg (2015)
  `https://gupea.ub.gu.se/handle/2077/40726`

Extension to mixed binary linear optimization problems

- Assume that the functions f and g_i, $i \in \mathcal{I}$, are linear and that the set $X := \{ \mathbf{x} \in \mathbb{B}^n \mid \mathbf{Dx} \geq \mathbf{d} \}$ is (mixed) binary
- The problem (1) can then be expressed as

$$
\begin{aligned}
z^* := \qquad \min \quad & \mathbf{c}^\top \mathbf{x}, && \text{(13a)}\\
\text{subject to} \quad \mathbf{Ax} & \geq \mathbf{b}, && \text{(13b)}\\
\mathbf{x} & \in X && \text{(13c)}
\end{aligned}
$$

- When relaxing the constraints (13b) the Lagrangian dual function is defined as

$$
q(\mathbf{u}) := \mathbf{b}^\top \mathbf{u} + \min_{\mathbf{x} \in X} \left\{ \left(\mathbf{c} - \mathbf{A}^\top \mathbf{u} \right)^\top \mathbf{x} \right\} \tag{14}
$$

- The Lagrangian dual problem:

$$
q^* := \max_{\mathbf{u} \geq \mathbf{0}} q(\mathbf{u}) \tag{15}
$$

Extension to mixed binary linear optimization problems

- Weak duality: $q^* \leq z^*$
- Strong duality does *not* hold in general, i.e., $q^* < z^*$
- The *convexified version* of (13) is defined as

$$
\begin{aligned}
z^*_{\mathrm{conv}} := \quad \min \quad & \mathbf{c}^\top \mathbf{x}, && \text{(16a)}\\
\text{subject to} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}, && \text{(16b)}\\
& \mathbf{x} \in \mathrm{conv}\, X && \text{(16c)}
\end{aligned}
$$

- The set of optimal solutions to (16): X^*_{conv}
- Strong duality: $q^* = z^*_{\mathrm{conv}}$

Convergence in the convexified version

The ergodic sequence $\{\bar{\mathbf{x}}^t\}$ of subproblem solutions converges to the optimal set of the convexified version (16): $\{\bar{\mathbf{x}}^t\} \to X^*_{\mathrm{conv}}$ [a]

[a] Theorem 3 in Aldenvik, Schierscher (2015)

A partition of the variables from an optimal binary solution

- Let $\mathcal{J}_0 \subset \{1, \ldots, n\}$ and $\mathcal{J}_1 \subset \{1, \ldots, n\}$ denote the subsets of the variables x_j, $j \in \{1, \ldots, n\}$, which possess the value 0 and 1, respectively, in *every* optimal solution to the convexified version (16).
- Define $\mathcal{J}_{\mathrm{f}} := \{1, \ldots, n\} \setminus (\mathcal{J}_0 \cup \mathcal{J}_1)$, corresponding to the variables x_j, $j \in \{1, \ldots, n\}$, having a fractional optimal value in some optimal solution to (16).
- For each $j \in \{1, \ldots, n\}$, the following relations hold:

$$j \in \mathcal{J}_0 \Longrightarrow \{\bar{x}_j^t\} \to 0$$
$$j \in \mathcal{J}_1 \Longrightarrow \{\bar{x}_j^t\} \to 1$$

  Consequently, if $\{\bar{x}_j^t\}$ has an accumulation point in $(0, 1)$, then $j \in \mathcal{J}_{\mathrm{f}}$.
- We let the set $\mathcal{J}_{\mathrm{f}}$ define the *core* of the problems (16) and (13) by fixing the rest of the variables to either 0 or 1

Solving (13) approximately using core problems

1. Choose values for the parameters $\tau \in \mathbb{Z}_+$ and $\varepsilon^1, \varepsilon^2 \in (0, \frac{1}{2})$, and let $t := 0$ and $\mathbf{u}^0 \in \Re^m_+$
2. Apply the subgradient method (8) to the Lagrangian dual problem (15) until $t = \tau$
3. If $\bar{x}_j^\tau \geq 1 - \varepsilon^1$, then fix the value $x_j \equiv 1$; If $\bar{x}_j^\tau \leq \varepsilon^1$, then fix the value $x_j \equiv 0$, $j = 1, \ldots, n$
4. If the core problem, defined by the non-fixed variables, is feasible, then solve it, either exactly or approximately. If it is not feasible, then decrease the values of ε^1 and/or ε^2, $j = 1, \ldots, n$, and repeat from step 3
5. If the best feasible solution to (13) found is satisfactory, then terminate the algorithm
6. Update (increase) the values of ε^1 and ε^2, $j = 1, \ldots, n$ and repeat from 2 with $t := 0$ and $\mathbf{u}^0 := \mathbf{u}^\tau$

Ways to utilize the core problems to solve (13) efficiently

- In Aldenvik, Schierscher (2015) also a Lagrangian heuristic as well as a Branch-and-bound algorithm with a Lagrangian heuristic are implemented
- The technique using *core problems* can be used to find a feasible solution to the original problem.
- Read also the slides about the case of convex optimization problems with *possibly empty feasible sets*