Ergodic, primal convergence in dual subgradient schemes for convex programming, II: the case of inconsistent primal problems

Mathematical Programming A, May 2017, Vol 163, Iss 1–2, pp 57–84

doi.org/10.1007/s10107-016-1055-x

Ann-Brith Strömberg

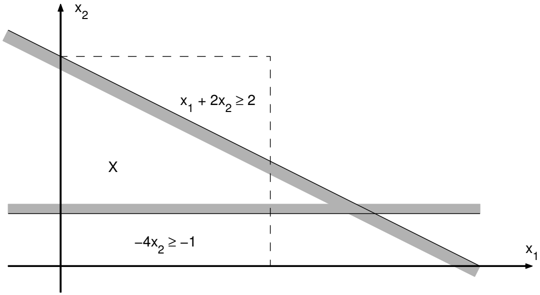
Joint work with Magnus Önnheim, Emil Gustavsson, Michael Patriksson, and Torbjörn Larsson

CHALMERS | UNIVERSITY OF GOTHENBURG

Large-Scale Optimization
2018–01–29

Introduction

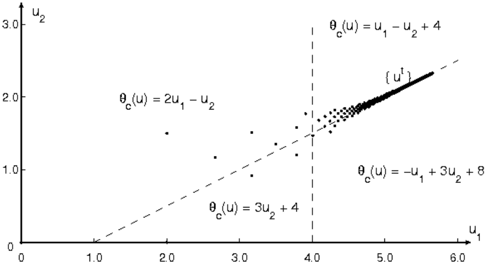
- We consider a possibly infeasible, i.e., inconsistent, convex optimization problem



Introduction

- We consider a possibly infeasible, i.e., inconsistent, convex optimization problem
- Apply a Lagrangian dual method to this problem

$\Rightarrow$ Divergence of the dual iterates



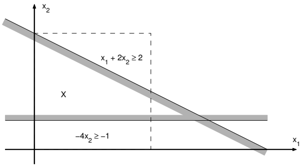
(a) The sequence $\{u^t\}$

Introduction

- We consider a possibly infeasible, i.e., inconsistent, convex optimization problem
- Apply a Lagrangian dual method to this problem

$\Rightarrow$ Divergence of the dual iterates

- Will the primal sequence of subproblem solutions still yield relevant information about the primal solution?



- Subproblem (extreme) solutions $\in \left\{ \binom{0}{0}, \binom{0}{1}, \binom{1}{0}, \binom{1}{1} \right\}$

Introduction

- We consider a possibly infeasible, i.e., inconsistent, convex optimization problem

- Apply a Lagrangian dual method to this problem

$\Rightarrow$ Divergence of the dual iterates

- Will the primal sequence of subproblem solutions still yield relevant information about the primal solution?

- Associate a primal-dual convex-concave program with a saddle point problem

- ... and homogenize the dual problem

Introduction

- Convergence results for a subgradient optimization algorithm applied to the Lagrange dual

Introduction

- Convergence results for a subgradient optimization algorithm applied to the Lagrange dual
- Construct an ergodic, i.e., averaged, sequence of primal subproblem solutions

$\Rightarrow$ Convergence to a saddle point

Introduction

- Convergence results for a subgradient optimization algorithm applied to the Lagrange dual
- Construct an ergodic, i.e., averaged, sequence of primal subproblem solutions

$\Rightarrow$ Convergence to a saddle point

- The dual iterates diverge in the direction of steepest ascent $\Longleftrightarrow$ the unique dual solution of the saddle-point problem

Introduction

- Convergence results for a subgradient optimization algorithm applied to the Lagrange dual
- Construct an ergodic, i.e., averaged, sequence of primal subproblem solutions

$\Rightarrow$ Convergence to a saddle point

- The dual iterates diverge in the direction of steepest ascent $\iff$ the unique dual solution of the saddle-point problem
- The primal sequence converges to the subset of the primal set X where the total infeasibility w.r.t. the relaxed constraints is at minimum

Introduction

- Convergence results for a subgradient optimization algorithm applied to the Lagrange dual
- Construct an ergodic, i.e., averaged, sequence of primal subproblem solutions

$\Rightarrow$ Convergence to a saddle point

- The dual iterates diverge in the direction of steepest ascent $\Longleftrightarrow$ the unique dual solution of the saddle-point problem
- The primal sequence converges to the subset of the primal set X where the total infeasibility w.r.t. the relaxed constraints is at minimum
- For LP: The original objective function is minimized over the subset of minimum infeasibility

Introduction

To start, we wish to solve a convex program:

$$\underset{x}{\text{minimize}}\ f(x), \tag{1a}$$

$$\text{subject to } g(x) \leq 0^m, \tag{1b}$$

$$x \in X \tag{1c}$$

- $\emptyset \neq X \subset \mathbb{R}^n$ – convex and compact set
- $g : \mathbb{R}^n \mapsto \mathbb{R}^m$, $f : \mathbb{R}^n \mapsto \mathbb{R}$ – convex functions
- Lagrange function w.r.t. relaxing the constraints (1b):

$$\mathcal{L}_f(x, u) := f(x) + u^\top g(x), \quad (x, u) \in \mathbb{R}^n \times \mathbb{R}^m \tag{2}$$

- Lagrangean dual function:

$$\theta_f(u) := \min_{x \in X} \mathcal{L}_f(x, u), \quad u \in \mathbb{R}^m \tag{3}$$

- Lagrangean dual problem:

$$\underset{u \in \mathbb{R}^m_+}{\text{maximize}}\ \theta_f(u) \tag{4}$$

Results for the case of consistency

- Assume $\{x \in X \mid g(x) \leq 0^m\} \neq \emptyset$

$\Rightarrow$ The optimal value of the convex program (1): $\theta_f^* > -\infty$

- Set of optimal solutions:
  $X_f^* := \{x \in X \mid g(x) \leq 0^m,\ f(x) \leq \theta_f^*\}$
- Lagrangean dual problem:

$$\theta_c^* := \max_{u \in \mathbb{R}_+^m} \ \theta_f(u) \tag{5}$$

- The dual optimal set is bounded (provided a Slater condition, i.e., $\{x \in X \mid g(x) < 0^m\} \neq \emptyset$):

$$U_f^* := \{u \in \mathbb{R}_+^m \mid \theta_f(u) \geq \theta_f^*\} \neq \emptyset$$

- Lagrangean subproblem solution at $u \in \mathbb{R}^m$:

$$x_f(u) \in X_f(u) := \operatorname*{argmin}_{x \in X} \{f(x) + u^\top g(x)\} \tag{6}$$

Dual subgradient algorithm

- Subdifferential of θ_f at $u \in \mathbb{R}^m$:

$$\partial\theta_f(u) := \left\{ \boldsymbol{\gamma} \in \mathbb{R}^m \,\middle|\, \theta_f(w) \le \theta_f(u) + \boldsymbol{\gamma}^T (w - u), \; w \in \mathbb{R}^m \right\}$$
$$= \{ g(x) \mid x \in X_f(u) \}$$

- Subgradient algorithm applied to the dual (5):

$$u^0 \ge 0; \; u^{t+1} := \left[u^t + \alpha_t g(x_f(u^t)) \right]_+ \; t = 0, 1, \ldots \qquad (7)$$

Convergence to a dual optimal set and point

Theorem

Apply the subgradient method (7) to the dual program (5). If the step lengths α_t fulfil the divergent series conditions

$$\alpha_t > 0 \ \forall t; \quad \{\alpha_t\} \to 0; \quad \left\{\sum_{s=0}^{t-1} \alpha_s\right\} \to \infty. \tag{8a}$$

Then[a] $\{\mathtt{dist}(u^t; U_f^*)\} \to 0$ *and* $\{\theta_f(u^t)\} \to \theta_f^*$
If, in addition,

$$\sum_{s=0}^{\infty} \alpha_s^2 < \infty \tag{8b}$$

holds, then[b] $\{u^t\} \to u^\infty \in U_f^*$.

---

[a] Ermol'ev, Yu.M. (1966): *Methods for solving nonlinear extremal problems*, Cybernetics 2(4), 1–14

[b] Shepilov, M.A. (1976): *Method of the generalized gradient for finding the absolute minimum of a convex function*, Cybernetics 12, 547–553

Convergence of an ergodic (averaged) sequence of primal subproblem solutions to the primal optimal set

- Cumulative step lengths:

$$A_t := \sum_{s=0}^{t-1} \alpha_s, \qquad t = 1, 2, \dots \tag{9}$$

- Averaged/ergodic primal sequence:

$$\overline{x}_f^t := \frac{1}{A_t} \sum_{s=0}^{t-1} \alpha_s x_f(u^s), \qquad t = 1, 2, \dots \tag{10}$$

Theorem

[a] *Apply the subgradient method (7), (8) to the dual program (5). Then,*

$$\{\mathtt{dist}\,(\overline{x}_f^t; X_f^*)\} \to 0 \quad \textit{and} \quad \{f(\overline{x}^t)\} \to \theta_f^*$$

---

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

The case of inconsistency

- Assume that $\{x \in X \mid g(x) \leq 0^m\} = \emptyset$
- Lagrangean dual problem:

$$\infty = \sup_{u \in \mathbb{R}^m_+} \theta_f(u) \tag{11}$$

$\Rightarrow$ The dual optimal set is empty: $U^*_f = \emptyset$

The case of inconsistency

- Assume that $\{x \in X \mid g(x) \leq 0^m\} = \emptyset$
- Lagrangean dual problem:

$$\infty = \sup_{u \in \mathbb{R}^m_+} \theta_f(u) \tag{11}$$

$\Rightarrow$ The dual optimal set is empty: $U^*_f = \emptyset$

- A cone of feasible and ascent directions for the dual (11)
  $C := \{w \in \mathbb{R}^m_+ \mid w^\top g(x) > 0, \; x \in X\}$

Theorem

A theorem of the alternative:

$$\{x \in X \mid g(x) \leq 0^m\} = \emptyset \iff C \neq \emptyset$$

The case of inconsistency

- Assume that $\{x \in X \mid g(x) \leq 0^m\} = \emptyset$
- Lagrangean dual problem:

$$\infty = \sup_{u \in \mathbb{R}^m_+} \theta_f(u) \tag{11}$$

$\Rightarrow$ The dual optimal set is empty: $\quad U^*_f = \emptyset$

- A cone of feasible and ascent directions for the dual (11)
  $C := \{w \in \mathbb{R}^m_+ \mid w^\top g(x) > 0, \; x \in X\}$

Theorem

A theorem of the alternative:

$$\{x \in X \mid g(x) \leq 0^m\} = \emptyset \quad \Longleftrightarrow \quad C \neq \emptyset$$

- Note: The cone C is independent of the objective function f

Divergence in the case of inconsistency

Theorem

Assume that $\{x \in X \mid g(x) \leq 0^m\} = \emptyset$.
Let the sequence $\{u^t\}$ *be generated by*

$$u^0 \geq 0; \quad u^{t+1} = \left[u^t + \alpha_t g(x_f(u^t))\right]_+$$

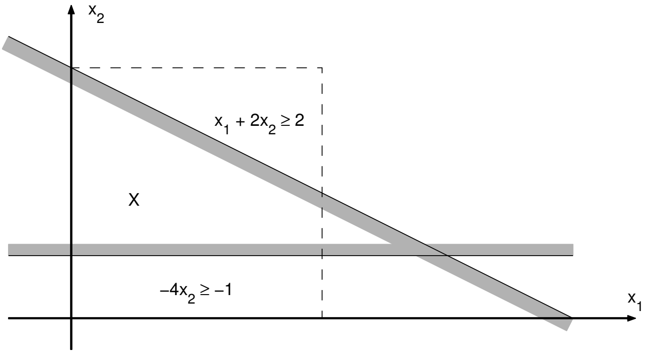
with $\alpha_t > 0$, $\{\alpha_t\} \to 0$, *and* $\left\{\sum_{s=0}^{t-1} \alpha_s\right\} \to \infty$, *applied to the dual program*

$$\sup_{u \in \mathbb{R}^m_+} \theta_f(u).$$

Then $\quad \{\|u^t\|\} \to \infty \quad$ *as* $\quad t \to \infty$

A numerical illustration — the primal space

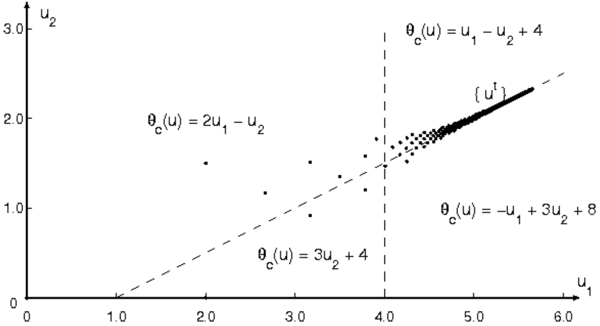
Let $m = n = 2$, $f(x) = 4x_1 + 2x_2$, $g(x) = \begin{pmatrix} 2 & -x_1 & -2x_2 \\ -1 & & +4x_2 \end{pmatrix}$, and $X = [0, 1]^2$.



A numerical illustration — the dual space

$$\theta_f(u) = \begin{cases} 2u_1 - u_2, & u_1 \leq 4,\ u_1 - 2u_2 \leq 1, \\ 3u_2 + 2, & u_1 \leq 4,\ u_1 - 2u_2 \geq 1, \\ u_1 - u_2 + 4, & u_1 \geq 4,\ u_1 - 2u_2 \leq 1, \\ -u_1 + 3u_2 + 6, & u_1 \geq 4,\ u_1 - 2u_2 \geq 1, \end{cases}$$

A couple of numbers in the figure are incorrect ...



(a) The sequence $\{u^t\}$

The homogeneous dual

- Assume $X \neq \emptyset$ but $\{x \in X \mid g(x) \leq 0^m\} = \emptyset$

The homogeneous dual

- Assume $X \neq \emptyset$ but $\{x \in X \mid g(x) \leq 0^m\} = \emptyset$
- We will study the case when $f(x) = 0$ and restrict the dual space to the unit ball

The homogeneous dual

- Assume $X \neq \emptyset$ but $\{x \in X \mid g(x) \leq 0^m\} = \emptyset$
- We will study the case when $f(x) = 0$ and restrict the dual space to the unit ball
- The Lagrange function at (x, u) for $u \neq 0$

$$\mathcal{L}_f(x, u) = \|u\| \left(\frac{f(x)}{\|u\|} + \frac{u^T g(x)}{\|u\|} \right) \tag{12}$$

- As the value of $\|u\|$ increases, the term $f(x)$ in the computation of $x_f(u)$ will tend to be negligible
- For $t \gg 1$ the subgradient method (7), (8a) tackles an approximation of the homogeneous dual – the Lagrange dual of a pure feasibility problem:

$$\underset{u \in \mathbb{R}^m_+}{\text{maximize}} \;\; \theta_0(u) \tag{13}$$

The homogeneous dual

$$\theta_0(u) = \min_{x \in X} \left\{ u^\top g(x) \right\}, \quad u \in \mathbb{R}^m \tag{14}$$

- The function θ_0 is superlinear – its hypograph is a nonempty and convex cone in $\mathbb{R}^{m+1}$

$\Rightarrow$ $\theta_0(\delta u) = \delta \theta_0(u)$ holds for all $\delta \geq 0$ and all $u \in \mathbb{R}^m$

- Directional derivative $\theta_0'(u; d)$, of θ_0 at u in direction d:

$$\theta_0'(0; d) = \theta_0(d) \;\; \text{holds for all} \;\; d \in \mathbb{R}^m \tag{15}$$

- The homogeneous dual (13) is interpreted as searching for a steepest feasible ascent direction of θ_0

$\Rightarrow$ Maximize θ_0 over the set $V = \left\{ u \in \mathbb{R}^m_+ \;\middle|\; \|u\| \leq 1 \right\}$

$$\theta_0^{V*} := \max_{u \in V} \theta_0(u) = \max_{d \in V} \theta_0'(0; d) \tag{16}$$

An associated saddle point problem

- By definition of θ_0 and (16):

$$\theta_0^{V*} = \max_{u \in V} \left\{ \min_{x \in X} \left\{ u^\top g(x) \right\} \right\} = \min_{x \in X} \left\{ \max_{u \in V} \left\{ g(x)^\top u \right\} \right\} \quad (17)$$

since $u^T g(x)$ is convex in x (for $u \in \mathbb{R}^m_+$) and linear in u, and since the sets X and V are convex and compact

- Define the mappings $X_0(\cdot) : V \mapsto 2^X$ and $V(\cdot) : X \mapsto 2^V$ by

$$X_0(v) := \operatorname*{argmin}_{x \in X} \left\{ v^\top g(x) \right\}, \quad v \in V$$

$$V(x) := \operatorname*{argmax}_{v \in V} \left\{ g(x)^\top v \right\}, \quad x \in X$$

- Saddle point property:

$$\bar{x} \in X_0(\bar{v}) \text{ and } \bar{v} \in V(\bar{x})$$

$$\Longleftrightarrow$$

$$(\bar{x}, \bar{v}) \text{ is a saddle point for } v^\top g(x) \text{ on } X \times V$$

Characterization of the set of saddle points

- LEMMA
  $x \in X \Rightarrow V(x) = \left\{ \frac{[g(x)]_+}{\|[g(x)]_+\|} \right\}$

- The set $X_0^* \times V^*$ of saddle points is characterized as:

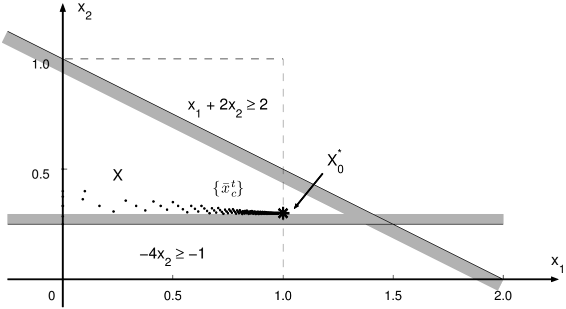
$$X_0(V^*) = X_0^* := \underset{x \in X}{\operatorname{argmin}} \|[g(x)]_+\| \quad \text{“minimum total infeasibility”}$$

$$V(x^*) = V^* := \underset{v \in V}{\operatorname{argmax}}\ \theta_0(v) \quad \text{“steepest ascent direction”}$$

- LEMMA
  $X_0^* \neq \emptyset$ is convex and compact
  $V^* = V(x^*)$ (a singleton) irrespective of $x^* \in X_0^*$

  Note: $V^* \in C$ (the cone of feasible ascent directions for θ_f)

Numerical illustrations — the primal space



Figur: $m = n = 2$, $f(x) = 4x_1 + 2x_2$, $g(x) = \begin{pmatrix} 2 & -x_1 & -2x_2 \\ -1 & & +4x_2 \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,
$X = [0,1]^2$, $X_0^* = \left(1, \frac{3}{10}\right)$

Dual and primal convergence in the inconsistent case — convergence in the homogeneous dual

Theorem

Let the sequence $\{u^t\}$ *be generated by the subgradient method*

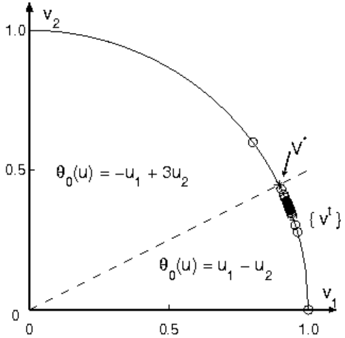
$$u^0 \geq 0; \quad u^{t+1} := \left[u^t + \alpha_t g(x_f(u^t))\right]_+$$

with $\alpha_t > 0$, $\{\alpha_t\} \to 0$, *and* $\left\{\sum_{s=0}^{t-1} \alpha_s\right\} \to \infty$. *Then*
$\{v^t\} := \left\{\frac{u^t}{\max_{s \leq t}\{1, \|u^s\|\}}\right\} \to V^*$ *and* $\{\theta_0(v^t)\} \to \theta_0^{V*}$

A numerical illustration — the homogeneous dual

$$\theta_0(u) = \begin{cases} u_1 - u_2, & u_1 \le 2u_2, \\ -u_1 + 3u_2, & u_1 \ge 2u_2, \end{cases} \qquad \theta_0^{V*} = \frac{1}{\sqrt{5}}$$

$$\{v^t\} = \left\{ \frac{u^t}{\max_{s \le t}\{1, \|u^s\|\}} \right\} \to V^* = \frac{1}{\sqrt{5}} \binom{2}{1}$$

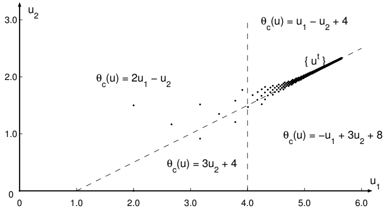


(b) The sequence $\{ v^t \}$

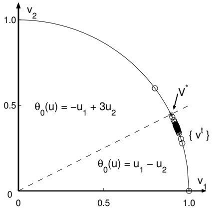
A numerical illustration — the homogeneous dual

$$\theta_f(u) = \begin{cases} 2u_1 - u_2, & u_1 \le 4,\ u_1 - 2u_2 \le 1, \\ 3u_2 + 2, & u_1 \le 4,\ u_1 - 2u_2 \ge 1, \\ u_1 - u_2 + 4, & u_1 \ge 4,\ u_1 - 2u_2 \le 1, \\ -u_1 + 3u_2 + 6, & u_1 \ge 4,\ u_1 - 2u_2 \ge 1, \end{cases}$$

$$\theta_0(u) = \begin{cases} u_1 - u_2, & u_1 \le 2u_2, \\ -u_1 + 3u_2, & u_1 \ge 2u_2, \end{cases}$$



(a) The sequence $\{u^t\}$



(b) The sequence $\{v^t\}$

$\{u^t\}$ diverges in the direction V^* of steepest ascent of θ_f over $\mathbb{R}^m_+$

Convergence to a saddle point

Theorem (Convergence of the primal ergodic sequence)

$\{\mathtt{dist}(\overline{x}_f^t; X_0(V^*))\} \to 0$ *as* $t \to \infty$ *where* $\overline{x}_f^t = \frac{1}{A_t}\sum_{s=0}^{t-1} \alpha_s x_f(u^s)$

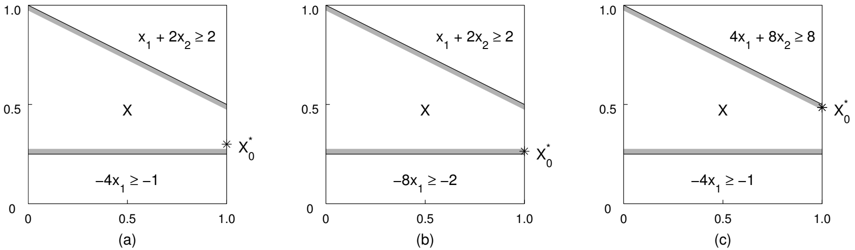
Theorem (Convergence of the homogeneous dual sequence)

$\{\mathtt{dist}\ (v^t; V(\overline{x}_f^\infty))\} \to 0$ *as* $t \to \infty$ *where* $\overline{x}_f^\infty$ *is an accumulation point of* $\{\overline{x}_f^t\}$

Theorem (Convergence to a saddle point)

$\left\{\mathtt{dist}\left((\overline{x}_f^t, v^t); X_0(V^*) \times V(\overline{x}_f^\infty)\right)\right\} \to 0$ *as* $t \to \infty$ *where* $\overline{x}_f^\infty \in X_0(V^*) = X_0^*$ *and* $V(\overline{x}_f^\infty) = V^* = \frac{[g(\overline{x}_f^\infty)]_+}{\left\|[g(\overline{x}_f^\infty)]_+\right\|}$

Further results: The effect of constraint scaling on the solution set X_0^*



Main references and further reading

- Larsson, Patriksson, Strömberg (1996): *Conditional subgradient optimization—theory and applications*, European Journal of Operational Research 88:382–403
- Larsson, Patriksson, Strömberg (1999): *Ergodic, primal convergence in dual subgradient schemes for convex programming*, Mathematical Programming 86:283–312
- Larsson, Patriksson, Strömberg (2003): *On the convergence of conditional ε-subgradient methods for convex programs and convex–concave saddle-point problems*, European Journal of Operational Research, 151:461–473
- Gustavsson, E., Patriksson, M., Strömberg, A.-B. (2015): *Primal convergence from dual subgradient methods for convex optimization*. Mathematical Programming 150(2), 365–390 (increase convergence speed)