Revisiting Augmented Lagrangian Duals

Marcelo Cordova(cordova.m@posgrad.ufsc.br)
Welington de Oliveira(welington.oliveira@mines-paristech.fr)
Claudia Sagastizábal(sagastiz@unicamp.br)

eprints for the optimization community

CeMEAI
CEPID - Center for Mathematical Sciences Applied to Industry

Claudia Sagastizábal
IMECC-UNICAMP Brazil

One World Optimization Seminar

UniVie, June 22nd, 2020

Context: when to use decomposition methods?

- problems too difficult to solve directly
- problems with partial separable structure
- problems of problems
- information not accesible

Context: when to use decomposition methods?

- problems too difficult to solve directly

  mixed 0-1, linear, quadratic, or nonlinear programs

Context: when to use decomposition methods?

- problems too difficult to solve directly

  mixed 0-1, linear, quadratic, or nonlinear programs

- problems with partial separable structure

  complicating variables: block diagonal 2nd-stage constraints, coupled with 1st stage

Context: when to use decomposition methods?

- problems too difficult to solve directly

  mixed 0-1, linear, quadratic, or nonlinear programs

- problems with partial separable structure

  complicating variables: block diagonal 2nd-stage constraints, coupled with 1st stage

  complicating constraints, separable objective

Context: when to use decomposition methods?

- problems too difficult to solve directly

  mixed 0-1, linear, quadratic, or nonlinear programs

- problems with partial separable structure

  complicating variables: block diagonal 2nd-stage constraints, coupled with 1st stage

  complicating constraints, separable objective

- problems of problems

  equilibrium problems, games, variational inequalities

Context: when to use decomposition methods?

- problems too difficult to solve directly
  mixed 0-1, linear, quadratic, or nonlinear programs

- problems with partial separable structure
  complicating variables: block diagonal 2nd-stage constraints,
  coupled with 1st stage
  complicating constraints, separable objective
- problems of problems
  equilibrium problems, games, variational inequalities

- information not accesible **in ML, commercial oracles**

Which decomposition method?

Which decomposition method?

Decomposition: what and how?

Primal

Decomposition: what and how?

Primal

Dual

Decomposition: what and how?

Primal

Dual

Primal-Dual

Decomposition: what and how?

Primal

Dual

Primal-Dual

It all depends on the output of interest

Illustration with a simple example

$$\begin{cases} \min & f_T(y_T) + f_H(y_H) \\ \text{s.t.} & y_T \in \mathcal{S}_T \,,\, y_H \in \mathcal{S}_H \\ & \\ & \end{cases}$$

Two power plants

$y_T \in \mathcal{S}_T$

$f_T(y_T)$

$y_H \in \mathcal{S}_H$

$f_H(y_H)$

Illustration with a simple example

$$\begin{cases} \min & f_T(y_T) + f_H(y_H) \\ \text{s.t.} & y_T \in \mathcal{S}_T\,, y_H \in \mathcal{S}_H \end{cases}$$

Two power plants

$y_T \in \mathcal{S}_T$

$f_T(y_T)$

$y_H \in \mathcal{S}_H$

$f_H(y_H)$

$\boxed{y_T + y_H = d}$ (demand)

Illustration with a simple example

$$\begin{cases} \min & f_T(y_T) + f_H(y_H) \\ \text{s.t.} & y_T \in \mathcal{S}_T \,, y_H \in \mathcal{S}_H \\ & y_T + y_H = d \end{cases}$$

Two power plants

$y_T \in \mathcal{S}_T$

$f_T(y_T)$

$y_H \in \mathcal{S}_H$

$f_H(y_H)$

A less simple example

Two power plants

$y_T \in \mathcal{S}_T$

$\langle \mathcal{F}, x \rangle + f_T(y_T)$

$x \in \{0, 1\}$ and $y_T \leq x\, y^{up}$

$y_H \in \mathcal{S}_H$

$f_H(y_H)$

A less simple example

$$\left\{\begin{array}{ll} \min & \langle \mathcal{F}, x\rangle + f_T(y_T) + f_H(y_H) \\ \text{s.t.} & y_T \in \mathcal{S}_T\,, y_H \in \mathcal{S}_H \\ & y_T + y_H = d \\ & x \in \{0,1\} \text{ and } y_T \le x\, y^{up} \iff (x, y_T) \in \mathcal{S}_T \end{array}\right.$$

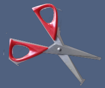
Two power plants

$y_T \in \mathcal{S}_T$
$\langle \mathcal{F}, x\rangle + f_T(y_T)$
$x \in \{0,1\}$ and $y_T \le x\, y^{up}$

$y_H \in \mathcal{S}_H$
$f_H(y_H)$

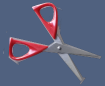
Primal scissors (Benders)

- **Good**, to split 0-1 variables from possible NLP relations



Primal scissors (Benders)

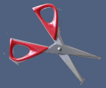
- **Good**, to split 0-1 variables from possible NLP relations
- Bad, because all technologies dealt with together



Primal scissors (Benders)

- **Good**, to split 0-1 variables from possible NLP relations
- Bad, because all technologies dealt with together

$$\text{for instance,}\left\{\begin{array}{ll} \min & \langle \mathcal{F}, x\rangle + f_T(y_T) + f_H(y_H) \\ \text{s.t.} & (x\ \ , y_T) \in \mathcal{S}_T\,, y_H \in \mathcal{S}_H \\ & y_T + y_H = d \end{array}\right.$$



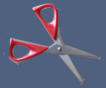
Primal scissors (Benders)

- **Good**, to split 0-1 variables from possible NLP relations
- **Bad**, because all technologies dealt with together

$$\text{for instance, } \left\{ \begin{array}{rl} \min & \langle \mathcal{F}, x \rangle + f_T(y_T) + f_H(y_H) \\ \text{s.t.} & (x_k, y_T) \in \mathcal{S}_T \, , y_H \in \mathcal{S}_H \\ & y_T + y_H = d \end{array} \right.$$

$\Longrightarrow$ operational subproblem in both y_T, y_H

for each given x_k (a master program defines x_{k+1})



Primal scissors (Benders)

- **Good**, to split 0-1 variables from possible NLP relations
- **Bad**, because all technologies dealt with together

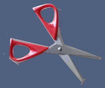
$$\text{for instance,} \begin{cases} \min & \langle \mathcal{F}, x \rangle + f_T(y_T) + f_H(y_H) \\ \text{s.t.} & (\boxed{x_k}, y_T) \in \mathcal{S}_T, y_H \in \mathcal{S}_H \\ & y_T + y_H = d \end{cases}$$

$\Longrightarrow$ operational subproblem in both y_T, y_H

for each given x_k (a master program defines x_{k+1})

To separate technologies

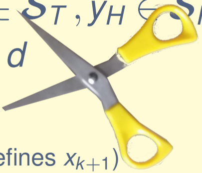
we need **dual** scissors



Primal scissors (Benders)

- **Good**, to split 0-1 variables from possible NLP relations
- **Bad**, because all technologies dealt with together

for instance, $$\left\{\begin{array}{rl} \min & \langle \mathcal{F}, x\rangle + f_T(y_T) + f_H(y_H) \\ \text{s.t.} & (\boxed{x_k}, y_T) \in \mathcal{S}_T\,, y_H \in \mathcal{S}_H \\ & y_T + y_H = d \end{array}\right.$$



$\Longrightarrow$ operational subproblem in both y_T, y_H

for each given x_k (a master program defines x_{k+1})

To separate technologies
we need **dual** scissors

acting on the demand constraint

Dual scissors: Lagrangian relaxation

$$\left\{\begin{array}{rl} \min & f_T(x, y_T) + f_H(y_H) \\ \text{s.t.} & (x\,, y_T) \in \mathcal{S}_T \\ & y_H \in \mathcal{S}_H \\ & y_T + y_H = d \qquad \leftrightarrow \mathbf{u} \end{array}\right. \Longrightarrow$$

$$\begin{aligned} L(x, y, u) \quad = \quad & f_T(x, y_T) + f_H(y_H) \\ & + \langle u, d - y_T - y_H \rangle \end{aligned}$$

Dual scissors: Lagrangian relaxation

$$\left\{\begin{array}{rl} \min & f_T(x, y_T) + f_H(y_H) \\ \text{s.t.} & (x\,, y_T) \in \mathcal{S}_T \\ & y_H \in \mathcal{S}_H \\ & y_T + y_H = d \qquad \leftrightarrow \mathbf{u} \end{array}\right. \Longrightarrow \begin{array}{rcl} L(x, y, u) & = & f_T(x, y_T) + f_H(y_H) \\ & & \quad + \langle u, d - y_T - y_H \rangle \\ & = & L_T(x, y_T, u) + L_H(y_H, u) \\ & & \quad + \langle u, d \rangle \end{array}$$

$$\Updownarrow$$

Dual scissors: Lagrangian relaxation

$$\left\{\begin{array}{rl} \min & f_T(x, y_T) + f_H(y_H) \\ \text{s.t.} & (x\,, y_T) \in \mathcal{S}_T \\ & y_H \in \mathcal{S}_H \\ & y_T + y_H = d \qquad \leftrightarrow \mathbf{u} \end{array}\right. \implies \begin{array}{rcl} L(x, y, u) & = & f_T(x, y_T) + f_H(y_H) \\ & & \quad + \langle u, d - y_T - y_H \rangle \\ & = & L_T(x, y_T, u) + L_H(y_H, u) \\ & & \quad + \langle u, d \rangle \end{array}$$

$$\Updownarrow$$

$$\left\{\begin{array}{rl} \min\limits_{x,y} \max\limits_{u} & L(x, y, u) \\ \text{s.t.} & (x\,, y_T) \in \mathcal{S}_T \\ & y_H \in \mathcal{S}_H \end{array}\right.$$

Dual scissors: Lagrangian relaxation

$$\begin{cases} \min & f_T(x, y_T) + f_H(y_H) \\ \text{s.t.} & (x\,, y_T) \in \mathcal{S}_T \\ & y_H \in \mathcal{S}_H \\ & y_T + y_H = d \qquad \leftrightarrow \mathbf{u} \end{cases} \implies \begin{aligned} L(x, y, u) &= f_T(x, y_T) + f_H(y_H) \\ &\quad + \langle u, d - y_T - y_H \rangle \\ &= L_T(x, y_T, u) + L_H(y_H, u) \\ &\quad + \langle u, d \rangle \end{aligned}$$

DUAL maxmin replaces minmax

$$\begin{cases} \max_u \min_y & L(x, y, u) \\ \text{s.t.} & (x\,, y_T) \in \mathcal{S}_T \\ & y_H \in \mathcal{S}_H \end{cases}$$

Dual scissors: Lagrangian relaxation

$$\left\{\begin{array}{rl} \min & f_T(x, y_T) + f_H(y_H) \\ \text{s.t.} & (x, y_T) \in \mathcal{S}_T \\ & y_H \in \mathcal{S}_H \\ & y_T + y_H = d \qquad \leftrightarrow \mathbf{u} \end{array}\right. \implies \begin{array}{rcl} L(x,y,u) & = & f_T(x, y_T) + f_H(y_H) \\ & & \quad + \langle u, d - y_T - y_H \rangle \\ & = & L_T(x, y_T, u) + L_H(y_H, u) \\ & & \quad + \langle u, d \rangle \end{array}$$

DUAL maxmin replaces minmax

$$\left\{\begin{array}{rl} \max\limits_u \min\limits_y & L(x,y,u) \\ \text{s.t.} & (x, y_T) \in \mathcal{S}_T \\ & y_H \in \mathcal{S}_H \end{array}\right. \implies \max_u \quad \theta_T(u) + \theta_H(u) + \langle u, d \rangle$$

Dual scissors: Lagrangian relaxation

$$\left\{\begin{array}{ll} \min & f_T(x,y_T)+f_H(y_H) \\ \text{s.t.} & (x\,,y_T)\in\mathcal{S}_T \\ & y_H\in\mathcal{S}_H \\ & y_T+y_H=d \qquad \leftrightarrow \mathbf{u} \end{array}\right. \Longrightarrow \quad \begin{array}{rcl} L(x,y,u) & = & f_T(x,y_T)+f_H(y_H) \\ & & +\langle u, d-y_T-y_H\rangle \\ & = & L_T(x,y_T,u)+L_H(y_H,u) \\ & & +\langle u,d\rangle \end{array}$$

DUAL maxmin replaces minmax

$$\left\{\begin{array}{ll} \max\limits_u \min\limits_y & L(x,y,u) \\ \text{s.t.} & (x\,,y_T)\in\mathcal{S}_T \\ & y_H\in\mathcal{S}_H \end{array}\right. \Longrightarrow \quad \begin{array}{ll} \max\limits_u & \theta_T(u)+\theta_H(u)+\langle u,d\rangle \\ & \theta_T(u):=\min L_T(x,y_T,u):(x,y_T)\in\mathcal{S}_T \\ \text{for} & \\ & \theta_H(u):=\min L_H(y_H,u):y_H\in\mathcal{S}_H \end{array}$$

Dual scissors: Lagrangian relaxation *à la bundle*

$$\left\{\begin{array}{ll} \min & f_T(y_T) + f_H(y_H) \\ \text{s.t.} & x \in \{0,1\}\,, y_T \in \mathcal{S}_T \\ & y_T \le x\, y^{up} \\ & y_H \in \mathcal{S}_H \\ & y_T + y_H = d \qquad \leftrightarrow \mathbf{u} \end{array}\right. \Longrightarrow \quad \begin{array}{rcl} L(y,u) & = & f_T(y_T) + f_H(y_H) \\ & & \quad + \langle u, d - y_T - y_H \rangle \\ & & \\ & = & L_T(y_T,u) + L_H(y_H,u) \\ & & \quad + \langle u, d \rangle \end{array}$$

DUAL maxmin replaces minmax

$$\left\{\begin{array}{ll} \max\limits_u \min\limits_y & L(x,y,u) \\ \text{s.t.} & (x\,, y_T) \in \mathcal{S}_T \\ & y_H \in \mathcal{S}_H \end{array}\right. \Longrightarrow \quad \max_u \quad \theta_T(u) + \theta_H(u) + \langle u, d \rangle$$

Dual scissors: Lagrangian relaxation *à la bundle*

$$\left\{\begin{array}{ll} \min & f_T(y_T)+f_H(y_H) \\ \text{s.t.} & x\in\{0,1\}\,, y_T\in\mathcal{S}_T \\ & y_T \le x\, y^{up} \\ & y_H \in \mathcal{S}_H \\ & y_T+y_H=d \qquad \leftrightarrow \mathbf{u} \end{array}\right. \Longrightarrow \begin{array}{rcl} L(y,u) & = & f_T(y_T)+f_H(y_H) \\ & & +\langle u, d-y_T-y_H\rangle \\ & = & L_T(y_T,u)+L_H(y_H,u) \\ & & +\langle u,d\rangle \end{array}$$

DUAL maxmin replaces minmax

$$\left\{\begin{array}{lll} \max\limits_u \min\limits_y & L(x,y,u) \\ \quad\text{s.t.} & (x,y_T)\in\mathcal{S}_T \\ & y_H\in\mathcal{S}_H \end{array}\right. \Longrightarrow \begin{array}{ll} \max\limits_u & \theta_T(u)+\theta_H(u)+\langle u,d\rangle \\ & \boldsymbol{\theta_T(u)} \approx \min L_T(x,y_T,u) : (x,y_T)\in\mathcal{S}_T \\ \text{for} & \\ & \boldsymbol{\theta_H(u)} \approx \min L_H(y_H,u) : y_H\in\mathcal{S}_H \end{array}$$

Dual scissors

- **Good**, to separate technologies + inexact information

Dual scissors

- **Good**, to separate technologies + inexact information
- **Good**, if output of interest if u (shadow price)

Dual scissors

- **Good**, to separate technologies + inexact information
- **Good**, if output of interest if u (shadow price)
- Bad, if output of interest is x, y: the final ones may be infeasible

Dual scissors

- **Good**, to separate technologies + inexact information
- **Good**, if output of interest if u (shadow price)
- **Bad**, if output of interest is x, y: the final ones may be infeasible

(dual approach solves the bi-dual of original problem)

Dual scissors

- **Good**, to separate technologies + inexact information
- **Good**, if output of interest if u (shadow price)
- **Bad**, if output of interest is x, y: the final ones may be infeasible

(dual approach solves the bi-dual of original problem)

Scissor features

-  Primal scissors — **primal feasibility**
-  Dual scissors — **separability**

Scissor features

-  Primal scissors **primal feasibility**
-  Dual scissors **separability**

Can we have both (x, y) and u???

Scissor features

-  Primal scissors — **primal feasibility**
-  Dual scissors — **separability**

Can we have both (x, y) and u???

Need **primal-dual** scissors

Primal-dual scissors: Augmented Lagrangians

$$L_r(x,y,u) = L(x,y,u) + \frac{r}{2}\|d - y_T - y_H\|^2$$

- **Good**, closes duality gap

Primal-dual scissors: Augmented Lagrangians

$$L_r(x, y, u) = L(x, y, u) + \frac{r}{2}\|d - y_T - y_H\|^2$$

- **Good**, closes duality gap
- Bad, puts back together the technologies

Primal-dual scissors: Augmented Lagrangians

$$L_r(x, y, u) = L(x, y, u) + \frac{r}{2}\|d - y_T - y_H\|^2$$

- **Good**, closes duality gap
- **Bad**, puts back together the technologies
- Also: inexact calculations make it difficult to manage parameter r

Primal-dual scissors: Augmented Lagrangians

$$L_r(x,y,u) = L(x,y,u) + \frac{r}{2}\|d - y_T - y_H\|^2$$

- **Good**, closes duality gap
- **Bad**, puts back together the technologies
- Also: inexact calculations make it difficult to manage parameter r

Need to **sharpen** our scissors

Primal-dual scissors: Sharp Augmented Lagrangians

$$L_r(x,y,u) = L(x,y,u) + \frac{r}{2}\|d - y_T - y_H\|^2$$

- **Good**, closes duality gap
- Bad, puts back together the technologies
- Also: inexact calculations make it difficult to manage parameter r

Need to **sharpen** our scissors

$$L^{\#}(x,y,u,r) = L(x,y,u) + r|d - y_T - y_H|_1$$

Primal-dual scissors: Sharp Augmented Lagrangians

$$L_r(x,y,u) = L(x,y,u) + \frac{r}{2}\|d - y_T - y_H\|^2$$

- **Good**, closes duality gap
- Bad, puts back together the technologies
- Also: inexact calculations make it difficult to manage parameter r

Need to **sharpen** our scissors

$$L^{\#}(x,y,u,\boxed{r}) = L(x,y,u) + r|d - y_T - y_H|_1$$

$\boxed{r}$ is a dual variable

Generalized Augmented Lagrangians (GAL)

$$\left\{\begin{array}{rl} \min & f_T(x,y_T)+f_H(y_H) \\ \text{s.t.} & (x,y_T)\in\mathcal{S}_T\,, y_H\in\mathcal{S}_H \\ & y_T+y_H=d \end{array}\right. \iff \left\{\begin{array}{rl} \min & \varphi(x) \\ \text{s.t.} & x\in X \\ & h(x)=0 \end{array}\right.$$

Generalized Augmented Lagrangians (GAL)

The perturbation function p

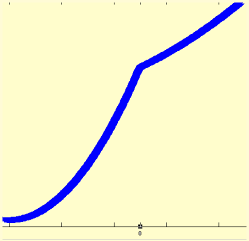
$$p(u) = \left\{ \begin{array}{ll} \inf & f_T(x, y_T) + f_H(y_H) \\ \text{s.t.} & (x, y_T) \in \mathcal{S}_T \,, y_H \in \mathcal{S}_H \\ & y_T + y_H = d + u \end{array} \right. \quad \Longleftrightarrow \quad p(u) = \left\{ \begin{array}{ll} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{array} \right.$$

Generalized Augmented Lagrangians (GAL)

The perturbation function p

$$p(u)=\left\{\begin{array}{ll}\inf & f_T(x,y_T)+f_H(y_H)\\ \text{s.t.} & (x,y_T)\in\mathcal{S}_T\,,y_H\in\mathcal{S}_H\\ & y_T+y_H=d+u\end{array}\right. \quad\Longleftrightarrow\quad p(u)=\left\{\begin{array}{ll}\inf & \varphi(x)\\ \text{s.t.} & x\in X\\ & h(x)=u\end{array}\right.$$

If in the example
$f_{T/H}$ are quadratic,
p is the minimum
of two quadratic functions

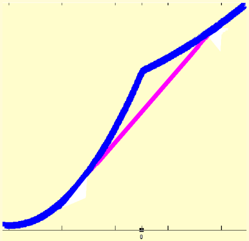


Generalized Augmented Lagrangians (GAL)

The perturbation function p

$$p(u) = \begin{cases} \inf & f_T(x, y_T) + f_H(y_H) \\ \text{s.t.} & (x, y_T) \in \mathcal{S}_T\,, y_H \in \mathcal{S}_H \\ & y_T + y_H = d + u \end{cases} \iff p(u) = \begin{cases} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{cases}$$

If in the example
$f_{T/H}$ are quadratic,
p is the minimum
of two quadratic functions

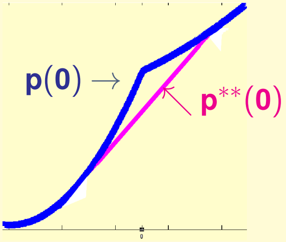


Generalized Augmented Lagrangians (GAL)

The perturbation function p

$$p(u) = \left\{ \begin{array}{ll} \inf & f_T(x, y_T) + f_H(y_H) \\ \text{s.t.} & (x, y_T) \in \mathcal{S}_T \,, y_H \in \mathcal{S}_H \\ & y_T + y_H = d + u \end{array} \right. \iff p(u) = \left\{ \begin{array}{ll} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{array} \right.$$

If in the example
$f_{T/H}$ are quadratic,
p is the minimum
of two quadratic functions

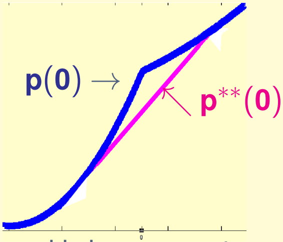


Generalized Augmented Lagrangians (GAL)

The perturbation function p

$$p(u) = \begin{cases} \inf & f_T(x, y_T) + f_H(y_H) \\ \text{s.t.} & (x, y_T) \in \mathcal{S}_T\,, y_H \in \mathcal{S}_H \\ & y_T + y_H = d + u \end{cases} \iff p(u) = \begin{cases} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{cases}$$

If in the example
$f_{T/H}$ are quadratic,
p is the minimum
of two quadratic functions



Need a “wedge” to close duality gap and bring magenta curve closer to blue one

The GAL according to R&W

$$p(u) = \begin{cases} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{cases}$$

The GAL according to R&W

$$p(u) = \left\{ \begin{array}{ll} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{array} \right. = \inf_{x \in X} \underbrace{\left(\varphi(x) + \mathbb{I}_{\{h(x)-u=0\}}(x) \right)}_{\mathcal{D}(x,u)}$$

The GAL according to R&W

$$p(u) = \begin{cases} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{cases} = \inf_{x \in X} \underbrace{\left(\varphi(x) + \mathbb{I}_{\{h(x)-u=0\}}(x)\right)}_{\mathcal{D}(x,u)}$$

- p as a marginal function of $\mathcal{D}$

The GAL according to R&W

$$p(u) = \left\{ \begin{array}{ll} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{array} \right. = \inf_{x \in X} \underbrace{\left(\varphi(x) + \mathbb{I}_{\{h(x)-u=0\}}(x) \right)}_{\mathcal{D}(x,u)}$$

- p as a marginal function of $\mathcal{D}$
- $\mathcal{D}$ is the conjugate of the Lagrangian: $\mathcal{D}^* = L$

The GAL according to R&W

$$p(u) = \left\{ \begin{array}{ll} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{array} \right. = \inf_{x \in X} \underbrace{\Big(\varphi(x) + \mathbb{I}_{\{h(x)-u=0\}}(x) \Big)}_{\mathcal{D}(x,u)}$$

- p as a marginal function of $\mathcal{D}$
- $\mathcal{D}$ is the conjugate of the Lagrangian: $\mathcal{D}^* = L$
- "Fix" $\mathcal{D}$ adding a "σ-term"

$$\mathcal{D}_\sigma \quad = \quad \mathcal{D} + \tfrac{1}{r}\sigma^*$$

The GAL according to R&W

$$p(u) = \begin{cases} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{cases} = \inf_{x \in X} \underbrace{\Big(\varphi(x) + \mathbb{I}_{\{h(x)-u=0\}}(x)\Big)}_{\mathcal{D}(x,u)}$$

- p as a marginal function of $\mathcal{D}$
- $\mathcal{D}$ is the conjugate of the Lagrangian: $\mathcal{D}^* = L$
- "Fix" $\mathcal{D}$ adding a "σ-term"

$$\begin{aligned} \mathcal{D}_\sigma &= \mathcal{D} + \tfrac{1}{r}\sigma^* \\ \mathcal{D}_\sigma^* &= \mathcal{D}^* \overset{+}{\vee} r\sigma \end{aligned}$$

The GAL according to R&W

$$p(u) = \begin{cases} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{cases} = \inf_{x \in X} \underbrace{\left(\varphi(x) + \mathbb{I}_{\{h(x)-u=0\}}(x)\right)}_{\mathcal{D}(x,u)}$$

- p as a marginal function of $\mathcal{D}$
- $\mathcal{D}$ is the conjugate of the Lagrangian: $\mathcal{D}^* = L$
- "Fix" $\mathcal{D}$ adding a "σ-term"

$$\begin{aligned} \mathcal{D}_\sigma &= \mathcal{D} + \tfrac{1}{r}\sigma^* \\ \mathcal{D}_\sigma^* &= \mathcal{D}^* \overset{+}{\vee} r\sigma \\ &= L_\sigma \text{ is the GAL} \end{aligned}$$

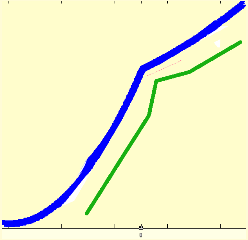
- $L_\sigma(x,u,r) = L(x,u) + r\sigma(h(x))$

The GAL according to R&W

$$p(u) = \begin{cases} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{cases} = \inf_{x \in X} \underbrace{\Big(\varphi(x) + \mathbb{I}_{\{h(x)-u=0\}}(x)\Big)}_{\mathcal{D}(x,u)}$$

- p as a marginal function of $\mathcal{D}$
- $\mathcal{D}$ is the conjugate of the Lagrangian: $\mathcal{D}^* = L$
- "Fix" $\mathcal{D}$ adding a "σ-term"

$$\begin{aligned} \mathcal{D}_\sigma &= \mathcal{D} + \tfrac{1}{r}\sigma^* \\ \mathcal{D}^*_\sigma &= \mathcal{D}^* \overset{+}{\vee} r\sigma \\ &= L_\sigma \text{ is the GAL} \end{aligned}$$



$\sigma = |\cdot|_1$

- $L_\sigma(x,u,r) = L(x,u) + r\sigma(h(x))$
- p_σ is the marginal function of $\mathcal{D}_\sigma$

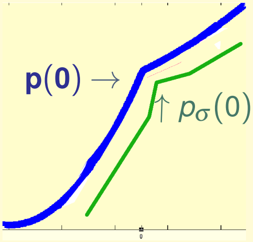
The GAL according to R&W

$$p(u) = \begin{cases} \inf & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = u \end{cases} = \inf_{x \in X} \underbrace{\left(\varphi(x) + \mathbb{I}_{\{h(x)-u=0\}}(x)\right)}_{\mathcal{D}(x,u)}$$

- p as a marginal function of $\mathcal{D}$
- $\mathcal{D}$ is the conjugate of the Lagrangian: $\mathcal{D}^* = L$
- "Fix" $\mathcal{D}$ adding a "σ-term"

$$\begin{aligned} \mathcal{D}_\sigma &= \mathcal{D} + \frac{1}{r}\sigma^* \\ \mathcal{D}_\sigma^* &= \mathcal{D}^* \overset{+}{\vee} r\sigma \\ &= L_\sigma \text{ is the GAL} \end{aligned}$$

- $L_\sigma(x,u,r) = L(x,u) + r\sigma(h(x))$
- p_σ is the marginal function of $\mathcal{D}_\sigma$



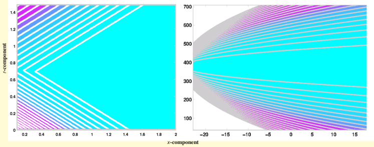
$\sigma = |\cdot|_1$

no duality gap!

Sharp and Proximal GAL according to R&W

$$L_\sigma(x,u,r)$$

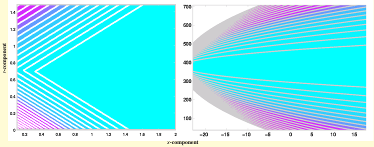
$$L(x,u)+r|h(x)|_1 \qquad \text{and} \qquad L(x,u)+\frac{r}{2}|h(x)|_2^2$$



Sharp and Proximal GAL according to R&W

$$L_\sigma(x,u,r)$$

$$L(x,u) + r|h(x)|_1 \qquad \text{and} \qquad L(x,u) + \frac{r}{2}|h(x)|_2^2$$

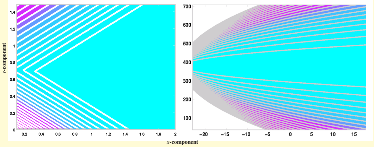


No duality gap, but two catches:

Sharp and Proximal GAL according to R&W

$$L_\sigma(x,u,r)$$

$$L(x,u) + r|h(x)|_1 \qquad \text{and} \qquad L(x,u) + \frac{r}{2}|h(x)|_2^2$$



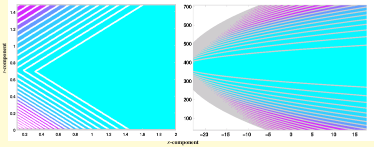
No duality gap, but two catches:

- separable $L(x,u) = L_T(x_T,u) + L_H(x_H,u)$
  turned into **non-separable** $L_\sigma(x,u,r) = L(x,u) + r|x_T + x_H - d|_1$

Sharp and Proximal GAL according to R&W

$$L_\sigma(x,u,r)$$

$$L(x,u) + r|h(x)|_1 \qquad \text{and} \qquad L(x,u) + \frac{r}{2}|h(x)|_2^2$$



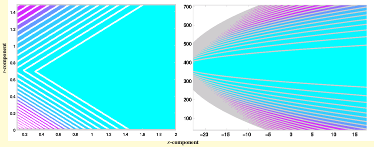
No duality gap, but two catches:

- separable $L(x,u) = L_T(x_T,u) + L_H(x_H,u)$
  turned into **non-separable** $L_\sigma(x,u,r) = L(x,u) + r|x_T + x_H - d|_1$
- requires exact evaluation of dual function $\theta_\sigma(u,r) = \min_{x \in X} L_\sigma(x,u,r)$
  a global optimization problem

Sharp and Proximal GAL according to R&W

$$L_\sigma(x,u,r)$$

$$L(x,u) + r|h(x)|_1 \qquad \text{and} \qquad L(x,u) + \frac{r}{2}|h(x)|_2^2$$



No duality gap, but two catches:

- separable $L(x,u) = L_T(x_T,u) + L_H(x_H,u)$
  turned into **non-separable** $L_\sigma(x,u,r) = L(x,u) + r|x_T + x_H - d|_1$
- requires exact evaluation of dual function $\theta_\sigma(u,r) = \min_{x\in X} L_\sigma(x,u,r)$
  a global optimization problem

Our work — eprints for the optimization community — **Revisiting Augmented Lagrangian Duals**

is a proposal to address these issues

GAL according to

Revisiting Augmented Lagrangian Duals

$$\text{GAL of}\ \begin{cases} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \end{cases}$$

GAL according to Revisiting Augmented Lagrangian Duals

$$\text{GAL of} \left\{ \begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \end{array} \right. \equiv \text{Lagrangian of} \left\{ \begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \\ & \sigma(h(x)) \leq 0 \end{array} \right.$$

GAL according to Revisiting Augmented Lagrangian Duals

$$
\text{GAL of} \begin{cases} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \end{cases} \equiv \text{Lagrangian of} \begin{cases} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \\ & \sigma(h(x)) \leq 0 \end{cases}
$$

From now on we consider $L(x, u, r)$ for problem on the right

GAL according to Revisiting Augmented Lagrangian Duals

$$
\text{GAL of} \left\{ \begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \end{array} \right. \quad \equiv \quad \text{Lagrangian of} \left\{ \begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \\ & \sigma(h(x)) \leq 0 \end{array} \right.
$$

From now on we consider $L(x, u, r)$ for problem on the right

to examine relations of approximate solutions to $\min_x L(x, u, r)$

$$\text{GAL of} \left\{ \begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \end{array} \right. \quad \equiv \quad \text{Lagrangian of} \left\{ \begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \\ & \sigma(h(x)) \leq 0 \end{array} \right.$$

From now on we consider $L(x,u,r)$ for problem on the right

to examine relations of approximate solutions to $\min_x L(x,u,r)$

- with ε-subgradients of the dual function $\theta(u,r)$
- with numerical schemes à la bundle
- with solutions to problem on the left

GAL according to Revisiting Augmented Lagrangian Duals

$$\text{GAL of } \left\{ \begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \end{array} \right. \quad \equiv \quad \text{Lagrangian of } \left\{ \begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \\ & \sigma(h(x)) \leq 0 \end{array} \right.$$

From now on we consider $L(x, u, r)$ for problem on the right

to examine relations of approximate solutions to $\min_x L(x, u, r)$

- with ε-subgradients of the dual function $\theta(u, r)$
- with numerical schemes à la bundle
- with solutions to problem on the left

GAL according to

$$\text{GAL of} \left\{ \begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \end{array} \right. \quad \equiv \quad \text{Lagrangian of} \left\{ \begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X \\ & h(x) = 0 \\ & \sigma(h(x)) \leq 0 \end{array} \right.$$

From now on we consider $L(x,u,r)$ for problem on the right

to examine relations of approximate solutions to $\min_x L(x,u,r)$

- with ε-subgradients of the dual function $\theta(u,r)$
- with numerical schemes à la bundle
- with solutions to problem on the left

$\boxed{\text{Any CQ for original problem yields multipliers } (u,r) \text{ for the } \sigma\text{-augmented problem}}$

Everett's theorem

σ **continuous, non-negative with unique minimizer at 0**

Everett's theorem

σ continuous, non-negative with unique minimizer at 0

Fix $(\tilde{u}, \tilde{r})$ and consider evaluating $\theta(\tilde{u}, \tilde{r}) = \min_x L(x, \tilde{u}, \tilde{r})$,

what primal problem is solved by minimizer $\tilde{x}$?

Everett's theorem σ continuous, non-negative with unique minimizer at 0

Fix $(\tilde{u}, \tilde{r})$ and consider evaluating $\theta(\tilde{u}, \tilde{r}) = \min_x L(x, \tilde{u}, \tilde{r})$,

what primal problem is solved by minimizer $\tilde{x}$?

NLP Everett's: A stationary point $\tilde{x}$ for $\theta(\tilde{u}, \tilde{r})$, i.e., solving

$$0 \in \partial_x L(x, \tilde{u}, \tilde{r}) + N_X(x), \text{ with } x \in X$$

is a stationary point for the original problem $\Longleftrightarrow$ $h(\tilde{x}) = 0$

Everett's theorem

σ continuous, non-negative with unique minimizer at 0

Fix $(\tilde{u}, \tilde{r})$ and consider evaluating $\theta(\tilde{u}, \tilde{r}) = \min_x L(x, \tilde{u}, \tilde{r})$,

what primal problem is solved by minimizer $\tilde{x}$?

NLP Everett's: A stationary point $\tilde{x}$ for $\theta(\tilde{u}, \tilde{r})$, i.e., solving

$$0 \in \partial_x L(x, \tilde{u}, \tilde{r}) + N_X(x)\,, \text{ with } x \in X$$

is a stationary point for the original problem $\Longleftrightarrow$ $h(\tilde{x}) = 0$

ε-Everett's: An ε-solution $\tilde{x}$ for $\theta(\tilde{u}, \tilde{r})$, i.e., satisfying

$$L(\tilde{x}, \tilde{u}, \tilde{r}) \leq L(x, \tilde{u}, \tilde{r}) + \varepsilon \text{ for all } x \in X$$

approximately solves the original problem ($\varphi(\tilde{x}) \leq \varphi^* + \varepsilon$) $\Longleftrightarrow$ $h(\tilde{x}) = 0$

Everett's theorem

σ continuous, non-negative with unique minimizer at 0

Fix $(\tilde{u}, \tilde{r})$ and consider evaluating $\theta(\tilde{u}, \tilde{r}) = \min_x L(x, \tilde{u}, \tilde{r})$,

what primal problem is solved by minimizer $\tilde{x}$?

NLP Everett's: A stationary point $\tilde{x}$ for $\theta(\tilde{u}, \tilde{r})$, i.e., solving

$$0 \in \partial_x L(x, \tilde{u}, \tilde{r}) + N_X(x)\,, \text{ with } x \in X$$

is a stationary point for the original problem $\Longleftrightarrow$ $h(\tilde{x}) = 0$

ε-Everett's: An ε-solution $\tilde{x}$ for $\theta(\tilde{u}, \tilde{r})$, i.e., satisfying

$$L(\tilde{x}, \tilde{u}, \tilde{r}) \leq L(x, \tilde{u}, \tilde{r}) + \varepsilon \text{ for all } x \in X$$

approximately solves the original problem $(\varphi(\tilde{x}) \leq \varphi^* + \varepsilon)$ $\Longleftrightarrow$ $h(\tilde{x}) = 0$

Everett's theorem

σ continuous, non-negative with unique minimizer at 0

Fix $(\tilde{u}, \tilde{r})$ and consider evaluating $\theta(\tilde{u}, \tilde{r}) = \min_x L(x, \tilde{u}, \tilde{r})$,

what primal problem is solved by minimizer $\tilde{x}$?

NLP Everett's: A stationary point $\tilde{x}$ for $\theta(\tilde{u}, \tilde{r})$, i.e., solving

$$0 \in \partial_x L(x, \tilde{u}, \tilde{r}) + N_X(x)\,, \text{ with } x \in X$$

is a stationary point for the original problem $\Longleftrightarrow$ $h(\tilde{x}) = 0$

ε-Everett's: An ε-solution $\tilde{x}$ for $\theta(\tilde{u}, \tilde{r})$, i.e., satisfying

$$L(\tilde{x}, \tilde{u}, \tilde{r}) \leq L(x, \tilde{u}, \tilde{r}) + \varepsilon \text{ for all } x \in X$$

approximately solves the original problem $(\varphi(\tilde{x}) \leq \varphi^* + \varepsilon)$ $\Longleftrightarrow$ $h(\tilde{x}) = 0$

- Summing up:
  - Augmentation (approximately) closes the duality gap, provided $h(\tilde{x}) = 0$

Everett's theorem

σ continuous, non-negative with unique minimizer at 0

Fix $(\tilde{u}, \tilde{r})$ and consider evaluating $\theta(\tilde{u}, \tilde{r}) = \min_x L(x, \tilde{u}, \tilde{r})$,

what primal problem is solved by minimizer $\tilde{x}$?

NLP Everett's: A stationary point $\tilde{x}$ for $\theta(\tilde{u}, \tilde{r})$, i.e., solving

$$0 \in \partial_x L(x, \tilde{u}, \tilde{r}) + N_X(x)\,, \text{ with } x \in X$$

is a stationary point for the original problem $\Longleftrightarrow$ $h(\tilde{x}) = 0$

ε-Everett's: An ε-solution $\tilde{x}$ for $\theta(\tilde{u}, \tilde{r})$, i.e., satisfying

$$L(\tilde{x}, \tilde{u}, \tilde{r}) \leq L(x, \tilde{u}, \tilde{r}) + \varepsilon \text{ for all } x \in X$$

approximately solves the original problem $(\varphi(\tilde{x}) \leq \varphi^* + \varepsilon)$ $\Longleftrightarrow$ $h(\tilde{x}) = 0$

- Summing up:
  - Augmentation (approximately) closes the duality gap, provided $h(\tilde{x}) = 0$
  - $h(x) = 0 \iff \sigma(h(x)) = 0$

Everett's theorem

σ **continuous, non-negative with unique minimizer at 0**

Fix $(\tilde{u},\tilde{r})$ and consider evaluating $\theta(\tilde{u},\tilde{r}) = \min_x L(x,\tilde{u},\tilde{r})$,

what primal problem is solved by minimizer $\tilde{x}$?

NLP Everett's: A stationary point $\tilde{x}$ for $\theta(\tilde{u},\tilde{r})$, i.e., solving

$$0 \in \partial_x L(x,\tilde{u},\tilde{r}) + N_X(x)\,,\ \text{with } x \in X$$

is a stationary point for the original problem $\Longleftrightarrow$ $h(\tilde{x}) = 0$

ε-Everett's: An ε-solution $\tilde{x}$ for $\theta(\tilde{u},\tilde{r})$, i.e., satisfying

$$L(\tilde{x},\tilde{u},\tilde{r}) \leq L(x,\tilde{u},\tilde{r}) + \varepsilon \text{ for all } x \in X$$

approximately solves the original problem $(\varphi(\tilde{x}) \leq \varphi^* + \varepsilon)$ $\Longleftrightarrow$ $h(\tilde{x}) = 0$

- Summing up:
  - Augmentation (approximately) closes the duality gap, provided $h(\tilde{x}) = 0$
  - $h(x) = 0 \iff \sigma(h(x)) = 0$
  - An inexact bundle method drives $\sigma(h(x^k))$ to 0

GAL and the σ-simplex

recall that $h(x) = 0 \iff \sigma(h(x)) = 0$

$$\theta(\tilde{u},\tilde{r}) = \min_{x\in X} L(x,\tilde{u},\tilde{r}) = \min_{x\in X} \varphi(x) + \langle \tilde{u}, h(x)\rangle + \tilde{r}\sigma(h(x))$$
$$\Longrightarrow \Big(h(x^{min}), \sigma(h(x^{min}))\Big) \in \partial_\varepsilon \theta(\tilde{u},\tilde{r})$$

GAL and the σ-simplex

recall that $h(x) = 0 \iff \sigma(h(x)) = 0$

$\theta(\tilde{u}, \tilde{r}) = \min_{x \in X} L(x, \tilde{u}, \tilde{r}) = \min_{x \in X} \varphi(x) + \langle \tilde{u}, h(x) \rangle + \tilde{r}\sigma(h(x))$

$\Longrightarrow \left(h(x^{min}), \sigma(h(x^{min})) \right) \in \partial_{\varepsilon} \theta(\tilde{u}, \tilde{r})$

Suppose X compact

Explicit calculus rule in

ε-Optimal solutions in nondifferentiable convex programming and some related questions

J. -J. Strodiot, V. Hien Nguyen & Norbert Heukemes

Mathematical Programming **25**, 307–328(1983)

GAL and the σ-simplex

recall that $h(x) = 0 \iff \sigma(h(x)) = 0$

$$\theta(\tilde{u}, \tilde{r}) = \min_{x \in X} L(x, \tilde{u}, \tilde{r}) = \min_{x \in X} \varphi(x) + \langle \tilde{u}, h(x) \rangle + \tilde{r}\sigma(h(x))$$

$$\Longrightarrow \left(h(x^{min}), \sigma(h(x^{min})) \right) \in \partial_\varepsilon \theta(\tilde{u}, \tilde{r})$$

Suppose X compact

Explicit calculus rule in

ε-Optimal solutions in nondifferentiable convex programming and some related questions

J. -J. Strodiot, V. Hien Nguyen & Norbert Heukemes

Mathematical Programming **25**, 307–328(1983)

$$\Longrightarrow \sigma\text{-simplex=conv}\left\{ \sigma(h(x^i)),\ x^i \text{ approximate minimizers} \right\}$$

GAL and the σ-simplex

recall that $h(x) = 0 \iff \sigma(h(x)) = 0$

$\theta(\tilde{u}, \tilde{r}) = \min_{x \in X} L(x, \tilde{u}, \tilde{r}) = \min_{x \in X} \varphi(x) + \langle \tilde{u}, h(x) \rangle + \tilde{r}\sigma(h(x))$

$\Longrightarrow \left(h(x^{min}), \sigma(h(x^{min})) \right) \in \partial_\varepsilon \theta(\tilde{u}, \tilde{r})$

Suppose X compact

Explicit calculus rule in

ε-Optimal solutions in nondifferentiable convex programming and some related questions

J. -J. Strodiot, V. Hien Nguyen & Norbert Heukemes

Mathematical Programming **25**, 307–328(1983)

$\Longrightarrow \sigma$-simplex=conv$\left\{ \sigma(h(x^i)),\ x^i \text{ approximate minimizers} \right\}$

Theorem: $0 \in \sigma$-simplex is equivalent to

- $0 \in \partial_\varepsilon \theta(\tilde{u}, \tilde{r})$
- One $x^{i_{best}}$ in the σ-simplex is primal feasible: $h(x^{i_{best}}) = 0$

$\Longrightarrow x^{i_{best}}$ approximate solution for original problem

PDBM: a primal-dual bundle method for GAL

Init Choose (u^1, r^1) and compute $x^1 \approx \min_{x \in X} L(x, u^1, r^1)$

Dual Solve bundle QP with a model for θ to obtain (u^+, r^+)

PDBM: a primal-dual bundle method for GAL

Init Choose (u^1, r^1) and compute $x^1 \approx \min_{x \in X} L(x, u^1, r^1)$

Dual Solve bundle QP with a model for θ to obtain (u^+, r^+)

Noise? Adjust prox-parameter and go to **Dual** if too much noise

PDBM: a primal-dual bundle method for GAL

Init Choose (u^1, r^1) and compute $x^1 \approx \min_{x \in X} L(x, u^1, r^1)$

Dual Solve bundle QP with a model for θ to obtain (u^+, r^+)

Noise? Adjust prox-parameter and go to **Dual** if too much noise

Primal Compute $x^+ \approx \min_{x \in X} L(x, u^+, r^+)$

PDBM: a primal-dual bundle method for GAL

Init Choose (u^1, r^1) and compute $x^1 \approx \min_{x \in X} L(x, u^1, r^1)$

Dual Solve bundle QP with a model for θ to obtain (u^+, r^+)

Noise? Adjust prox-parameter and go to **Dual** if too much noise

Primal Compute $x^+ \approx \min_{x \in X} L(x, u^+, r^+)$

Stop if $\sigma(h(x^+))$ is small

PDBM: a primal-dual bundle method for GAL

Init Choose (u^1, r^1) and compute $x^1 \approx \min_{x \in X} L(x, u^1, r^1)$

Dual Solve bundle QP with a model for θ to obtain (u^+, r^+)

Noise? Adjust prox-parameter and go to **Dual** if too much noise

Primal Compute $x^+ \approx \min_{x \in X} L(x, u^+, r^+)$

Stop if $\sigma(h(x^+))$ is small

Bundle Classify (u^+, r^+) as serious or null step, update the model

PDBM: a primal-dual bundle method for GAL

Init Choose (u^1, r^1) and compute $x^1 \approx \min_{x \in X} L(x, u^1, r^1)$

Dual Solve bundle QP with a model for θ to obtain (u^+, r^+)

Noise? Adjust prox-parameter and go to **Dual** if too much noise

Primal Compute $x^+ \approx \min_{x \in X} L(x, u^+, r^+)$

Stop if $\sigma(h(x^+))$ is small

Bundle Classify (u^+, r^+) as serious or null step, update the model

Loop to **Dual**

PDBM: a primal-dual bundle method for GAL

Init Choose (u^1, r^1) and compute $x^1 \approx \min_{x \in X} L(x, u^1, r^1)$

Dual Solve bundle QP with a model for θ to obtain (u^+, r^+)

Noise? Adjust prox-parameter and go to **Dual** if too much noise

Primal Compute $x^+ \approx \min_{x \in X} L(x, u^+, r^+)$

Stop if $\sigma(h(x^+))$ is small

Bundle Classify (u^+, r^+) as serious or null step, update the model

Loop to **Dual**

Theorem:

1. Noise attenuation loop is finite
2. There is a primal feasible limit point $\bar{x}^{i_{best}}$ that is approximately optimal
3. If the dual sequence has accumulation points, they solve approximately the dual problem
4. For existence of dual accumulation points, see “Convex proximal bundle methods in depth” MP2014

Solving difficult problems with PDBM for GAL

DC problems with explicit nonconvexity (exact solution of subproblems)

$$\begin{cases} \min\limits_{x \in X = \mathbb{R}^n} & \varphi(x) := \frac{1}{2}\langle x, Qx\rangle + \langle x, q\rangle - \max\limits_{i \in \{1,\ldots,N\}} \{\langle x, \alpha_i\rangle + \beta_i\} \\ \text{s.t.} & h(x) := Ax - b = 0 \end{cases}$$

Solving difficult problems with PDBM for GAL

DC problems with explicit nonconvexity (exact solution of subproblems)

$$\begin{cases} \min\limits_{x \in X = \mathbb{R}^n} & \varphi(x) := \frac{1}{2}\langle x, Qx\rangle + \langle x, q\rangle - \max\limits_{i \in \{1,\ldots,N\}} \{\langle x, \alpha_i\rangle + \beta_i\} \\ \text{s.t.} & h(x) := Ax - b = 0 \end{cases}$$

augmented with $\sigma(\cdot) = \frac{1}{2}\|\cdot\|_2^2$ yields an easy dual function $\theta(u, r) = \min_{i \in \{1,\ldots,N\}} \theta_i(u, r)$

Solving difficult problems with PDBM for GAL

DC problems with explicit nonconvexity (exact solution of subproblems)

$$\begin{cases} \min\limits_{x \in X = \mathbb{R}^n} & \varphi(x) := \frac{1}{2}\langle x, Qx\rangle + \langle x, q\rangle - \max\limits_{i\in\{1,\dots,N\}}\{\langle x, \alpha_i\rangle + \beta_i\} \\ \text{s.t.} & h(x) := Ax - b = 0 \end{cases}$$

augmented with $\sigma(\cdot) = \frac{1}{2}\|\cdot\|_2^2$ yields an easy dual function $\theta(u,r) = \min_{i\in\{1,\dots,N\}} \theta_i(u,r)$
Each θ_i is a QP having the same quadratic term for all i

$$\theta_i(u,r) = \min_x \frac{1}{2}\langle x, Qx\rangle + \langle x, q\rangle + \langle x, \alpha_i\rangle + \beta_i + \langle Ax - b, u\rangle + \frac{1}{2} r\|Ax - b\|_2^2$$

Solving difficult problems with PDBM for GAL

DC problems with explicit nonconvexity (exact solution of subproblems)

$$\begin{cases} \min\limits_{x \in X = \mathbb{R}^n} & \varphi(x) := \frac{1}{2}\langle x, Qx\rangle + \langle x, q\rangle - \max\limits_{i \in \{1,\dots,N\}} \{\langle x, \alpha_i\rangle + \beta_i\} \\ \text{s.t.} & h(x) := Ax - b = 0 \end{cases}$$

augmented with $\sigma(\cdot) = \frac{1}{2}\|\cdot\|_2^2$ yields an easy dual function $\theta(u,r) = \min_{i \in \{1,\dots,N\}} \theta_i(u,r)$
Each θ_i is a QP having the same quadratic term for all i

$$\theta_i(u,r) = \min_x \frac{1}{2}\langle x, Qx\rangle + \langle x, q\rangle + \langle x, \alpha_i\rangle + \beta_i + \langle Ax - b, u\rangle + \frac{1}{2} r\|Ax - b\|_2^2$$

	PDBM		MSM	(Gas02)	ENUM	
	avg	stdev	avg	stdev	avg	stdev
$\Delta\varphi$	-1E-03	3E-03	-7E-02	3E-01	0	0
$h(\bar{x})$	4E-08	3E-08	1E-03	6E-03		
#primal	25	42	105	115		
CPU (s)	5	8	21	36	209	515

Solving difficult problems with PDBM for GAL

Unit-commitment problems (inexact solution of subproblems, using ADMM)

$$
\left\{
\begin{array}{ll}
\min & \displaystyle\sum_{i \in I} \Big(\langle \mathcal{F}, x_i \rangle + C_i(y_i) \Big) \\
\text{s.t.} & (x_i, y_i) \in \mathcal{S}_i \\
 & \displaystyle\sum_i y_i = D
\end{array}
\right.
$$

Solving difficult problems with PDBM for GAL

Unit-commitment problems (inexact solution of subproblems, using ADMM)

$$
\left\{\begin{array}{ll}
\min & \sum_{i\in I}\Big(\langle \mathcal{F}, x_i\rangle + C_i(y_i)\Big)\\
\text{s.t.} & (x_i, y_i) \in \boldsymbol{\mathcal{S}}_i\\
 & \sum_i y_i = D
\end{array}\right.
\iff
\left\{\begin{array}{ll}
\min & \sum_{i\in I}\varphi_i(x_i, y_i)\\
\text{s.t.} & (x_i, y_i) \in \boldsymbol{\mathcal{S}}_i\\
 & \sum_i z_i = D\\
 & h(x,y) = z - y = 0
\end{array}\right.
$$

Solving difficult problems with PDBM for GAL

Unit-commitment problems (inexact solution of subproblems, using ADMM)

$$
\left\{
\begin{array}{ll}
\min & \sum_{i \in I} \Big(\langle \mathcal{F}, x_i \rangle + C_i(y_i) \Big) \\
\text{s.t.} & (x_i, y_i) \in \boldsymbol{\mathcal{S}}_i \\
 & \sum_i y_i = D
\end{array}
\right.
\iff
\left\{
\begin{array}{ll}
\min & \sum_{i \in I} \varphi_i(x_i, y_i) \\
\text{s.t.} & (x_i, y_i) \in \boldsymbol{\mathcal{S}}_i \\
 & \sum_i z_i = D \\
 & h(x, y) = z - y = 0
\end{array}
\right.
$$

augmented with $\sigma(\cdot) = |\cdot|_1$ gives

$$
L(x, y, z, u, r) = L_{0\text{-}1}(x, y, u) + L_{cont}(z, u) + r|y - z|_1
$$

Solving difficult problems with PDBM for GAL

Unit-commitment problems (inexact solution of subproblems, using ADMM)

$$\left\{\begin{array}{ll} \min & \sum_{i\in I}\Big(\langle \mathcal{F}, x_i\rangle + C_i(y_i)\Big) \\ \text{s.t.} & (x_i,y_i)\in \boldsymbol{\mathcal{S}}_i \\ & \sum_i y_i = D \end{array}\right. \iff \left\{\begin{array}{ll} \min & \sum_{i\in I}\varphi_i(x_i,y_i) \\ \text{s.t.} & (x_i,y_i)\in \boldsymbol{\mathcal{S}}_i \\ & \sum_i z_i = D \\ & h(x,y) = z - y = 0 \end{array}\right.$$

augmented with $\sigma(\cdot) = |\cdot|_1$ gives

$$\begin{aligned} L(x,y,z,u,r) &= L_{0\text{-}1}(x,y,u) + L_{cont}(z,u) + r|y-z|_1 \\ &\approx \underbrace{L_{0\text{-}1}(x,y,u) + \frac{r}{2}|y - z^{fixed}|_1} \quad + \quad \underbrace{L_{cont}(z,u) + \frac{r}{2}|y^{fixed} - z|_1} \end{aligned}$$

Solving difficult problems with PDBM for GAL

Unit-commitment problems (inexact solution of subproblems, using ADMM)

$$\left\{\begin{array}{ll} \min & \sum_{i\in I}\Big(\langle \mathcal{F}, x_i\rangle + C_i(y_i)\Big) \\ \text{s.t.} & (x_i, y_i) \in \mathcal{S}_i \\ & \sum_i y_i = D \end{array}\right. \iff \left\{\begin{array}{ll} \min & \sum_{i\in I} \varphi_i(x_i, y_i) \\ \text{s.t.} & (x_i, y_i) \in \mathcal{S}_i \\ & \sum_i z_i = D \\ & h(x,y) = z - y = 0 \end{array}\right.$$

augmented with $\sigma(\cdot) = |\cdot|_1$ gives

$$\begin{aligned} L(x,y,z,u,r) =& L_{0\text{-}1}(x,y,u) + L_{cont}(z,u) + r|y-z|_1 \\ \approx& \underbrace{L_{0\text{-}1}(x,y,u) + \frac{r}{2}|y - z^{fixed}|_1}_{\sum_i \min_{(x_i,y_i)\in\mathcal{S}_i}} + \underbrace{L_{cont}(z,u) + \frac{r}{2}|y^{fixed} - z|_1}_{\min_{\sum_i z_i = D}} \end{aligned}$$

Solving difficult problems with PDBM for GAL

Unit-commitment problems (inexact solution of subproblems, using ADMM)

Very good performance, provided r_0 **is well chosen** (not too large)

Results for 56 synthetic instances, horizon from 1 to 7 days, hourly discretization

	PDBM		MSM	(Gas02)
	avg	stdev	avg	stdev
$Gap(\%)$	4.5	6.5	14.6	12.1
$h(\bar{x})$	4.5E-03	7.2E-04	5.3E-03	1.1E-03
#primal	105	52	208	131
CPU (s)	96	121	148	140

Some works on GAL

- E. Golshtein and N. Tretyakov, Modified Lagrangians and Monotone Maps in Optimization, Wiley **1996**
- R. Rockafellar and R. Wets. Variational Analysis. Springer, **1998**
- A. M. Rubinov, B. M. Glover and X. Q. Yang, "Decreasing Functions with Applications to Penalization", **SiOPT 1999**
- R. N. Gasimov. "Augmented Lagrangian Duality and Nondifferentiable Optimization Methods in Nonconvex Programming",**JOGO 2002** $\longleftarrow$ *MSM*
- X. X. Huang and X. Q. Yang. "A Unified Augmented Lagrangian Approach to Duality and Exact Penalization",**MOR 2003**
- R. Burachik, R. Gasimov, N. Ismayilova, C. Kaya. "On a Modified Subgradient Algorithm for Dual Problems via Sharp Augmented Lagrangian",**JOGO 2006**
- R. Burachik and A. Rubinov. "Abstract Convexity and Augmented Lagrangians",**SiOPT 2007**
- A. Nedič, A. Ozdaglar, and A. Rubinov. "Abstract convexity for nonconvex optimization duality", **OPT 2007**
- R. Burachik, A. N. Iusem, and J. G. Melo. "A primal dual modified subgradient algorithm with sharp Lagrangian",**JOGO 2010**
- R. Burachik. "On primal convergence for augmented Lagrangian duality", **OPT 2011**
- R. Burachik, A. Iusem, and J. Melo. "An Inexact Modified Subgradient Algorithm for Primal-Dual Problems via Augmented Lagrangians", **JOTA 2013**
- N. L. Boland and A. C. Eberhard. "On the augmented Lagrangian dual for integer programming", **MP 2015**
- M. J. Feizollahi, S. Ahmed, and A. Sun. "Exact augmented Lagrangian duality for mixed integer linear programming",**MP 2017**
- A. M. Bagirov, G. Ozturk, and R. Kasimbeyli, "A sharp augmented Lagrangian-based method in constrained non-convex optimization", **OMS 2018**
- X. Gu, S. Ahmed, and S. S. Dey. "Exact Augmented Lagrangian Duality for Mixed Integer Quadratic Programming",**SiOPT 2020**

Some works on GAL

- E. Golshtein and N. Tretyakov, Modified Lagrangians and Monotone Maps in Optimization, Wiley **1996**
- R. Rockafellar and R. Wets. Variational Analysis. Springer, **1998**
- A. M. Rubinov, B. M. Glover and X. Q. Yang, "Decreasing Functions with Applications to Penalization", **SiOPT 1999**
- R. N. Gasimov. "Augmented Lagrangian Duality and Nondifferentiable Optimization Methods in Nonconvex Programming",**JOGO 2002** $\longleftarrow$ *MSM*
- X. X. Huang and X. Q. Yang. "A Unified Augmented Lagrangian Approach to Duality and Exact Penalization",**MOR 2003**
- R. Burachik, R. Gasimov, N. Ismayilova, C. Kaya. "On a Modified Subgradient Algorithm for Dual Problems via Sharp Augmented Lagrangian",**JOGO 2006**
- R. Burachik and A. Rubinov. "Abstract Convexity and Augmented Lagrangians",**SiOPT 2007**
- A. Nedič, A. Ozdaglar, and A. Rubinov. "Abstract convexity for nonconvex optimization duality", **OPT 2007**
- R. Burachik, A. N. Iusem, and J. G. Melo. "A primal dual modified subgradient algorithm with sharp Lagrangian",**JOGO 2010**
- R. Burachik. "On primal convergence for augmented Lagrangian duality", **OPT 2011**
- R. Burachik, A. Iusem, and J. Melo. "An Inexact Modified Subgradient Algorithm for Primal-Dual Problems via Augmented Lagrangians", **JOTA 2013**
- N. L. Boland and A. C. Eberhard. "On the augmented Lagrangian dual for integer programming", **MP 2015**
- M. J. Feizollahi, S. Ahmed, and A. Sun. "Exact augmented Lagrangian duality for mixed integer linear programming",**MP 2017**
- A. M. Bagirov, G. Ozturk, and R. Kasimbeyli, "A sharp augmented Lagrangian-based method in constrained non-convex optimization", **OMS 2018**
- X. Gu, S. Ahmed, and S. S. Dey. "Exact Augmented Lagrangian Duality for Mixed Integer Quadratic Programming",**SiOPT 2020**

Questions?