New Penalty Approaches for Bilevel Optimization Problems arising in Transportation Network Design

Jörg Fliege, Konstantinos Kaparis, & Huifu Xu

Melbourne, 2013

Contents

Current Section

The Problem

General setting: transportation network design & use.
At least two decision makers u ("upper") and ℓ ("lower").

1. Decision maker u (the network designer) makes decisions x_u on investment & maintenance costs, pricing, etc.
2. Decision maker(s) ℓ (network users) make network usage decisions x_ℓ. (For simplicitly, here only one lower-level decision maker. Can be generalized to the general case.)

Decision maker u tries to minimize its cost function $f_u(x_u, x_\ell)$; decision maker ℓ tries to minimize its own cost function $f_\ell(x_u, x_\ell)$.

Decision variables are

upper-level decision vector	$x_u \in \mathbb{R}^{n_u}$,
lower-level decision vector	$x_\ell \in \mathbb{R}^{n_\ell}$.

The Problem & Notation

For a *fixed* $x_u \in \mathbb{R}^{n_u}$, consider the parameterized *lower level problem*

$$\begin{array}{ll} \min\limits_{x_\ell} & f_\ell(x_u, x_\ell) \\ \text{subject to} & g_\ell(x_u, x_\ell) \le 0. \end{array} \qquad (P(x_u))$$

The *bilevel optimization problem* is now

$$\begin{array}{ll} \min\limits_{x_u, x_\ell} & f_u(x_u, x_\ell) \\ \text{subject to} & g_u(x_u, x_\ell) \le 0, \\ & x_\ell \text{ solves } P(x_u). \end{array} \qquad (1)$$

(Optimistic formulation.)

Note: upper level constraints g_u can also depend on x_ℓ.

Current Section

Optimizing over Abstract Sets

Let $C \subseteq \mathbb{R}^n$ be closed and $f \in \mathbf{C}^1(\mathbb{R}^n, \mathbb{R})$. Consider the problem

$$\begin{array}{rl} \min\limits_{x} & f(x) \\ \text{subject to} & x \in C. \end{array} \qquad (P)$$

Let $\|\cdot\|$ be the euclidean norm. For arbitrary $y \in \mathbb{R}^n$, denote by $\text{proj}_C(y)$ the *projection of y onto C*, i. e.

$$\text{proj}_C(y) := \arg\min_z \{\|y - z\| \mid z \in C\}.$$

Then, the first-order optimality condition for (P) holds if and only if

$$x \in \text{proj}_C\left(x - \nabla f(x)\right).$$

(See Eaves 1971, Harker & Pang 1990, Sun 1996, Fl. & V. 2004.)

Optimizing over Abstract Sets

Idea:
Solve

$$\mathrm{proj}_C\left(x - \nabla f(x)\right) = x$$

instead of

$$\begin{array}{ll} \min\limits_x & f(x) \\ \text{subject to} & x \in C. \end{array} \qquad (P)$$

Disadvantage: reformulation is nonsmooth.

Advantage: only knowlege of proj_C is assumed, and not of C.
(Especially, no explicit knowledge of functions g_i, h_j with
$C = \{x \mid g_i(x) \leq 0, h_j(x) = 0 \ \forall i \forall j\}$ required.)

Advantage: can easily be generalized if lower-level problem is equilibrium problem.

Optimizing over Abstract Sets

Situations where proj_C might be easier to handle than explicit constraint functions:

1. Information of C resides in a distributed computing environment: proj_C easy to compute, but Lagrangian hard to assemble. (Fl. 2006, 2010)
2. C a particular cone:
   1. C convex with "nice" dual C° (use Moreau: $x = \text{proj}_C(x) + \text{proj}_{C^\circ}(x)$): C *isotone cone* or *simplicial cone*, known only by extreme rays. (Nemeth et al '10, Ekart et al '10)
   2. C copositive cone? (Sponsel 2011)
   3. C epigraph of some matrix norm (spectral, nuclear, 1-norm, ∞-norm). (Ding et al 2010)
3. $C \subset \mathbb{R}^n$ polyhedron with m faces, $n \ll m$. (Llanas et al '00)
4. C complement of open polyhedron. (Mangasarian '00)
5. C set of correlation matrices. (Higham 2002)
6. $C = \{Y \in \mathbb{R}^{m \times m} \mid Y = Y^\top, Y_{i,i} = 1 \; \forall i\}$. (Qi & Sun, 2006)

Optimizing over Abstract Sets

Situations where proj_C might be easier to handle than explicit constraint functions:

1. Information of C resides in a distributed computing environment: proj_C easy to compute, but Lagrangian hard to assemble. (Fl. 2006, 2010)
2. C a particular cone:
    1. C convex with "nice" dual C° (use Moreau: $x = \text{proj}_C(x) + \text{proj}_{C^\circ}(x)$): C *isotone cone* or *simplicial cone*, known only by extreme rays. (Nemeth et al '10, Ekart et al '10)
    2. C copositive cone? (Sponsel 2011)
    3. C epigraph of some matrix norm (spectral, nuclear, 1-norm, ∞-norm). (Ding et al 2010)
3. $C \subset \mathbb{R}^n$ polyhedron with m faces, $n \ll m$. (Llanas et al '00)
4. C complement of open polyhedron. (Mangasarian '00)
5. C set of correlation matrices. (Higham 2002)
6. $C = \{Y \in \mathbb{R}^{m \times m} \mid Y = Y^\top, Y_{i,i} = 1 \ \forall i\}$. (Qi & Sun, 2006)

Optimizing over Abstract Sets

Situations where proj_C might be easier to handle than explicit constraint functions:

1. Information of C resides in a distributed computing environment: proj_C easy to compute, but Lagrangian hard to assemble. (Fl. 2006, 2010)
2. C a particular cone:
   1. C convex with "nice" dual C° (use Moreau: $x = \text{proj}_C(x) + \text{proj}_{C^\circ}(x)$): C *isotone cone* or *simplicial cone*, known only by extreme rays. (Nemeth et al '10, Ekart et al '10)
   2. C copositive cone? (Sponsel 2011)
   3. C epigraph of some matrix norm (spectral, nuclear, 1-norm, ∞-norm). (Ding et al 2010)
3. $C \subset \mathbb{R}^n$ polyhedron with m faces, $n \ll m$. (Llanas et al '00)
4. C complement of open polyhedron. (Mangasarian '00)
5. C set of correlation matrices. (Higham 2002)
6. $C = \{Y \in \mathbb{R}^{m\times m} \mid Y = Y^\top, Y_{i,i} = 1\ \forall i\}$. (Qi & Sun, 2006)

Optimizing over Abstract Sets

Situations where proj_C might be easier to handle than explicit constraint functions:

1. Information of C resides in a distributed computing environment: proj_C easy to compute, but Lagrangian hard to assemble. (Fl. 2006, 2010)
2. C a particular cone:
   1. C convex with "nice" dual C° (use Moreau: $x = \text{proj}_C(x) + \text{proj}_{C^\circ}(x)$): C *isotone cone* or *simplicial cone*, known only by extreme rays. (Nemeth et al '10, Ekart et al '10)
   2. C copositive cone? (Sponsel 2011)
   3. C epigraph of some matrix norm (spectral, nuclear, 1-norm, ∞-norm). (Ding et al 2010)
3. $C \subset \mathbb{R}^n$ polyhedron with m faces, $n \ll m$. (Llanas et al '00)
4. C complement of open polyhedron. (Mangasarian '00)
5. C set of correlation matrices. (Higham 2002)
6. $C = \{Y \in \mathbb{R}^{m \times m} \mid Y = Y^\top, Y_{i,i} = 1 \; \forall i\}$. (Qi & Sun, 2006)

Optimizing over Abstract Sets

Situations where proj_C might be easier to handle than explicit constraint functions:

1. Information of C resides in a distributed computing environment: proj_C easy to compute, but Lagrangian hard to assemble. (Fl. 2006, 2010)
2. C a particular cone:
   1. C convex with "nice" dual C° (use Moreau: $x = \text{proj}_C(x) + \text{proj}_{C^\circ}(x)$): C *isotone cone* or *simplicial cone*, known only by extreme rays. (Nemeth et al '10, Ekart et al '10)
   2. C copositive cone? (Sponsel 2011)
   3. C epigraph of some matrix norm (spectral, nuclear, 1-norm, ∞-norm). (Ding et al 2010)
3. $C \subset \mathbb{R}^n$ polyhedron with m faces, $n \ll m$. (Llanas et al '00)
4. C complement of open polyhedron. (Mangasarian '00)
5. C set of correlation matrices. (Higham 2002)
6. $C = \{Y \in \mathbb{R}^{m\times m} \mid Y = Y^\top, Y_{i,i} = 1\ \forall i\}$. (Qi & Sun, 2006)

Optimizing over Abstract Sets

Situations where proj_C might be easier to handle than explicit constraint functions:

1. Information of C resides in a distributed computing environment: proj_C easy to compute, but Lagrangian hard to assemble. (Fl. 2006, 2010)
2. C a particular cone:
   1. C convex with "nice" dual C° (use Moreau: $x = \text{proj}_C(x) + \text{proj}_{C^\circ}(x)$): C *isotone cone* or *simplicial cone*, known only by extreme rays. (Nemeth et al '10, Ekart et al '10)
   2. C copositive cone? (Sponsel 2011)
   3. C epigraph of some matrix norm (spectral, nuclear, 1-norm, ∞-norm). (Ding et al 2010)
3. $C \subset \mathbb{R}^n$ polyhedron with m faces, $n \ll m$. (Llanas et al '00)
4. C complement of open polyhedron. (Mangasarian '00)
5. C set of correlation matrices. (Higham 2002)
6. $C = \{Y \in \mathbb{R}^{m \times m} \mid Y = Y^\top, Y_{i,i} = 1\ \forall i\}$. (Qi & Sun, 2006)

Optimizing over Abstract Sets

Situations where proj_C might be easier to handle than explicit constraint functions:

1. Information of C resides in a distributed computing environment: proj_C easy to compute, but Lagrangian hard to assemble. (Fl. 2006, 2010)
2. C a particular cone:
   1. C convex with "nice" dual C° (use Moreau: $x = \text{proj}_C(x) + \text{proj}_{C^\circ}(x)$): C *isotone cone* or *simplicial cone*, known only by extreme rays. (Nemeth et al '10, Ekart et al '10)
   2. C copositive cone? (Sponsel 2011)
   3. C epigraph of some matrix norm (spectral, nuclear, 1-norm, ∞-norm). (Ding et al 2010)
3. $C \subset \mathbb{R}^n$ polyhedron with m faces, $n \ll m$. (Llanas et al '00)
4. C complement of open polyhedron. (Mangasarian '00)
5. C set of correlation matrices. (Higham 2002)
6. $C = \{Y \in \mathbb{R}^{m \times m} \mid Y = Y^\top, Y_{i,i} = 1 \ \forall i\}$. (Qi & Sun, 2006)

Optimizing over Abstract Sets

Situations where proj_C might be easier to handle than explicit constraint functions:

1. Information of C resides in a distributed computing environment: proj_C easy to compute, but Lagrangian hard to assemble. (Fl. 2006, 2010)
2. C a particular cone:
   1. C convex with "nice" dual C° (use Moreau: $x = \text{proj}_C(x) + \text{proj}_{C^\circ}(x)$): C *isotone cone* or *simplicial cone*, known only by extreme rays. (Nemeth et al '10, Ekart et al '10)
   2. C copositive cone? (Sponsel 2011)
   3. C epigraph of some matrix norm (spectral, nuclear, 1-norm, ∞-norm). (Ding et al 2010)
3. $C \subset \mathbb{R}^n$ polyhedron with m faces, $n \ll m$. (Llanas et al '00)
4. C complement of open polyhedron. (Mangasarian '00)
5. C set of correlation matrices. (Higham 2002)
6. $C = \{Y \in \mathbb{R}^{m \times m} \mid Y = Y^\top, Y_{i,i} = 1 \ \forall i\}$. (Qi & Sun, 2006)

Current Section

Reformulation of the lower level problem I

Use the reformulation on the lower level problem:

$$\begin{aligned} n &= n_\ell, \\ x &= x_\ell, \\ f &= f_\ell(x_u, \cdot), \\ C &= C(x_u) := \{z \in \mathbb{R}^{n_\ell} \mid g_\ell(x_u, z) \leq 0\} \end{aligned}$$

and define the nonsmooth function

$$P(x_u, x_\ell) := \mathrm{proj}_{C(x_u)}\left(x_\ell - \nabla_{x_\ell} f_\ell(x_u, x_\ell)\right) - x_\ell.$$

A reformulation of the bilevel problem is then

$$\begin{aligned} \min_{x_u, x_\ell} \quad & f_u(x_u, x_\ell) \\ \text{subject to} \quad & g_u(x_u, x_\ell) \leq 0, \\ & P(x_u, x_\ell) = 0. \end{aligned}$$

Current Section

Smoothness of Reformulation I

How smooth is $P(x_u, x_\ell) = \text{proj}_{C(x_u)}\left(x_\ell - \nabla_{x_\ell} f_\ell(x_u, x_\ell)\right) - x_\ell$**?**
I.e. let f_ℓ be sufficiently smooth. How smooth is $\text{proj}_{C(x_u)}(\ldots)$ w.r.t. (x_u, x_ℓ)?

Three easy special cases for fixed x_u:

- $\text{proj}_{C(x_u)}(\cdot) = \text{id}$ within $\text{int}(C(x_u))$.
- $y \in \text{bd}(C(x_u))$; direction $d \in \mathbb{R}^{n_\ell}$ given:

$$(\text{proj}_{C(x_u)})'_+(y; d) = \text{proj}_{T(x_u, y)}(d),$$

  where $T(x_u, y)$ is the tangent cone of $C(x_u)$ at y (Zarantonello 1971).
- Let $C(x_u)$ have a $\mathbf{C}^2$-boundary. Then, $\text{proj}_{C(x_u)}(\cdot) \in \mathbf{C}^1(\mathbb{R}^{n_\ell} \setminus C(x_u))$, and explicit representations of the derivative exist (Holmes 1973).

Smoothness of Reformulation I

How smooth is $P(x_u, x_\ell) = \mathrm{proj}_{C(x_u)}\left(x_\ell - \nabla_{x_\ell} f_\ell(x_u, x_\ell)\right) - x_\ell$**?**
I.e. let f_ℓ be sufficiently smooth. How smooth is $\mathrm{proj}_{C(x_u)}(\dots)$ w.r.t. (x_u, x_ℓ)?

Three easy special cases for fixed x_u:

- $\mathrm{proj}_{C(x_u)}(\cdot) = \mathrm{id}$ within $\mathrm{int}(C(x_u))$.
- $y \in \mathrm{bd}(C(x_u))$; direction $d \in \mathbb{R}^{n_\ell}$ given:

$$(\mathrm{proj}_{C(x_u)})'_+(y; d) = \mathrm{proj}_{T(x_u, y)}(d),$$

  where $T(x_u, y)$ is the tangent cone of $C(x_u)$ at y (Zarantonello 1971).
- Let $C(x_u)$ have a $\mathbf{C}^2$-boundary. Then, $\mathrm{proj}_{C(x_u)}(\cdot) \in \mathbf{C}^1(\mathbb{R}^{n_\ell} \setminus C(x_u))$, and explicit representations of the derivative exist (Holmes 1973).

Smoothness of Reformulation I

How smooth is $P(x_u, x_\ell) = \text{proj}_{C(x_u)}\left(x_\ell - \nabla_{x_\ell} f_\ell(x_u, x_\ell)\right) - x_\ell$**?**
I.e. let f_ℓ be sufficiently smooth. How smooth is $\text{proj}_{C(x_u)}(\dots)$ w.r.t. (x_u, x_ℓ)?

Three easy special cases for fixed x_u:

- $\text{proj}_{C(x_u)}(\cdot) = \text{id}$ within $\text{int}(C(x_u))$.
- $y \in \text{bd}(C(x_u))$; direction $d \in \mathbb{R}^{n_\ell}$ given:

$$(\text{proj}_{C(x_u)})'_+(y; d) = \text{proj}_{T(x_u, y)}(d),$$

  where $T(x_u, y)$ is the tangent cone of $C(x_u)$ at y (Zarantonello 1971).
- Let $C(x_u)$ have a $\mathbf{C}^2$-boundary. Then, $\text{proj}_{C(x_u)}(\cdot) \in \mathbf{C}^1(\mathbb{R}^{n_\ell} \setminus C(x_u))$, and explicit representations of the derivative exist (Holmes 1973).

Smoothness of Reformulation I

Theorem (Directional Differentiability) *Assume the following:*

1. $f_\ell \in \mathbf{C}^2(\mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}, \mathbb{R})$. *and* $g_\ell \in \mathbf{C}^2(\mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}, \mathbb{R}^{m_\ell})$.
2. *For each* $x_u \in \mathbb{R}^{n_u}$, $g_{\ell,i}(x_u, \cdot)$ *is convex.*
3. *Slater's condition for each lower level problem: for each* $x_u \in \mathbb{R}^{n_u}$, *there exists a* $z \in \mathbb{R}^{n_\ell}$ *with* $g_{\ell,i}(x_u, z) < 0$ *for all* i.
4. *There exists a constant* $\alpha > 0$, *such that, for all* (x_u, x_ℓ):

$$\|(\nabla_y g_\ell(x_u, y(x_u, x_\ell)))_{[:, i: g_{\ell,i}(x_u, P(x_u, x_\ell) - x_\ell) = 0]} v\| \geq \alpha \|v\|$$

for all $v \in \mathbb{R}^{\{i: g_{\ell,i}(x_u, P(x_u, x_\ell) - x_\ell) = 0\}}$.

Then, P is directionally differentiable at (x_u, x_ℓ) *in an arbitrary direction* $d \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}$ *and the forward and backward directional differentials can be computed by solving some explicitly known QPs.*

Smoothness of Reformulation I

Theorem (Gateaux Differentiability) *Let the same assumptions as in the last theorem hold and let the function g_ℓ not depend on x_u. Then, the function P is Gateaux differentiable if and only if strict complementarity holds:*

$$\{i : g_{\ell,i}(P(x_u, x_\ell) - x_\ell) = 0\} = \{j : \lambda_j(x_u, x_\ell) > 0\},$$

where $\lambda_j(x_u, x_\ell)$ are the Lagrangians of the projection problem

$$\begin{array}{ll} \min\limits_{y} & \|y - x_\ell + \nabla_{x_\ell} f_\ell(x_u, x_\ell)\| \\ \textit{subject to} & g_\ell(y) \leq 0. \end{array}$$

Again, differentials can be computed by solving some explicitly known QPs.

Exact Penalties for Reformulation

Theorem *Let $\nabla_{x_\ell} f_\ell$ be piecewise analytic; let f_u be Lipschitz continuous. Let $C : x_u \mapsto C(x_u)$ be continuous and convex for all x_u and let the mapping have the following property: for each $x_u \in \mathbb{R}^{n_u}$ and for each $y \in \mathrm{bd}(C(x_u))$ let there be a neighbourhood U of y such that there exists finitely many analytic and strongly convex functions $g_i(x_u, \cdot)$ such that*

$$C(x_u) \cap U = \{x_\ell \mid g_i(x_u, x_\ell) \leq 0 \; \forall \, i\}.$$

Let $\{(x_u, x_\ell) \mid g_u(x_u, x_\ell) \leq 0\}$ be compact and subanalytic. Then, there exists a constant $\beta^* > 0$ such that for all $\beta \geq \beta^*$ we have that

$$\|P(x_u, x_\ell)\|_1^{1/\beta}$$

is an exact penalty function for the reformulated problem.

Current Section

Numerical Results

Very preliminary results.

1. Purpose: sanity check. Does the reformulation make sense at all?
2. *Lazy* approach: reformulated problem solved with SLP/SQP code with ℓ_1-penalty for constraints and nonsmooth step length algorithm. (Previously implemented for ESA, European Space Agency.)
3. Differentials approximated by finite differences.

Numerical Results

Random bilinear problems with $n_u = n_\ell = 10$, $m_u = 1$, $m_\ell = 2$, feasibility & optimality tolerance 1e-6:

prob.	1	2	3	4	5	6	7	8	9	10
SLP iter	65	25	17	19	27	60	117	234	97	7
SQP iter	10f	43	6f	4f	11	15	5f	12	14	5

- All problems solved to specified accuracy by SLP.
- Central differences perform better than forward differences. (In contrary to theory?!)
- SQP performance sensitive to upper and lower starting point: code can jam at an infeasible point, restoration phase then unsuccessful.

Numerical Results

Test problems from literature, reformulated problems solved with IPOPT, all other settings as before.

problem	iter	feval
Shimizu & Aiyoshi I	170	932
Shimizu & Aiyoshi II	19	23
Bard1	27	78
Bard2	3000*	*
Aiyoshi & Shimuzu	12	18
Ye, Zhu, & Zhu	31	111

Current Section

Lessons learnt & Future Plans

- Reformulation provides flexible framework for bilevel problems.
- Can be approached with a variety of algorithms. What is the best approach to solve the reformulated problem?
- No assumption on uniqueness of lower level solutions.
- Further tests necessary to ascertain performance of the approach.
- Generalization to multilevel problems possible.
- Generalization to multiobjective lower level problems?

Questions?

Further information:

J.Fliege@soton.ac.uk
www.cormsis.soton.ac.uk