Two approaches to solve a particular class of bilevel problems

Joint work with Antoine Oustry, Claudia D'Ambrosio, Leo Liberti

Martina Cerulli

mcerulli@lix.polytechnique.fr

LIX - École polytechnique - CNRS

March 2, 2021

Bilevel programming

$$\min_{x,\,y} F(x,y)$$
$$s.t.\ G(x,y) \leq 0$$
$$y \in \arg\min_{y'} \{f(x,y') | g(x,y') \leq 0\}$$

Optimal value function transformation

One way to reformulate the bilevel problem is considering the so-called *optimal value function* of the lower-level problem:

$$\varphi(x) = \min_{y'} \{ f(x, y') | g(x, y') \leq 0 \},$$

obtaining:

Optimal value reformulation

$$\begin{aligned} \min_{x,\, y} \; & F(x, y) \\ s.t. \; & G(x, y) \leq 0 \\ & g(x, y) \leq 0 \\ & f(x, y) \leq \varphi(x). \end{aligned}$$

KKT transformation

If the lower level problem is convex, and a regularity condition as Slater's condition (or, equivalently, the (MFCQ)) is satisfied at all feasible points, it can be replaced by its KKT conditions, obtaining:

$$
\begin{aligned}
\min_{x,\, y}\ & F(x,y) \\
s.t.\ & G(x,y) \leq 0 \\
& \nabla_y f(x,y) + \lambda^\top \nabla_y g(x,y) = 0 \\
& g(x,y) \leq 0, \quad \lambda \geq 0 \\
& \lambda^\top \nabla_y g(x,y) = 0
\end{aligned}
$$

If the lower level problem is not convex, this formulation is a relaxation of the bilevel problem.

Our bilevel formulation

$$
\begin{aligned}
\min_{x \in \mathbb{R}^m} \quad & F(x) \\
s.t. \quad & G(x) \leq 0 \qquad\qquad (BP) \\
& h(x) \leq \min_{y \in \mathbb{R}^n} \{ \frac{1}{2} y^\top Q(x) y + q(x)^\top y | Ay \leq b \}
\end{aligned}
$$

Our bilevel formulation

$$\begin{aligned} \min_{x \in \mathbb{R}^m} \quad & F(x) \\ s.t. \quad & G(x) \leq 0 \\ & h(x) \leq \min_{y \in \mathbb{R}^n} \{ \frac{1}{2} y^\top Q(x) y + q(x)^\top y | Ay \leq b \} \end{aligned} \qquad (BP)$$

- no *argmin* operator but an inequality constraint that links the upper and the lower-level problems
- only the variable x in the upper level; $F(x)$, $G(x)$, $h(x)$ continuous in x
- a quadratic lower-level problem P_x with a feasible set $\mathcal{F} = \{y \in \mathbb{R}^n : Ay \leq b\} = \{y \in \mathbb{R}^n : a_j^T y \leq b_j, \forall j = 1, \ldots, r\}$, which is assumed not to depend on x

Our bilevel formulation

$$\begin{aligned} \min_{x\in\mathbb{R}^m} \quad & F(x) \\ s.t. \quad & G(x) \leq 0 \qquad\qquad (BP) \\ & h(x) \leq \min_{y\in\mathbb{R}^n} \{\frac{1}{2}y^\top Q(x)y + q(x)^\top y | Ay \leq b\} \end{aligned}$$

- no *argmin* operator but an inequality constraint that links the upper and the lower-level problems
  This class of programs arises in many applications requiring semi-infinite programming (SIP) problems, i.e. optimization problems with a finite number of variables and an infinite number of parametrized constraints of the type $\forall y \in Y$, $0 \leq f(x,y)$. In fact, if this is the case, it is sufficient to impose that this inequality holds for the minimum over all $y \in Y$ of $f(x,y)$ in the following way:

$$0 \leq \min_{y\in Y} f(x,y), \qquad (1)$$

reformulating the SIP program into a bilevel problem.

Our bilevel formulation

$$\begin{aligned} \min_{x \in \mathbb{R}^m} \quad & F(x) \\ s.t. \quad & G(x) \leq 0 \\ & h(x) \leq \min_{y \in \mathbb{R}^n} \{\frac{1}{2} y^\top Q(x) y + q(x)^\top y | Ay \leq b\} \end{aligned} \qquad (BP)$$

- no *argmin* operator but an inequality constraint that links the upper and the lower-level problems
- only the variable x in the upper level; $F(x)$, $G(x)$, $h(x)$ continuous in x
- a quadratic lower-level problem P_x with a feasible set $\mathcal{F} = \{y \in \mathbb{R}^n : Ay \leq b\} = \{y \in \mathbb{R}^n : a_j^T y \leq b_j, \forall j = 1, \ldots, r\}$, which is assumed not to depend on x

Our bilevel formulation

$$\begin{aligned} \min_{x\in\mathbb{R}^m} \quad & F(x) \\ s.t. \quad & G(x) \leq 0 \qquad\qquad (BP) \\ & h(x) \leq \min_{y\in\mathbb{R}^n} \{\frac{1}{2} y^\top Q(x) y + q(x)^\top y | Ay \leq b\} \end{aligned}$$

- no *argmin* operator but an inequality constraint that links the upper and the lower-level problems
- only the variable x in the upper level; $F(x)$, $G(x)$, $h(x)$ continuous in x
- a quadratic lower-level problem P_x with a feasible set $\mathcal{F} = \{y \in \mathbb{R}^n : Ay \leq b\} = \{y \in \mathbb{R}^n : a_j^T y \leq b_j, \forall j = 1, \ldots, r\}$, which is assumed not to depend on x

Dual approach

SDP reformulation of the LL

Assumption

The LL feasible set

$$\mathcal{F} = \{y \in \mathbb{R}^n : Ay \leq b\} = \{y \in \mathbb{R}^n : a_j^T y \leq b_j, \forall j = 1, \dots, r\}$$

is compact and included in the centered l_2-ball with radius $\rho > 0$, which is known.

We define the following matrices:

- $\mathcal{Q}(x) = \begin{pmatrix} \frac{1}{2} Q(x) & \frac{q(x)}{2} \\ \frac{q(x)}{2}^T & 0 \end{pmatrix}$,
- $\mathcal{A}_j = \begin{pmatrix} 0_n & \frac{a_j}{2} \\ \frac{a_j^T}{2} & 0 \end{pmatrix}, \quad \forall j = 1, \dots, r.$

SDP reformulation of the LL

With this notation, under Assumption 1, the problem

$$\begin{array}{llll} \min\limits_{Y \in \mathbb{R}^{(n+1)\times(n+1)}} & \langle \mathcal{Q}(x), Y \rangle & & \\ \text{s.t.} & \langle \mathcal{A}_j, Y \rangle & \leq & b_j, \qquad \forall j = 1, \ldots, r \\ & \operatorname{Tr}(Y) & \leq & 1 + \rho^2 \\ & Y_{n+1,n+1} & = & 1 \\ & Y & \succeq & 0 \\ & \operatorname{rank}(Y) & = & 1 \end{array}$$

is an exact reformulation of the quadratic lower level P_x, because any feasible matrix Y has the form $Y = \begin{pmatrix} y \\ 1 \end{pmatrix} \begin{pmatrix} y \\ 1 \end{pmatrix}^\top$ with $y \in \mathcal{F}$ such that $\langle \mathcal{Q}(x), Y \rangle = f(x, y)$.

SDP reformulation/relaxation of the LL

If we relax the non-convex constraint rank$(Y) = 1$ we obtain:

$$
\begin{array}{llll}
\min\limits_{Y \in \mathbb{R}^{(n+1)\times(n+1)}} & \langle \mathcal{Q}(x), Y \rangle & & \\
\text{s.t.} & \langle \mathcal{A}_j, Y \rangle & \leq & b_j, \qquad \forall j = 1, \ldots, r \\
& \text{Tr}(Y) & \leq & 1 + \rho^2 \\
& Y_{n+1,n+1} & = & 1 \\
& Y & \succeq & 0
\end{array}
$$

It is a relaxation of the quadratic problem P_x (we define it SDP_x). It is a reformulation of P_x if $Q \succeq 0$, despite the relaxation of the rank constraint.

SDP dual of SDP_x

Let E be a $(n+1) \times (n+1)$ matrix such that

$$E_{ij} = \begin{cases} 1 & \text{if } i = j = n+1 \\ 0 & \text{otherwise} \end{cases}$$

and I_{n+1} the identity matrix of size $(n+1) \times (n+1)$. The following SDP problem:

$$\begin{array}{ll} \max\limits_{\lambda \in \mathbb{R}^r_+,\, \alpha \in \mathbb{R}_+,\, \beta \in \mathbb{R}} & -b^\top \lambda - \alpha(1+\rho^2) - \beta \\ \text{s.t.} & \mathcal{Q}(x) + \sum\limits_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E \succeq 0, \end{array} \tag{$DSDP_x$}$$

is a dual problem of the problem (SDP_x).

Proof

The Lagrangian of the SDP problem is defined over $Y \in S_{n+1}^{+}(\mathbb{R})$, $\lambda \in \mathbb{R}_{+}^{r}$, $\alpha \in \mathbb{R}_{+}$, $\beta \in \mathbb{R}$ and reads

$$\begin{aligned} & L_x(Y, \lambda, \alpha, \beta) = \\ & = \langle \mathcal{Q}(x), Y \rangle + \sum_{j=1}^{r} [\lambda_j (\langle \mathcal{A}_j, Y \rangle - b_j)] + \alpha(\mathrm{Tr}(Y) - 1 - \rho^2) + \beta(Y_{n+1,n+1} - 1) \\ & = -\sum_{j=1}^{r} \lambda_j b_j - \alpha(1+\rho^2) - \beta + \left\langle \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E,\ Y \right\rangle. \end{aligned}$$

Proof

The Lagrangian of the SDP problem is defined over $Y \in S_{n+1}^{+}(\mathbb{R})$, $\lambda \in \mathbb{R}_{+}^{r}$, $\alpha \in \mathbb{R}_{+}$, $\beta \in \mathbb{R}$ and reads

$$
\begin{aligned}
&L_x(Y, \lambda, \alpha, \beta) = \\
&= \langle \mathcal{Q}(x), Y \rangle + \sum_{j=1}^{r} \left[\lambda_j \left(\langle \mathcal{A}_j, Y \rangle - b_j\right)\right] + \alpha(\mathrm{Tr}(Y) - 1 - \rho^2) + \beta(Y_{n+1,n+1} - 1) \\
&= -\sum_{j=1}^{r} \lambda_j b_j - \alpha(1 + \rho^2) - \beta + \left\langle \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E,\ Y \right\rangle .
\end{aligned}
$$

And its Lagrangian dual problem:

$$
\max_{\substack{\lambda \in \mathbb{R}_{+}^{r} \\ \alpha \in \mathbb{R}_{+} \\ \beta \in \mathbb{R}}} \quad \min_{Y \in S_{n+1}^{+}(\mathbb{R})} L_x(Y, \lambda, \alpha, \beta).
$$

Proof

The Lagrangian of the SDP problem is defined over $Y \in S_{n+1}^+(\mathbb{R})$, $\lambda \in \mathbb{R}_+^r$, $\alpha \in \mathbb{R}_+$, $\beta \in \mathbb{R}$ and reads

$$
\begin{aligned}
&L_x(Y, \lambda, \alpha, \beta) = \\
&= \langle \mathcal{Q}(x), Y \rangle + \sum_{j=1}^{r} \left[\lambda_j \left(\langle \mathcal{A}_j, Y \rangle - b_j\right)\right] + \alpha(\mathrm{Tr}(Y) - 1 - \rho^2) + \beta(Y_{n+1,n+1} - 1) \\
&= -\sum_{j=1}^{r} \lambda_j b_j - \alpha(1+\rho^2) - \beta + \left\langle \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E,\ Y \right\rangle .
\end{aligned}
$$

And its Lagrangian dual problem:

$$
\max_{\substack{\lambda \in \mathbb{R}_+^r \\ \alpha \in \mathbb{R}_+ \\ \beta \in \mathbb{R}}} \left(- \left(\sum_{j=1}^{r} \lambda_j b_j + \alpha(1+\rho^2) + \beta \right) + \min_{Y \in S_{n+1}^+(\mathbb{R})} \left\langle \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E, Y \right\rangle \right)
$$

Proof

The Lagrangian of the SDP problem is defined over $Y \in S^+_{n+1}(\mathbb{R})$, $\lambda \in \mathbb{R}^r_+$, $\alpha \in \mathbb{R}_+$, $\beta \in \mathbb{R}$ and reads

$$\begin{aligned} & L_x(Y, \lambda, \alpha, \beta) = \\ & = \langle \mathcal{Q}(x), Y \rangle + \sum_{j=1}^{r} [\lambda_j (\langle \mathcal{A}_j, Y \rangle - b_j)] + \alpha(\mathrm{Tr}(Y) - 1 - \rho^2) + \beta(Y_{n+1,n+1} - 1) \\ & = -\sum_{j=1}^{r} \lambda_j b_j - \alpha(1 + \rho^2) - \beta + \left\langle \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E,\ Y \right\rangle . \end{aligned}$$

And its Lagrangian dual problem:

$$\max_{\substack{\lambda \in \mathbb{R}^r_+ \\ \alpha \in \mathbb{R}_+ \\ \beta \in \mathbb{R}}} \left(-\left(\sum_{j=1}^{r} \lambda_j b_j + \alpha(1 + \rho^2) + \beta \right) + \boxed{\min_{Y \in S^+_{n+1}(\mathbb{R})} \left\langle \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E, Y \right\rangle} \right)$$

Proof

To prove strong duality, we prove that Slater condition holds for the dual problem

$$\begin{aligned} \max_{\lambda\in\mathbb{R}^r_+,\, \alpha\in\mathbb{R}_+,\, \beta\in\mathbb{R}} \quad & -b^\top\lambda - \alpha(1+\rho^2) - \beta \\ \text{s.t.} \quad & \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E \succeq 0, \end{aligned} \qquad (DSDP_x)$$

We denote by m_x the minimum eigenvalue of $\mathcal{Q}(x)$. By definition of m_x, matrix $\mathcal{Q}(x) + (1-m_x)I_{n+1}$ is positive definite. This is why $(0, 1-m_x, 0)$ is a strictly feasible point of ($DSDP_x$), i.e. Slater condition holds.

Primal-dual pair of the SDP reformulation/relaxation

Primal problem - SDP_x

$$\begin{array}{llll} \min\limits_{Y \in \mathbb{R}^{(n+1)\times(n+1)}} & \langle \mathcal{Q}(x), Y \rangle & & \\ \text{s.t.} & \langle \mathcal{A}_j, Y \rangle & \leq & b_j, \quad \forall j \\ & \mathrm{Tr}(Y) & \leq & 1+\rho^2 \\ & Y_{n+1,n+1} & = & 1 \\ & Y & \succeq & 0 \end{array}$$

Dual problem - $DSDP_x$

$$\begin{array}{ll} \max\limits_{\substack{\lambda \in \mathbb{R}^r_+ \\ \alpha \in \mathbb{R}_+ \\ \beta \in \mathbb{R}}} & -b^\top \lambda - \alpha(1+\rho^2) - \beta \\ \text{s.t.} & \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E \succeq 0 \end{array}$$

Primal-dual pair of the SDP reformulation/relaxation

Primal problem - SDP_x

$$
\begin{array}{llll}
\min\limits_{Y \in \mathbb{R}^{(n+1)\times(n+1)}} & \langle \mathcal{Q}(x), Y \rangle & & \\
\text{s.t.} & \langle \mathcal{A}_j, Y \rangle & \leq & b_j, \quad \forall j \\
& \mathrm{Tr}(Y) & \leq & 1 + \rho^2 \\
& Y_{n+1,n+1} & = & 1 \\
& Y & \succeq & 0
\end{array}
$$

Dual problem - $DSDP_x$

$$
\begin{array}{ll}
\max\limits_{\substack{\lambda \in \mathbb{R}^r_+ \\ \alpha \in \mathbb{R}_+ \\ \beta \in \mathbb{R}}} & -b^\top \lambda - \alpha(1 + \rho^2) - \beta \\
\text{s.t.} & \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E \succeq 0
\end{array}
$$

$\mathrm{val}(SDP_x) \leq \mathrm{val}(P_x)$

$\mathrm{val}(SDP_x) = \mathrm{val}(DSDP_x)$

Primal-dual pair of the SDP reformulation/relaxation

Primal problem - SDP_x

$$\begin{array}{llll} \min_{Y \in \mathbb{R}^{(n+1)\times(n+1)}} & \langle \mathcal{Q}(x), Y \rangle & & \\ \text{s.t.} & \langle \mathcal{A}_j, Y \rangle & \leq & b_j, \quad \forall j \\ & \mathrm{Tr}(Y) & \leq & 1+\rho^2 \\ & Y_{n+1,n+1} & = & 1 \\ & Y & \succeq & 0 \end{array}$$

Dual problem - $DSDP_x$

$$\begin{array}{ll} \max_{\substack{\lambda \in \mathbb{R}^r_+ \\ \alpha \in \mathbb{R}_+ \\ \beta \in \mathbb{R}}} & -b^\top \lambda - \alpha(1+\rho^2) - \beta \\ \text{s.t.} & \mathcal{Q}(x) + \sum_{j=1}^r \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E \succeq 0 \end{array}$$

$\mathrm{val}(SDP_x) \leq \mathrm{val}(P_x)$

$\mathrm{val}(SDP_x) = \mathrm{val}(DSDP_x)$

$$h(x) \leq \mathrm{val}(DSDP_x) \iff h(x) \leq \mathrm{val}(SDP_x) \implies h(x) \leq \mathrm{val}(P_x)$$

Primal-dual pair of the SDP reformulation/relaxation

Primal problem - SDP_x

$$\begin{array}{llll} \min\limits_{Y \in \mathbb{R}^{(n+1)\times(n+1)}} & \langle \mathcal{Q}(x), Y \rangle & & \\ \text{s.t.} & \langle \mathcal{A}_j, Y \rangle & \leq & b_j, \quad \forall j \\ & \mathrm{Tr}(Y) & \leq & 1 + \rho^2 \\ & Y_{n+1,n+1} & = & 1 \\ & Y & \succeq & 0 \end{array}$$

Dual problem - $DSDP_x$

$$\begin{array}{ll} \max\limits_{\substack{\lambda \in \mathbb{R}^r_+ \\ \alpha \in \mathbb{R}_+ \\ \beta \in \mathbb{R}}} & -b^\top \lambda - \alpha(1+\rho^2) - \beta \\ \text{s.t.} & \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E \succeq 0 \end{array}$$

$\mathrm{val}(SDP_x) = \mathrm{val}(P_x)$ $\qquad$ $\mathrm{val}(SDP_x) = \mathrm{val}(DSDP_x)$

$$h(x) \leq \mathrm{val}(DSDP_x) \iff h(x) \leq \mathrm{val}(SDP_x) \iff h(x) \leq \mathrm{val}(P_x)$$

If $Q(x) \succeq 0$ for all feasible x

An important step

Our bilevel formulation

$$\begin{aligned} \min_{x \in \mathbb{R}^m} \quad & F(x) \\ \text{s.t.} \quad & G(x) \leq 0 \\ & h(x) \leq \text{val}(P_x) \end{aligned}$$

An important step

A restriction/reformulation

$$\begin{array}{lll} \min_{x \in \mathbb{R}^m} & F(x) & \\ \text{s.t.} & G(x) & \leq 0 \\ & h(x) & \leq \text{val}(DSDP_x) \end{array}$$

because either

$$h(x) \leq \text{val}(DSDP_x) \iff h(x) \leq \text{val}(SDP_x) \implies h(x) \leq \text{val}(P_x)$$

or

$$h(x) \leq \text{val}(DSDP_x) \iff h(x) \leq \text{val}(SDP_x) \iff h(x) \leq \text{val}(P_x)$$

An important step

A restriction/reformulation

$$\begin{aligned}
\min_{x\in\mathbb{R}^m} \quad & F(x) \\
\text{s.t.} \quad & G(x) \leq 0 \\
& h(x) \leq \max_{\substack{\lambda\in\mathbb{R}^r_+ \\ \alpha\in\mathbb{R}_+ \\ \beta\in\mathbb{R}}} \{-b^\top\lambda - \alpha(1+\rho^2) - \beta : \\
& \qquad \mathcal{Q}(x) + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E \succeq 0\}
\end{aligned}$$

An important step

A restriction/reformulation

$$\begin{aligned}
\min_{x\in\mathbb{R}^m} \quad & F(x) \\
\text{s.t.} \quad & G(x) \preceq 0 \\
& h(x) \leq \max_{\substack{\lambda\in\mathbb{R}^r_+ \\ \alpha\in\mathbb{R}_+ \\ \beta\in\mathbb{R}}} \{-b^\top\lambda - \alpha(1+\rho^2) - \beta : \\
& \qquad \mathcal{Q}(x) + \sum_{j=1}^r \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E \succeq 0\}
\end{aligned}$$

We can replace it by:

$$\left.\begin{aligned}
h(x) \leq -b^\top\lambda - \alpha(1+\rho^2) - \beta \\
\mathcal{Q}(x) + \sum_{j=1}^r \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E \succeq 0
\end{aligned}\right\} \quad (*)$$

SDP restriction/reformulation of our BP

Given the lower-level dual variables λ, α, β:

$$\begin{aligned}
\min_{x,\lambda,\alpha,\beta} \quad & F(x) \\
\text{s.t.} \quad & G(x) \le 0 \\
& h(x) \le -\lambda^\top b - \alpha(1+\rho^2) - \beta \\
& \mathcal{Q}(x) + \textstyle\sum_j \lambda_j \mathcal{A}_j + \alpha I_{n+1} + \beta E \succeq 0 \\
& x \in \mathbb{R}^m, \lambda \in \mathbb{R}^r_+, \ \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}
\end{aligned} \qquad (BPR)$$

We remark that the single-level problem (BPR)

- is convex if $Q(x)$ and $q(x)$ depend on x linearly, while $F(x)$, $G(x)$ and $h(x)$ are convex,
- is a SDP problem if $Q(x)$ and $q(x)$ depend on x linearly, while $F(x)$, $G(x)$ and $h(x)$ are convex and semidefinite representable.

Cutting plane approach

Cutting plane approach

Let $k = 1$. Initialize the relaxation R_k of the bilevel problem (BP), obtained by considering the upper-level problem only.
while true **do**
    Solve R_k, obtaining the optimal solution x^k.
    Compute an optimal solution y^k of the LL problem for $x = x^k$.
    if $h(x^k) \leq \frac{1}{2}(y^k)^\top Q(x^k)y^k + q(x^k)^\top y^k$ **then**
        The algorithm terminates and (x^k, y^k) is the optimal solution of the bilevel formulation.
    else
        Define R_{k+1} as R_k with the adjoined inequality:

$$h(x) \leq \frac{1}{2}(y^k)^\top Q(x)y^k + q(x)^\top y^k.$$

        $k := k + 1$
    end if
end while

Applications

Constrained Quadratic Regression

Constrained Quadratic Regression

$$z = \frac{1}{2} w^\top \bar{Q} w + \bar{q}^\top w + \bar{c} + \epsilon$$

- $w \in \mathbb{R}^n$ features vector
- $z \in \mathbb{R}$ output
- $\bar{Q} \in \mathbb{R}^{n \times n}$ s.t. $\bar{Q} = \bar{Q}^\top$
- $\bar{q} \in \mathbb{R}^n$
- $\bar{c} \in \mathbb{R}$
- $\epsilon \sim \mathcal{N}(0, \sigma^2)$ Gaussian error

Constrained Quadratic Regression

$$z = \frac{1}{2} w^\top \bar{Q} w + \bar{q}^\top w + \bar{c} + \epsilon$$

- $w \in \mathbb{R}^n$ features vector
- $z \in \mathbb{R}$ output
- $\bar{Q} \in \mathbb{R}^{n \times n}$ s.t. $\bar{Q} = \bar{Q}^\top$
- $\bar{q} \in \mathbb{R}^n$
- $\bar{c} \in \mathbb{R}$
- $\epsilon \sim \mathcal{N}(0, \sigma^2)$ Gaussian error

Constrained Quadratic Regression

$$z = \frac{1}{2} w^\top \bar{Q} w + \bar{q}^\top w + \bar{c} + \epsilon$$

- $w \in \mathbb{R}^n$ features vector
- $z \in \mathbb{R}$ output
- $\bar{Q} \in \mathbb{R}^{n \times n}$ s.t. $\bar{Q} = \bar{Q}^\top$
- $\bar{q} \in \mathbb{R}^n$
- $\bar{c} \in \mathbb{R}$
- $\epsilon \sim \mathcal{N}(0, \sigma^2)$ Gaussian error

Constrained Quadratic Regression

$$z = \frac{1}{2} w^\top \bar{Q} w + \bar{q}^\top w + \bar{c} + \epsilon$$

- $w \in \mathbb{R}^n$ features vector
- $z \in \mathbb{R}$ output
- $\bar{Q} \in \mathbb{R}^{n \times n}$ s.t. $\bar{Q} = \bar{Q}^\top$
- $\bar{q} \in \mathbb{R}^n$
- $\bar{c} \in \mathbb{R}$
- $\epsilon \sim \mathcal{N}(0, \sigma^2)$ Gaussian error

Constrained Quadratic Regression

$$z = \frac{1}{2} w^\top \bar{Q} w + \bar{q}^\top w + \bar{c} + \epsilon$$

- $w \in \mathbb{R}^n$ features vector
- $z \in \mathbb{R}$ output
- $\bar{Q} \in \mathbb{R}^{n \times n}$ s.t. $\bar{Q} = \bar{Q}^\top$
- $\bar{q} \in \mathbb{R}^n$
- $\bar{c} \in \mathbb{R}$
- $\epsilon \sim \mathcal{N}(0, \sigma^2)$ Gaussian error

Constrained Quadratic Regression

$$z = \frac{1}{2} w^\top \bar{Q} w + \bar{q}^\top w + \bar{c} + \epsilon$$

- $w \in \mathbb{R}^n$ features vector
- $z \in \mathbb{R}$ output
- $\bar{Q} \in \mathbb{R}^{n \times n}$ s.t. $\bar{Q} = \bar{Q}^\top$
- $\bar{q} \in \mathbb{R}^n$
- $\bar{c} \in \mathbb{R}$
- $\epsilon \sim \mathcal{N}(0, \sigma^2)$ Gaussian error

Let us suppose that the parameters of this model are unknown, but we are given a dataset $(w_i, z_i)_{1 \leq i \leq P} \in (\mathbb{R}^n \times \mathbb{R})^P$.

Constrained quadratic regression problem

Problem:

finding the maximum likelihood estimator for $\bar{Q} \in \mathbb{R}^{n\times n}$, $\bar{q} \in \mathbb{R}^n$, $\bar{c} \in \mathbb{R}$

$$\Updownarrow$$

computing the triplet $(Q, q, c) \in \mathbb{R}^{n\times n} \times \mathbb{R}^n \times \mathbb{R}$ that minimizes the least-squares error $\sum_{i=1}^{P}(z_i - \frac{1}{2}w_i^T Q w_i - q^T w_i - c)^2$.

Constrained quadratic regression problem

Problem:

finding the maximum likelihood estimator for $\bar{Q} \in \mathbb{R}^{n\times n}$, $\bar{q} \in \mathbb{R}^n$, $\bar{c} \in \mathbb{R}$

$$\Updownarrow$$

computing the triplet $(Q, q, c) \in \mathbb{R}^{n\times n} \times \mathbb{R}^n \times \mathbb{R}$ that minimizes the least-squares error $\sum_{i=1}^{P}(z_i - \frac{1}{2}w_i^T Q w_i - q^T w_i - c)^2$.

Assumption

- *the features vector belongs to a given polytope* $\mathcal{F} \subset \mathbb{R}^n$
- *the noiseless value* $\frac{1}{2}y^T \bar{Q} y + \bar{q}^T y + \bar{c}$ *is non-negative for any* $y \in \mathcal{F}$

Problem formulation

Semi-infinite formulation

$$\begin{aligned}\min_{Q,q,c} \quad & \sum_{i=1}^{P}(z_i - \tfrac{1}{2}w_i^T Q w_i - q^T w_i - c)^2 \\ \text{s.t.} \quad & Q = Q^T \\ & \tfrac{1}{2}y^T Q y + q^T y + c \geq 0, \quad \forall y \in \mathcal{F} \\ & Q \in \mathbb{R}^{n\times n},\ q \in \mathbb{R}^n,\ c \in \mathbb{R}.\end{aligned}$$

Bilevel formulation

$$\begin{aligned}\min_{Q,q,c} \quad & \sum_{i=1}^{P}(z_i - \tfrac{1}{2}w_i^T Q w_i - q^T w_i - c)^2 \\ \text{s.t.} \quad & Q = Q^T \\ & \min_{y\in\mathcal{F}}\{\tfrac{1}{2}y^T Q y + q^T y\} \geq -c \\ & Q \in \mathbb{R}^{n\times n},\ q \in \mathbb{R}^n,\ c \in \mathbb{R}.\end{aligned}$$

Dual restriction/reformulation

Q not PSD $\Longrightarrow$ we obtain an **upper bound** of the BP

$$
\begin{aligned}
\min_{Q,q,c,\lambda,\alpha,\beta} \quad & \sum_{i=1}^{P} \left(z_i - \tfrac{1}{2} w_i^\top Q w_i - q^\top w_i - c\right)^2 \\
\text{s.t.} \quad & Q = Q^\top \\
& -\lambda^\top b - \alpha(1+\rho^2) - \beta \geq -c \\
& \tfrac{1}{2} \begin{pmatrix} Q + 2\alpha I_n & q \\ q^\top & 2(\beta+\alpha) \end{pmatrix} + \sum_{j=1}^{r} \lambda_j \mathcal{A}_j \succeq 0 \\
& Q \in \mathbb{R}^{n\times n},\ q \in \mathbb{R}^n,\ c \in \mathbb{R} \\
& \lambda \in \mathbb{R}^r_+,\ \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}.
\end{aligned}
$$

In general, it is a restriction of the original bilevel problem formulation since Q may not necessarily be PSD.

KKT relaxation/reformulation

Q not PSD $\Longrightarrow$ we obtain a **lower bound** of the BP

$$\begin{array}{rl} \min\limits_{Q,q,c,y,\gamma} & \sum\limits_{i=1}^{P}(z_i - \frac{1}{2}w_i^\top Q w_i - q^\top w_i - c)^2 \\ \text{s.t.} & Q = Q^\top \\ & \frac{1}{2}y^\top Q y + q^\top y \geq -c \\ & Ay \leq b \\ & Qy + q + A^\top \gamma = 0 \\ & \gamma^\top (Ay - b) = 0 \\ & Q \in \mathbb{R}^{n\times n},\ q \in \mathbb{R}^n,\ c \in \mathbb{R},\ y \in \mathbb{R}^n,\ \gamma \in \mathbb{R}^r_+, \end{array}$$

In general, it is a relaxation of the original bilevel problem formulation since Q may not necessarily be PSD.

Zero-sum game with quadratic payoff

Zero-sum game with quadratic payoff

Let us consider an undirected graph $\mathcal{G} = (V, E)$ (with $n = |V|$). Each player positions a resource on each node $i \in V$. After normalization, we can consider that the action set of both players is $\Delta_n = \{x \in \mathbb{R}_+ \ : \ \sum_{i=1}^{n} x_i = 1\}$.

2 players zero-sum game: $P_1(x, y) = -P_2(x, y)$, being $P_i(x, y)$ the payoff of player i related to the pair of strategies (x, y).

$\Longrightarrow$ We need to specify just one *game payoff* $P(x, y)$

Game payoff

The game payoff $P(x, y)$ related to the pair of strategies $(x, y) \in \Delta_n \times \Delta_n$ is

$$P(x, y) = -x^\top My + c_1(x) - c_2(x, y),$$

given by the sum of:

- the opposite of a term describing the "proximity" between x and y in the graph, with $M \in \mathbb{R}^{n \times n}$ is the matrix defined as $M_{ij} = 1$ if $i = j$ or $\{i, j\} \in E$, and $M_{ij} = 0$ otherwise
- the quadratic costs that player 1 has to pay to deploy his resources on the graph: $c_1(x) = \frac{1}{2}x^\top Q_1 x + q_1^\top x$
- the opposite of the quadratic costs that player 2 has to pay to deploy her resources on the graph, and that is influenced by player 1 strategy: $c_2(x, y) = \frac{1}{2}y^\top Q_2(x)y + q_2^\top y$.

Game payoff

The game payoff $P(x, y)$ related to the pair of strategies $(x, y) \in \Delta_n \times \Delta_n$ is

$$P(x, y) = -x^\top My + c_1(x) - c_2(x, y),$$

given by the sum of:

- the opposite of a term describing the "proximity" between x and y in the graph, with $M \in \mathbb{R}^{n\times n}$ is the matrix defined as $M_{ij} = 1$ if $i = j$ or $\{i, j\} \in E$, and $M_{ij} = 0$ otherwise
- the quadratic costs that player 1 has to pay to deploy his resources on the graph: $c_1(x) = \frac{1}{2}x^\top Q_1 x + q_1^\top x$
- the opposite of the quadratic costs that player 2 has to pay to deploy her resources on the graph, and that is influenced by player 1 strategy: $c_2(x, y) = \frac{1}{2}y^\top Q_2(x)y + q_2^\top y$.

Game payoff

The game payoff $P(x, y)$ related to the pair of strategies $(x, y) \in \Delta_n \times \Delta_n$ is

$$P(x, y) = -x^\top My + c_1(x) - c_2(x, y),$$

given by the sum of:

- the opposite of a term describing the "proximity" between x and y in the graph, with $M \in \mathbb{R}^{n \times n}$ is the matrix defined as $M_{ij} = 1$ if $i = j$ or $\{i, j\} \in E$, and $M_{ij} = 0$ otherwise
- the quadratic costs that player 1 has to pay to deploy his resources on the graph: $c_1(x) = \frac{1}{2}x^\top Q_1 x + q_1^\top x$
- the opposite of the quadratic costs that player 2 has to pay to deploy her resources on the graph, and that is influenced by player 1 strategy: $c_2(x, y) = \frac{1}{2}y^\top Q_2(x)y + q_2^\top y$.

Game payoff

The game payoff $P(x, y)$ related to the pair of strategies $(x, y) \in \Delta_n \times \Delta_n$ is

$$P(x, y) = -x^\top My + c_1(x) - c_2(x, y),$$

given by the sum of:

- the opposite of a term describing the "proximity" between x and y in the graph, with $M \in \mathbb{R}^{n\times n}$ is the matrix defined as $M_{ij} = 1$ if $i = j$ or $\{i, j\} \in E$, and $M_{ij} = 0$ otherwise
- the quadratic costs that player 1 has to pay to deploy his resources on the graph: $c_1(x) = \frac{1}{2}x^\top Q_1 x + q_1^\top x$
- the opposite of the quadratic costs that player 2 has to pay to deploy her resources on the graph, and that is influenced by player 1 strategy: $c_2(x, y) = \frac{1}{2}y^\top Q_2(x)y + q_2^\top y$.

This zero-sum game can then be written as

$$\min_{x\in\Delta_n}\max_{y\in\Delta_n} -x^\top My + \frac{1}{2}x^\top Q_1 x + q_1^\top x - \frac{1}{2}y^\top Q_2(x)y - q_2^\top y$$

Bilevel formulation

From player 1's perspective, this problem can be cast as the following bilevel formulation:

Bilevel formulation

$$
\begin{aligned}
\min_{x,v} \quad & v \\
\text{s.t.} \quad & \sum_{i=1}^{n} x_i = 1 \\
& v \geq \max_{y \in \Delta_n} -x^\top M y + \tfrac{1}{2} x^\top Q_1 x + q_1^\top x - \tfrac{1}{2} y^\top Q_2(x) y - q_2^\top y \\
& x \in \mathbb{R}^n_+, \; v \in \mathbb{R}.
\end{aligned}
$$

Bilevel formulation

From player 1's perspective, this problem can be cast as the following bilevel formulation:

Bilevel formulation

$$
\begin{aligned}
\min_{x,v} \quad & v \\
\text{s.t.} \quad & \sum_{i=1}^{n} x_i = 1 \\
& -v + \tfrac{1}{2}x^\top Q_1 x + q_1^\top x \leq \min_{y \in \Delta_n} \tfrac{1}{2} y^\top Q_2(x) y + (q_2 + M^\top x)^\top y \\
& x \in \mathbb{R}^n_+, \ v \in \mathbb{R}.
\end{aligned}
$$

Dual restriction/reformulation

Q not PSD for all feasible x $\Longrightarrow$ we obtain an **upper bound** of the BP.

$$
\begin{aligned}
\min_{x,v,\lambda,\alpha,\beta} \quad & v \\
\text{s.t.} \quad & \sum_{i=1}^{n} x_i = 1 \\
& -v + \frac{1}{2} x^\top Q_1 x + q_1^\top x \leq -\lambda_1 - 2\alpha - \beta \\
& \frac{1}{2} \begin{pmatrix} Q_2(x) + 2\alpha I_n & q_2 + M^\top x - \sum_{j=3}^{n+2} \lambda_j e_j + (\lambda_1 - \lambda_2) 1 \\ (q_2 + M^\top x - \sum_{j=3}^{n+2} \lambda_j e_j + (\lambda_1 - \lambda_2) 1)^\top & 2(\beta + \alpha) \end{pmatrix} \succeq 0 \\
& x \in \mathbb{R}_+^n, \; v \in \mathbb{R} \\
& \lambda \in \mathbb{R}_+^{n+2}, \alpha \in \mathbb{R}_+, \beta \in \mathbb{R},
\end{aligned}
$$

KKT relaxation/reformulation

Q not PSD for all x $\implies$ we obtain a **lower bound** of the BP

$$
\begin{aligned}
\min_{x,v,y,\gamma_1,\gamma_2} \quad & v \\
\text{s.t.} \quad & \sum_{i=1}^{n} x_i = 1 \\
& -v + \tfrac{1}{2}x^\top Q_1 x + q_1^\top x \le \tfrac{1}{2}y^\top Q_2(x)y + (q_2 + M^\top x)^\top y \\
& \sum_{i=1}^{n} y_i = 1 \\
& Q_2(x)y + q_2 + M^\top x + \gamma_1 1 - I_n\gamma_2 = 0 \\
& -\gamma_2^\top (I_n y) = 0 \\
& x \in \mathbb{R}^n_+,\ v \in \mathbb{R},\ y \in \mathbb{R}^n_+,\ \gamma_1 \in \mathbb{R},\ \gamma_2 \in \mathbb{R}^n_+.
\end{aligned}
$$

Thanks for your attention!