

Contents

Acknowledgments

Abstract

List of Symbols

List of Figures

List of Tables

0.1 Application of Outrata-theory to bilevel problem 1

German Summary

References

0.1 Application of Outrata-theory to bilevel problem

Remark: For multigroup (optimization in more than one variable) only “C-formulation” possible, not λ -formulation.

procedure taken from Outrata et al. in [1]

Follow steps in Appendix A:

A.1 Problem

bilevel problem has to be adapted to fit the theory

→ general: U_{ad} compact and convex, this means adapt constraint for the C_g ; (no disadvantage compared to grid search)

→ Inner level problem: introduce implicit bias

By adding a column of ones to the data matrix X the last component of $\tilde{w} := (w, w_b)^\top$ is the bias

Disadvantage: bias now also in regularization

Advantage: problem fits theory (for Hingequad)

→ Outer level problem: has to be continuously differentiable

Take hingequad function

reformulation of outer level problem as constrained smooth optimization problem (for both loss functions possible) seems to be not possible as outer level problem (A.1) in [1] is unconstrained in z except for the implicit function constraint

⇒ Bilevel problem:

1.

$$\begin{aligned}
\min_{\tilde{w}} \quad & \mathcal{L}_{\text{hingequad}}(\tilde{w}) = \frac{1}{T} \sum_{t=1}^T \frac{1}{|\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} \max \left\{ 1 - y_i \langle \tilde{w}^t, x^i \rangle, 0 \right\}^2 \\
\text{s.t.} \quad & \bar{c} \leq C_g \leq \bar{C}, \quad \text{for } \bar{c}, \bar{C} \in \mathbb{R}_+, \quad \bar{c} < \bar{C} \quad g = 1, \dots, G \\
& \text{for } t = 1, \dots, T \\
& \left\{ \begin{array}{l} \tilde{w}^t \in \arg \min_{\tilde{w}} \left\{ \frac{1}{2} \left(\|\tilde{w}\|_2^2 + \sum_{g=1}^G C_g \sum_{i \in \tilde{\mathcal{N}}_t^g} \xi_i^2 \right) \right\} \\ \text{s.t. for } i \in \bigcup_{g=1}^G \mathcal{N}_t^g : \quad y_i \langle \tilde{w}, x^i \rangle \geq 1 - \xi_i \end{array} \right\}
\end{aligned} \tag{0.1}$$

corresponding variables:

$x \rightarrow (C_1, \dots, C_G)$

y, z (both used for the same variable in Outrata) $\rightarrow (\tilde{w}, \xi)^\top$

A.2 Assumptions

- U_{ad} compact, nonempty ✓
- $\mathcal{L}_{\text{hingequad}}$ is continuously differentiable ✓
- S is single valued on \tilde{A} (open set containing U_{ad})
this follows from theorem 4.8 in [1, p. 82] ✓
derivation see below
- the considered GE is strongly regular (at all points (x, z) with $x \in \tilde{A}, z = S(x)$)
this follows from theorem 5.8 in [1, p. 96] ✓
derivation see below

Derivations

To assert the above assumptions rewrite lower level problem as given in A.3.1:

size of variables: $\tilde{w} \in \mathbb{R}^{\text{feat}+1}, \xi \in \mathbb{R}^N$

$$\begin{aligned} \min_{(\tilde{w}, \xi)} f(\tilde{w}, \xi) &:= \frac{1}{2} \left\langle \begin{pmatrix} \tilde{w} \\ \xi \end{pmatrix}, \begin{pmatrix} \mathbb{I} & \\ & C \end{pmatrix} \begin{pmatrix} \tilde{w} \\ \xi \end{pmatrix} \right\rangle \\ \text{s.t.} \quad &\begin{pmatrix} -y_1(x^1)^\top & -1 & & \\ & \vdots & \ddots & \\ -y_N(x^N)^\top & & & -1 \end{pmatrix} \begin{pmatrix} \tilde{w} \\ \xi \end{pmatrix} \leq \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix} \\ &\text{for } i \in \bigcup_{g=1}^G \mathcal{N}_t^g \end{aligned} \quad (0.2)$$

with $C := (\underbrace{C_1, \dots, C_1}_{|\mathcal{N}_t^1| \text{ times}}, \dots, \underbrace{C_G, \dots, C_G}_{|\mathcal{N}_t^G| \text{ times}})^\top$ and $N := |\bigcup_{g=1}^G \mathcal{N}_t^g|$.

Definition 0.1 ([1, Definition 4.2, p. 79]) The function F is said to be strongly monotone on Ω if there exists an $\alpha > 0$ such that

$$\langle F(v) - F(w), v - w \rangle \geq \alpha \|v - w\|^2 \quad \text{for all } v, w \in \Omega.$$

Theorem 0.2 ([1, Theorem 4.4 (ii), p. 79]) *If F is strongly monotone on Ω , then the equilibrium problem has exactly one solution.*

Check strong monotonicity for $F = \nabla f$:

$$\left\langle \begin{pmatrix} \mathbb{I} \\ C \end{pmatrix} (v - w), v - w \right\rangle = \sum_{i=1}^N \lambda_i (v_i - w_i)^2 \geq \min\{1, \lambda_{\min}\} \|v - w\|^2, \quad \forall v, w \in \mathbb{R}^{feat+1+N}.$$

Here the $\lambda_i > 0$ are the eigenvalues and λ_{\min} the largest eigenvalue of the positive definite matrix $\begin{pmatrix} \mathbb{I} \\ C \end{pmatrix}$. As the C_g are constrained away from 0 all eigenvalues are strictly larger than zero. Thus the theorem applies.

The set Ω is the constraint set of the lower level problem. It is given in the form of (4.3) in the book.

Theorem 0.3 ([1, Theorem 4.8, p. 82]) *Let F be strongly monotone on Ω given by (4.3) and $(\bar{w}, \bar{\xi})$ be the unique solution of the lower level problem. Assume that the linear independence constraint qualification holds at $(\bar{w}, \bar{\xi})$. Then the GE (4.14) (special form of the GE, provided in this case) possesses a unique solution $(\hat{y}, \hat{\lambda})$ with $\hat{y} = (\bar{w}, \bar{\xi})$.*

Theorem above essentially states that there is a unique solution of the KKT equations. Compare also to theorem 16.14 in [2].

Frage: wie ist $\bar{y} := y(\bar{x})$ in der (ESCQ) auf Seite 92 gemeint???

Ist \bar{y} die optimale Lösung des lower level Problems für den Parameter \bar{x} oder ist es einfach irgendein y und es soll lediglich die Zugehörigkeit in der (ESCQ) zum Parameter \bar{x} ausdrücken?

Theorem 0.4 ([1, Theorem 5.8, p. 96]) *Suppose that for some $x_0 \in \tilde{\mathcal{A}}$ the gradients $\nabla_y h^i(x_0, y_0)$, for $i = 1, \dots, l$ and $\nabla_y g^j(x_0, y_0)$, for all $j \in \{1, \dots, s\}$ corresponding to active inequalities at the point (x_0, y_0) are linearly independent.*

Additionally suppose that the Jacobian with respect to y of the Lagrangian $(J)_y \mathcal{L}(x_0, y_0, \mu_0, \lambda_0)$ is strictly copositive with respect to $\ker(\mathcal{J}_y H(x_0, y_0)) \cap \ker(\underbrace{\mathcal{J}_y G_{I+}}_{\text{only active inequalities with } \lambda^j > 0})(x_0, y_0)$.

Then the GE is strongly regular.

Linear independence:

no equality constraints

Jacobian of the inequality constraints:
$$\begin{pmatrix} -y_1(x^1)^\top & -1 & & \\ \vdots & & \ddots & \\ -y_N(x^N)^\top & & & -1 \end{pmatrix}$$

linearly independent because of identity in the last part of the matrix

Lagrangian of the lower level problem:

$$\mathcal{L}(\tilde{C}, (\tilde{w}, \xi), \lambda) = \begin{pmatrix} \mathbb{I} & \\ & C \end{pmatrix} \begin{pmatrix} \tilde{w} \\ \xi \end{pmatrix} + (\lambda_1, \dots, \lambda_N) \begin{pmatrix} -y_1(x^1)^\top & -1 & & \\ & \vdots & \ddots & \\ -y_N(x^N)^\top & & & -1 \end{pmatrix}$$

Jacobian with respect to (\tilde{w}, ξ) :

$$\nabla_{(\tilde{w}, \xi)} \mathcal{L}(\tilde{C}, (\tilde{w}, \xi), \lambda) = \begin{pmatrix} \mathbb{I} & \\ & C \end{pmatrix}$$

As shown before this matrix is uniformly positive definite. This means that it is also strictly copositive.

References

- [1] J. Outrata, M. Kočvara, and J. Zowe. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*. Springer US, 1998.
- [2] M. Ulbrich and S. Ulbrich. *Nichtlineare Optimierung*. Springer Basel AG, 2012.