Bilevel Hyperparameter Optimization for Nonlinear Support Vector Machines

Stefano Coniglio STEFANO.CONIGLIO@UNIBG.IT
Department of Economics
University of Bergamo
Bergamo, Via dei Caniana 2, 24127, Italy

Anthony Dunn ANTHONYDUNN@DAS-LTD.CO.UK
Decision Analysis Services Ltd
Grove House, Basingstoke RG24 8AG, UK

Qingna Li QNL@BIT.EDU.CN
School of Mathematics and Statistics
Key Laboratory of Mathematical Theory
and Computation in Information Security
Beijing Institute of Technology
Beijing, 100081, People's Republic of China

Alain Zemkoho A.B.ZEMKOHO@SOTON.AC.UK
School of Mathematical Sciences
University of Southampton
Southampton, University Rd, SO17 1BJ, UK

Contents

Abstract

While the problem of tuning the hyperparameters of a support vector machine (SVM) via cross-validation is easily understood as a bilevel optimization problem, so far, the corresponding literature has mainly focused on the linear-kernel case. In this paper, we establish a theoretical framework for the development of bilevel optimization-based methods for tuning the hyperparameters of an SVM in the case where a nonlinear kernel is adopted, which affords the ability to capture highly-complex relationships between the points in the data set. By leveraging a Karush-Kuhn-Tucker (KKT)/mathematical program with equilibrium constraints (MPEC) reformulation of the (lower-level) training problem, we develop a theoretical framework for the SVM hyperparameter-tuning problem that established under which assumptions and conditions suitable qualification conditions including the Mangasarian–Fromovitz, the linear-independence, and the strong second order sufficient conditions are satisfied. We then illustrate the need for this theoretical framework in the context of the well-known Scholtes relaxation algorithm for solving the MPEC reformulation of our bilevel hyperparameter problem for SVMs. Numerical experiments are conducted to demonstrate the potential of this algorithm for examples of nonlinear SVM problems.

Keywords: support vector classification, hyperparameter optimization, bilevel optimization, mathematical program with equilibrium constraints, constraint qualifications, strong second order sufficient conditions, Scholtes relaxation method

1 Introduction

Support vector machines (SVMs) Hearst et al. (1998) are among the most-used models in machine learning for tackling both regression and classification tasks. While the most basic version of the general SVM model (in which the data set is linearly separable and the separation margin is maximized) features no hyperparameters, many extremely-popular variants such as soft-margin and kernel SVMs feature one or more hyperparameters—parameters of the model which must be selected by the user before fitting the model to data in the *training* phase. The standard approach to selecting values for such hyperparameters in such a way that over-fitting the training data is avoided is to perform cross validation, wherein a portion of the data set (the *validation set*) is held out from training and the hyperparameter values are selected such that, after the model is fit to the training data, the model's predictions for the hereto unseen validation set minimize a given loss function. For both classification and regression tasks, the loss function that is typically used is the L_1-norm of the vector of misclassification errors.

The problem of choosing the best hyperparameter values within a cross-validation framework is easily understood as a bilevel optimization problem: it asks for minimizing a loss function of the model's predictions evaluated on the validation set (*the upper-level or leader's objective function*) subject to the model parameters minimizing another loss function evaluated on the the training set (*lower-level* or *follower's objective function*).

In the general case, such a bilevel program is complex to formulate and often prohibitively hard solve to (local or global) optimality. For this reason, many popular methods for hyperparameter selection (such as grid search, random search, and Bayesian optimization) rely on various schemes for sampling different hyperparameter combinations for which the training problem is solved and its solution evaluated to measure the loss function of the resulting trained model's predictions on the validation set Bishop and Nasrabadi (2006).

While some of such methods work reasonably well in practice, they all are, in essence, heuristics designed for solving the above-mentioned bilevel optimization problem and, as such, one may except that a better performance would be achieved if one were to develop *ad hoc* techniques for solving it that exploit its bilevel nature within a well-formalized mathematical optimization framework. Numerous attempts have indeed been made to formulate and solve the problem in such a way but, to the best of our knowledge, none studied the case where a nonlinear kernel is employed.

Hyperparameter tuning via bilevel optimization has also been used for classification models other than SVMs such as ℓ_p regression Okuno et al. (2021); Nguyen et al. (2023). In the bilevel optimization literature on hyperparameter tuning, most of the works consider the linear-kernel case with focus on either the support vector regression (SVR) or the support-vector classification (SVC) case Bennett et al. (2008, 2006); Moore et al. (2009); Kunapuli et al. (2008a,b); Li et al. (2022a,b); Wang and Li (2023). The nonlinear-kernel case is mentioned only in Kunapuli et al. (2008a,b), but it is not studied beyond emphasizing its importance in many practical applications. Interestingly, some of the aforementioned works show that, especially when using variants of the SVM model featuring additional hyperparameters as in Kunapuli et al. (2008b), bilevel methods are able to outperform sampling-based methods such as grid search.

1.1 Aim and scope of the paper

Since the nonlinear feature-space mapping does not admit a finite algebraic representation, in the formulation we propose we state the training problem (lower-level problem) in its dual form. While such a choice substantially complicates the analysis, thanks to the so-called *kernel trick* (see, e.g., Bishop and Nasrabadi (2006) the adoption of a dual formulation for the training problem leads us to a very general formulation that can be flexibly adopted to many other choices of a nonlinear feature-space mapping. Among many options for the choice of a nonlinear kernel, for the study of this paper we focus on the Radial Basis Function (RBF) kernel (also known as Gaussian kernel). Such a kernel has been used in many practical applications Prajapati and Patle (2010) and has been shown to outperform many other kernels such as the linear, polynomial, and sigmoid kernel in numerous cases Nanda et al. (2018); Yekkehkhany et al. (2014); Feizizadeh et al. (2017); Hong et al. (2017); Tbarki et al. (2016); Garrett et al. (2003). In particular, we remark that most of our analysis is independent of the adoption of the RBF kernel, which makes extending it to other (nonlinear) kernels rather easy.

The bilevel optimization problem we formulate features the SVM hyperparameters and the cross-validation function as, respectively, upper-level variables and upper-level objective function, while it features the variables of the dual of the nonlinear SVM training problem and the corresponding loss function as lower-level variables and objective function. Our formulation also includes a robust formulation for the bias b, which, while not directly available due to the adoption of the dual formulation, is crucial for stating the upper-level objective function.

Due to the way cross validation is defined, our bilevel problem features a single upper-level problem and a multitude of lower-level problems (one per split—see further) and, as such, it could be classified as a single-leader multi-follower optimization problem.

We then use the Karush-Kuhn-Tucker (KKT)/mathematical program with equilibrium constraints (MPEC) reformulation (a classical tool to transform a bilevel optimization problem into a single-level problem—see, e.g., Dempe and Zemkoho (2012, 2013) and references therein) to build a single-level formulation which is the core of the theoretical analysis we carry out in the paper.

In the field of MPEC, three qualification conditions are crucial for the theoretical understanding of the problem and the development of numerical methods; that is, the MPEC Mangasarian-Fromovitz constraint qualification (MPEC-MFCQ), the MPEC linear independence constraint qualification (MPEC-LICQ), and the MPEC strong second order sufficient condition (MPEC-SSOSC); see, e.g., Dempe and Zemkoho (2012, 2013); Flegel (2005); Ye et al. (1997); Guo et al. (2013); Kanzow and Schwartz (2013); Hoheisel et al. (2013); Scholtes (2001). Hence, the fundamental question that we address in this paper is the following one:

> *Can the MPEC-MFCQ, MPEC-LICQ, and MPEC-SSOSC be satisfied for the MPEC reformulation of our bilevel hyperparameter optimization problem for nonlinear SVMs?*

In Section 4, we prove that the MPEC-MFCQ is automatically satisfied for any feasible point of our MPEC reformulation. As for the MPEC-LICQ and MPEC-SSOSC, in, respectively, Sections 5 and 6 we carefully consider an exhaustive set of scenarios and construct conditions and numerical assumptions under which these conditions either fail or hold. Such a study is crucial for the development of efficient numerical algorithms for our problem, as we illustrate in Section 7 in the context of an application of the Scholtes relaxation method. In particular, thanks to our experimental results carried out on a number of real world data sets shows that our proposed bilevel optimization formulation for selecting the hyperparameter of a nonlinear SVM (with an RBF kernel) has the potential to outperform grid search.

1.2 Main contributions of the paper

In summary, the main contributions of the paper are as follows:

1. We introduce a bilevel optimization formulation of the hyperparameter optimization problem for a nonlinear SVM. A tractable transformation of this problem based on dual formulation of the SVM is then constructed and a framework ensuring that it is locally and globally equivalent to the corresponding KKT/MPEC reformulation is established.

2. We show that MPEC-MFCQ is automatically satisfied for the MPEC reformulation of the our bilevel hyperparameter optimization problem for nonlinear SVMs.

3. We provide an exhaustive analysis that specifically establish conditions under which the MPEC-LICQ holds or fails for the our bilevel hyperparameter optimization problem for nonlinear SVMs.

4. Unlike the MPEC-MFCQ and MPEC-LICQ, which only take the feasible set the problem under consideration, the MPEC-SSOSC involves the objective function. In our analysis, we exploit the structure of the feasible set of our bilevel hyperparameter optimization problem for nonlinear SVMs to establish conditions ensuring the MPEC-SSOC is satisfied or fail. We also extend our analysis under certain differentiability assumptions on the upper-level objective function of our problem.

5. We illustrate the need for the study of the MPEC-MFCQ, MPEC-LICQ, and MPEC-SSOSC in the context of the a Scholtes-relaxation algorithm which we design to solve the problem.

1.3 Structure of the paper

The remainder of the paper is structured as follows. In the next section, we introduce the nonlinear SVM (lower-level problem) and its dual form, construct the cross-validation loss function (upper-level objective) using the dual variables of the lower-level problem, and subsequently formulate the whole problem of computing optimal hyperparameter values for a nonlinear SVM via k-fold CV as a bilevel program. In Section 3, we introduce the KKT/MPEC single-level reformulation of the problem. Sections 4, 5, and 6 are devoted to the analysis of the MPEC-MFCQ, MPEC-LICQ, and MPEC-SSOSC, respectively. The need for these conditions is illustrated in Section 7 by showing how they can be utilized in the analysis of the Scholtes relaxation algorithm, which is then implemented and run on a collection of real-world data sets. Conclusions and possible future works are discussed in Section 8. For the sake of readability, almost all the proofs of the paper are relegated to the appendices.

2 Mathematical Formulation of the Problem

We start this section by establishing the notation we use throughout the paper. Given positive integer $n \in \mathbb{N}$, we denote the set of integers from 1 to n by $[n] := \{1, \ldots, n\}$. $\mathcal{I}_{n\times n}$ denotes the identity matrix in $\mathbb{R}^{n\times n}$, while $\mathbf{0}_{m\times n}$ denotes the zero matrix in $\mathbb{R}^{m\times n}$ and

$\mathbf{0}_n$ the zero vector in $\mathbb{R}^n$. Given a matrix $X \in \mathbb{R}^{m\times n}$ and two sets $\tau \subseteq [m]$ and $\kappa \subseteq [n]$, $X_{(\tau,\kappa)}$ denotes the submatrix of X with row and column indices in τ and κ, while $X(\tau,\cdot)$ and $X(\cdot,\tau)$ represent the submatrices of X containing, respectively, the rows and columns indexed by τ. Given a vector $x \in \mathbb{R}^n$, x_τ denotes the subvector of x with index set τ.

2.1 SVM training problem and k-fold cross validation

Let us consider N data points $(X_j, y_j)_{j\in[N]}$, where $X_i \in \mathbb{R}^l$ is the l-dimensional feature vector of the data point of index j and $y_j \in \{-1, 1\}$ is a binary label indicating whether the point belongs to one of two given classes $\mathcal{A}$ and $\mathcal{B}$.

In its basic version, a support vector machine (SVM) is a maximum margin (often called hard-margin) binary classifier for which the training problem is to find a hyperplane of equation $\omega^\top x + b = 0$ which separates the two classes of points. Here, $\omega \in \mathbb{R}^l$ is the vector orthogonal to the hyperplane, and $b \in \mathbb{R}$ is the intercept. The hyperplane should be such that $\omega^\top x + b > 0$ if $x \in \mathcal{A}$ and $\omega^\top x + b < 0$ if $x \in \mathcal{B}$ and the separation margin (i.e., the slab of width $\frac{1}{2}\|\omega\|_2$ centered about the hyperplane that contains no data points) associated with it should be as large as possible (Cortes and Vapnik, 1995).

Since in practice it is rarely the case that the two classes $\mathcal{A}$ and $\mathcal{B}$ are linearly separable (i.e. separable by a hyperplane), the notion of a soft-margin SVM is often adopted. The goal of a soft-margin SVM is to find a hyperplane that, at one time, maximizes the separation margin and minimizes the total misclassification error, the latter being defined as the sum over all points that end up on the wrong side of the hyperplane of their absolute residual $|\omega^\top x - b|$. The training problem for a soft-margin SVM reads as

$$\min_{\omega,\, b} \frac{1}{2}\|\omega\|^2 + C \sum_{j\in[N]} \max\left\{0, 1 - y_j\left(\omega^\top X_j + b\right)\right\},$$

where the hyperparameter $C \geq 0$ controls the relative contribution of each of the two terms: the inverse of the margin and the total misclassification error. We note that the case $C = 0$ is of no practical interest. This is because, with $C = 0$, the problem achieves an optimal solution (ω^*, b^*) of value 0 by setting $\omega^* = 0$ for any choice of $b^* \in \mathbb{R}$. Such a solution induces a degenerate hyperplane with normal vector $\omega^* = 0$ leading to an infinite separation margin $\frac{1}{\|\omega^*\|}$.

When a k-fold cross-validation framework (with $k \geq 2$) is in place, the data set is partitioned into k *folds*, from which k splits of the data set are generated. For each split $i \in [k]$, $\bar{n}^{(i)}$ and $\hat{n}^{(i)}$ denote the number of validation points and training points (Note that $\hat{n}^{(i)} + \bar{n}^{(i)} = N$ holds). Each split of index $i \in [k]$ features the i^{th} fold as validation set, which we denote by $\{(\bar{X}_j^{(i)}, \bar{y}_j^{(i)})\}_{j\in[\bar{n}^{(i)}]}$ (where, for each $j \in [\bar{n}^{(i)}]$, $\bar{y}_j^{(i)} \in \{1, -1\}$ is the label associated with the validation points $\bar{X}_j^{(i)}$) and the union of the $k-1$ remaining folds as training set, which we denote by $\{(\hat{X}_j^{(i)}, \hat{y}_j^{(i)})\}_{j\in[\hat{n}^{(i)}]}$ (where, for each $j \in [\hat{n}^{(i)}]$, $\hat{y}_j^{(i)} \in \{1, -1\}$ is the label associated with the training point $\hat{X}_j^{(i)}$).

Differently from the basic SVM model that we introduced before, we assume that, for each split $i \in [k]$, each training point $\hat{X}_j^{(i)}$, $j \in [\hat{n}^{(i)}]$, is embedded into a higher-dimensional *feature space* of dimension l^* thanks to the *feature-space mapping* $\phi_\gamma : \mathbb{R}^l \to \mathbb{R}^{l^*}$, where $l^* > l$ and $\gamma > 0$ is a parameter of the mapping (a hyperparameter of the SVM model).

When the RBF kernel is adopted, ϕ_γ is an infinite-dimensional function ($l^* = \infty$) whereas in the linear-SVM case, ϕ_γ is the identity function. Adopting this notation, the optimization problem of training a soft-margin SVM on each split $i \in [k]$ reads:

$$\min_{\omega^{(i)},\, b^{(i)}} \frac{1}{2}\|\omega^{(i)}\|^2 + C \sum_{j \in [\hat{n}^{(i)}]} \Xi_{\phi_\gamma}\left(\hat{y}_j^{(i)}, \omega^{(i)}, \hat{X}_j^{(i)}, b^{(i)}\right), \tag{1}$$

where the function Ξ_{ϕ_γ} is defined by $\Xi_{\phi_\gamma}(y, \omega, X, b) := \max\left\{0, 1 - y\left(\omega^\top \phi_\gamma(X) + b\right)\right\}$.

2.2 Bilevel optimization problem

We now formally introduce the bilevel hyperparameter optimization problem for nonlinear SVMs. In it, each lower-level problem belongs to the family of training problems introduced in (1). In the upper-level problem, the hyperparameters C and γ are tuned in such a way that the average generalization error evaluated over the k validation sets (across the k splits) is minimized. This is done by minimizing the following upper-level objective function:

$$\mathcal{L}_{\text{SVM}}(\omega, b, \bar{X}, \bar{y}) := \frac{1}{k}\sum_{i \in [k]} \sum_{j \in [\bar{n}^{(i)}]} \frac{1}{\bar{n}^{(i)}} \Xi_{\phi_\gamma}\left(\bar{y}_j^{(i)}, \omega^{(i)}, \bar{X}_j^{(i)}, b^{(i)}\right), \tag{2}$$

whose value decreases as the performance of the SVM model on the k validation sets increases. For each of the k splits, the lower-level problem will be problem (1) (training problem) and the upper-level problem will be to minimize the loss function $\mathcal{L}$ defined in equation (2).

Combining the k instances of problem (1) for the lower-level and equation (2) for the upper level, we obtain the following formulation of the bilevel hyperparameter optimization problem for nonlinear SVMs:

$$\begin{aligned} \min_{C,\, \gamma \geq 0,\, \omega,\, b} \quad & F(C, \gamma, \omega, b) := \frac{1}{k}\sum_{i \in [k]} \sum_{j \in [\bar{n}^{(i)}]} \frac{1}{\bar{n}^{(i)}} \Xi_{\phi_\gamma}\left(\bar{y}_j^{(i)}, \omega^{(i)}, \bar{X}_j^{(i)}, b^{(i)}\right) \\ \text{s.t.} \quad & (\omega^{(i)}, b^{(i)}) \in \operatorname*{argmin}_{\omega^{(i)},\, b^{(i)}} \left\{ \frac{1}{2}\|\omega^{(i)}\|^2 \right. \\ & \qquad\qquad \left. + C \sum_{j \in [\hat{n}^{(i)}]} \Xi_{\phi_\gamma}\left(\hat{y}_j^{(i)}, \omega^{(i)}, \hat{X}_j^{(i)}, b^{(i)}\right) \right\} \quad \text{for} \quad i \in [k], \end{aligned} \tag{3}$$

where $\omega := \left(\omega^{(1)}, \ldots, \omega^{(k)}\right) \in \mathbb{R}^{kl^*}$ and $b := \left(b^{(1)}, \ldots, b^{(k)}\right) \in \mathbb{R}^k$.

Differently from the single-level case where, with $C = 0$, any solution (ω^*, b^*) with $\omega^* = 0$ is optimal for any choice of $b^* \in \mathbb{R}$, in the bilevel case $b^{(i)*}$ must be chosen in such a way that the out-of-sample upper-level loss function is minimized. Such a value of $b^{(i)*}$ can be computed in closed-form according to the following proposition:

Proposition 1. *With $C = 0$ and for any choice of $\gamma \in \mathbb{R}_+$, problem (3) admits the following optimal solution:*

$$\textit{for all } i \in [k],\ \ \omega^{(i)*} = 0 \quad \textit{and} \quad b^{(i)*} = \begin{cases} +1 & \textit{if } |B^{(i)}| < |A^{(i)}|, \\ -1 & \textit{if } |B^{(i)}| > |A^{(i)}|, \\ \textit{any value in } [-1, 1] & \textit{if } |B^{(i)}| = |A^{(i)}|, \end{cases}$$

where $A^{(i)} := \left\{ j \in [\bar{n}^{(i)}] : \; y_j^{(i)} = 1 \right\}$ *and* $B^{(i)} := \left\{ j \in [\bar{n}^{(i)}] : \; y_j^{(i)} = -1 \right\}$.

In order to efficiently solve the problem and, in particular, to facilitate the use of the Karush-Kuhn-Tucker (KKT) reformulation to transform it into a single-level optimization problem, the first issue to address here is the nonsmoothness of the max operator appearing in each of the lower-level training problems. To proceed, we transform problem (3) into a bilevel program featuring a constrained lower-level problem per split of index $i \in [k]$ along the lines of Bishop and Nasrabadi (2006). We do so by applying a commonly-used lifting operation thanks to which, for each lower-level problem of index $i \in [k]$, the max operator is removed from the objective function at the cost of introducing a linear number of variables and constraints, one per data point and folder, where, for each split $i \in [k]$, the variable $\xi_j^{(i)} \geq 0$ denotes the misclassification error of data point $j \in [\hat{n}^{(i)}]$. This leads to the following reformulation:

$$\min_{C, \gamma \geq 0} \; F(C, \gamma, \omega, b) \quad \text{s.t.} \quad \left(\omega^{(i)}, b^{(i)}, \xi^{(i)}\right) \in S^{(i)}(C) \;\; \text{for} \;\; i \in [k], \tag{4}$$

where, for each split $i \in [k]$, $S^{(i)}(C)$ denotes the set of optimal solutions to the following reformulation of the ith lower-level problem:

$$\begin{aligned} \min_{\omega^{(i)}, b^{(i)}, \xi^{(i)}} \quad & \tfrac{1}{2}(\omega^{(i)})^\top \omega^{(i)} + C \sum_{j \in [\hat{n}^{(i)}]} \xi_j^{(i)} & \\ \text{s.t.} \quad & \hat{y}_j^{(i)} \left((\omega^{(i)})^\top \phi_\gamma(\hat{X}_j^{(i)}) + b^{(i)} \right) \geq 1 - \xi_j^{(i)} & \text{for } j \in [\hat{n}^{(i)}], \\ & \xi_j^{(i)} \geq 0 & \text{for } j \in [\hat{n}^{(i)}]. \end{aligned} \tag{5}$$

This problem is completely equivalent to problem (1).

Since a finite algebraic expression of ϕ_γ is not available if ϕ_γ is infinite-dimensional, one can define the *kernel function*

$$\mathbb{K}(X_r, X_s) := \phi_\gamma(X_r)^\top \phi_\gamma(X_s) \quad \text{for any pair} \quad (r, s) \in [N] \times [N] \tag{6}$$

in terms of the inner product between the maps of any two data points of index r and s into the (higher- or) infinite-dimensional feature space. In this work, we consider the Radial Basis Function (RBF) kernel defined by:

$$\mathbb{K}(X_r, X_s) := \exp(-\gamma \|X_r - X_s\|^2) \quad \text{for any pair} \;\; (r, s) \in [N] \times [N]. \tag{7}$$

Here $\gamma \geq 0$ is the hyperparameter. We recall that, as mentioned above, our analysis can be easily adapted to many other types of nonlinear kernels besides the RBF one.

By relying on the so-called *kernel trick* (see, e.g., Bishop and Nasrabadi (2006); Chung et al. (2003)), equations (6) and (7) can now be applied to the dual of the i^{th}-split training problem (5) to obtain a completely explicit formulation which does not include the (infinite-dimensional) map ϕ_γ:

$$\begin{aligned} \min_{\alpha^{(i)} \in \mathbb{R}^{\hat{n}^{(i)}}} \quad & \mathcal{H}^{(i)}\left(\gamma, \alpha^{(i)}\right) := \tfrac{1}{2}(\alpha^{(i)})^\top Q^i(\gamma) \alpha^{(i)} - (\alpha^{(i)})^\top \mathbf{e}^{(i)} \\ \text{s.t.} \quad & 0 \leq \alpha_j^{(i)} \leq C \quad \text{for } j \in [\hat{n}^{(i)}], \\ & \left(\alpha^{(i)}\right)^\top \hat{y}^{(i)} = 0. \end{aligned} \tag{LLP^i}$$

In this formulation, $\alpha^{(i)} \in \mathbb{R}^{\hat{n}^{(i)}}$ denotes the dual variable of the lower-level problem defined in (5), $\hat{y}^{(i)} := \left(\hat{y}_1^{(i)}, \cdots, \hat{y}_{\hat{n}^{(i)}}^{(i)}\right) \in \mathbb{R}^{\hat{n}^{(i)}}$, and $\mathbf{e}^{(i)}$ is the all-one vector in the space $\mathbb{R}^{n^{(i)}}$. $Q^i(\gamma) \in \mathbb{R}^{\hat{n}^{(i)} \times \hat{n}^{(i)}}$ is defined as

$$(Q^i(\gamma))_{rs} := \hat{y}_r^{(i)} \hat{y}_s^{(i)} \exp\left(-\gamma \|\hat{X}_r^{(i)} - \hat{X}_s^{(i)}\|_2^2\right) \quad \text{for } r, s \in [\hat{n}^{(i)}],\ i \in [k]. \tag{8}$$

Note that $C = 0$ implies $\alpha^{(i)} = 0$, which leads to a 0-valued objective function, in line with what we observed before for the primal problem. In particular, Lagrangian duality shows that, with $0 = \alpha = C = 0$, the dual multipliers μ satisfy $\mu = 0$, which allows for nonegative ξ (and, in particular, arbitrarily large ξ's, which coincides with entirely ignoring the misclassification error in the primal problem.

It is important to remark that, in problem LLP^i, C and γ are hyperparameters whose values must be known before the lower-level training problem can be solved. In the context of linear SVMs, the regularization parameter C it typically the only hyperparameter. However, as it can be seen from (LLP^i) and equation (8), the adoption of an RBF kernel leads to the introduction of the second hyperparameter γ.

From now on, problem (LLP^i) will serve as the lower-level problem for each split $i \in [k]$.

Proposition 2. *For a given $C \in \mathbb{R}_+$, let $\Omega_C : \mathbb{R} \to \mathbb{R}$ be any function satisfying $\Omega_C(\zeta) = 0$ for $\zeta \leq 0$ and $\zeta \geq C$ and $\Omega_C(\zeta) > 0$ all $\zeta \in (0, C)$. For any point (C, ω, b) that is feasible for problem* (4)*, there exists a vector $(C, \gamma, \alpha) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n$ such that the value of the function F defined in* (3) *coincides with that of the following function $\mathcal{F}$:*

$$\mathcal{F}(C, \gamma, \alpha) := \begin{cases} \dfrac{1}{k} \displaystyle\sum_{i \in [k]} \dfrac{1}{\bar{n}^{(i)}} \sum_{j \in [\bar{n}^{(i)}]} \max \Bigg\{ 0, 1 - \bar{y}_j^{(i)} \mathbb{H}^{(i)}\left(\bar{X}_j^{(i)}, \gamma\right) \\ \qquad + \dfrac{1}{\sum_{\ell \in [\hat{n}^{(i)}]} \Omega_C\left(\alpha_\ell^i\right)} \displaystyle\sum_{\ell \in [\hat{n}^{(i)}]} \Omega_C(\alpha_\ell^i) \left(\hat{y}_\ell^{(i)} - \mathbb{H}^{(i)}\left(\hat{X}_\ell^{(i)}, \gamma\right)\right) \Bigg\} & \text{if } C > 0, \\ \dfrac{1}{k} \displaystyle\sum_{i \in [k]} \dfrac{1}{\bar{n}^{(i)}} 2 \min\left\{ |A^{(i)}|, |B^{(i)}| \right\} & \text{if } C = 0, \end{cases}$$

where $\alpha := \left(\alpha^{(1)}, \ldots, \alpha^{(k)}\right) \in \mathbb{R}^{\hat{n}^{(1)} \times \cdots \times \hat{n}^{(k)}}$, $n := \sum_{i=1}^k \hat{n}^{(i)}$, and for all $i \in [k]$ and $X \in \mathbb{R}^l$, $A^{(i)}$ and $B^{(i)}$ are defined as in Proposition 1, while

$$\mathbb{H}^{(i)}(X, \gamma) \quad := \quad \sum_{j \in \hat{n}^{(i)}} \alpha_j^{(i)} \hat{y}_j^{(i)} \exp\left(-\gamma \|\hat{X}_j^{(i)} - X\|^2\right).$$

We remark that the proposition holds for any choice of Ω_C that satisfies our assumptions, irrespective of its norm. That is, besides the indicator function Ω_C satisfying $\Omega_C(\zeta) = 0$ for $\zeta \leq 0$ and $\zeta \geq C$ and $\Omega_C(\zeta) = 1$ for all $\zeta \in (0, C)$, which can be an example, we can also select a scenario, where the value of Ω_C on the interval $(0, C)$ is arbitrarily large.

When $C > 0$ and assuming that there is at least a $j \in [\hat{n}^{(i)}]$ with $\alpha_j^{(i)} \in (0, C)$, the second term in the upper-level objective function $\mathcal{F}$ coincides with the bias $b^{(i)}$. It is obtained starting from the following formula (cf., for instance, Bishop and Nasrabadi (2006)):

$$b^{(i)} = \hat{y}_j^{(i)} - \left(\omega^{(i)}\right)^\top \phi_\gamma\left(\hat{X}_j^{(i)}\right), \quad \forall i \in [k], \ \ \forall j \in \left[\hat{n}^{(i)}\right] \ \ \text{s.t.} \ \ \alpha_j^{(i)} \in (0, C),$$

and re-expressing it w.r.t. the dual variables (see the proof of Proposition 2 in the appendix for more details), which results in:

$$b^{(i)} = \hat{y}_j^{(i)} - \sum_{t \in [\hat{n}^{(i)}]} \alpha_t^{(i)} \hat{y}_t^{(i)} \exp\left(-\gamma \left\|\hat{X}_t^{(i)} - \hat{X}_j^{(i)}\right\|^2\right), \tag{9}$$

for all $i \in [k]$ and for *any* $j \in \left[\hat{n}^{(i)}\right]$ such that $\alpha_j^{(i)} \in (0, C)$. The "any" part of the last statement requires the introduction of an indicator function which not only identifies, for each $i \in [k]$, all the indices $j \in \hat{n}^{(i)}$ where the condition $\alpha_j^{(i)} \in (0, C)$ is satisfied, but also selects one (and one only) of them. To circumvent the need for arbitrarily choosing an index $j \in \hat{n}^{(i)}$ among those satisfying $\alpha_j^{(i)} \in (0, C)$ and, at the same time, obtain a formulation that is numerically stable, we optimize for defining (w.l.o.g.) $b^{(i)}$ as the weighted average of the right-hand sides of the previous expression over all the $\alpha_j^{(i)}$ that satisfy the condition $\alpha_j^{(i)} \in (0, C)$. Letting Ω_C be the indicator function $\Omega_C : [0, C] \to \{0, 1\}$ with $\Omega_C(x) = 1$ if $0 < x < C$ and $\Omega_C(x) = 0$ otherwise, we have that, for each $i \in [k]$ the ratio $1/\sum_{l \in [\hat{n}^{(i)}]} \Omega_C\left(\alpha_\ell^i\right)$ coincides with the number of such variables belonging to $(0, C)$. Thus, the expression for $b^{(i)}$ used in $\mathcal{F}$ coincides the average of the right-hand sides of equation (9). This avoids the need for selecting a suitable index while also guaranteeing a numerically more stable formulation due to averaging the value that one would calculate for $b^{(i)}$ over all the possible indices $j \in \hat{n}^{(i)}$ that could be used for it.

If $C > 0$, and there is some $i \in [k]$ such that $\alpha_j^{(i)} \in \{0, C\}$ for all $j \in [\hat{n}^{(i)}]$, we use the following way to determine $b^{(i)}$. To proceed, we introduce the sets

$$\mathbb{I}_=(\alpha^{(i)}) := \left\{j \in [\hat{n}^{(i)}] \mid \alpha_j^{(i)} = 0\right\} \quad \text{and} \quad \mathbb{I}_<(\alpha^{(i)}) := \left\{j \in [\hat{n}^{(i)}] \mid \alpha_j^{(i)} = C\right\}. \tag{10}$$

Proposition 3. *Let $\alpha^{(i)}$ be the optimal solution of* (LLP^i) *for $i \in [k]$. If $\alpha_j^{(i)} \in \{0, C\}$ for all $j \in [\hat{n}^{(j)}]$, for some lower-level problem i, then the optimal value of $b^{(i)}$ for lower-level problem* (LLP^i) *can be any value in $[b_{min}^{(i)}, b_{max}^{(i)}]$, where*

$$b_{min}^{(i)} := \max_{j \in \mathbb{I}_=(\alpha^{(i)})} \left\{\hat{y}_j^{(i)} - \mathbb{H}^{(i)}(\hat{X}_j^{(i)}, \gamma)\right\}, \ \ b_{max}^{(i)} := \min_{j \in \mathbb{I}_<(\alpha^{(i)})} \left\{\hat{y}_j^{(i)} - \mathbb{H}^{(i)}(\hat{X}_j^{(i)}, \gamma)\right\}. \tag{11}$$

Based on the above result and problem (LLP^i) for each split $i \in [k]$, the bilevel hyperparameter optimization problem for a nonlinear SVM with RBF kernel can be replaced by the following problem, which will be our focus in the remainder of the paper and to which we refer by the shorthand BHO as in Bilevel Hyperparameter Optimization (problem):

$$\begin{aligned} \min_{C, \gamma, \alpha} \quad & \mathcal{F}(C, \gamma, \alpha) \\ \text{s.t.} \quad & (C, \gamma, \alpha) \in \mathbb{R}_+^2 \times \mathbb{R}^n, \ \ \alpha^{(i)} \in S_D^{(i)}(C, \gamma) \quad \text{for} \quad i \in [k], \end{aligned} \tag{BHO}$$

where $n := \sum_{i=1}^{k} \hat{n}^{(i)}$ and, for each split $i \in [k]$, we have

$$S_D^{(i)}(C,\gamma) \quad := \quad \underset{\alpha^{(i)} \in \mathbb{R}^{\hat{n}^{(i)}}}{\operatorname{argmin}} \left\{ \mathcal{H}^{(i)}(\gamma, \alpha^{(i)}) : \ \alpha^{(i)} \in [0, C]^{\hat{n}^{(i)}}, \ (\alpha^{(i)})^\top \hat{y}^{(i)} = 0 \right\}.$$

Remark 4. *It is important to note that both C and γ are likely to be positive for any global solution in the context of any practically relevant data set. This is because the occurrence of data sets where solutions with either $C = 0$ or $\gamma = 0$ are optimal is extremely unlikely. Indeed, the existence of an optimal solution with $C = 0$ would imply that entirely ignoring the misclassification error leads to a better out-of-sample loss than doing otherwise, which is rather unlikely for realistic data sets. Similarly, the adoption of $\gamma = 0$ would lead to an all-one kernel matrix. Thus, the existence of an optimal solution with $\gamma = 0$ would imply that adopting a trivial feature-space map thanks to which the distance between every pair of data points is identical (which, incidentally, makes the data points completely indistinguishable) leads to best out-of-sample loss, which is absurd.*

Based on this remark, we would like to highlight our basic setting in terms of data sets under consideration. Clearly, as we assume that the data sets involved in the model described here are based on real-world scenarios, it is very unlikely that all the data points can lie on the separating hyperplane and all be support vectors. Hence, the following basic settings will be used throughout the paper for data when trained by our nonlinear SVM:

Basic Settings.

(i) *For each lower-level problem $i \in [k]$, there exists at least one training data with label* 1*, which is not a support vector in the kernel space, and there also exists at least one training data with label* -1*, which is not a support vector in kernel space.*

(ii) *For each lower-level problem $i \in [k]$, there exists at least two training data with label* 1*, which are support vectors in kernel space, and there also exists at least two training data with label* -1*, which are support vectors in kernel space.*

In the subsequent sections, we will build a single-level reformulation of (BHO) and study its theoretical properties.

3 MPEC reformulation

3.1 Single-level reformulation

As it is common in the bilevel optimization literature, the first step in the process of developing a theoretical framework or numerical solution scheme for a bilevel program is to reformulate it into a single-level optimization problem. There are three standard approaches to do so: the implicit function approach, the optimal value (function) approach, and the Karush-Kuhn-Tucker (KKT) reformulation approach. We refer the reader to Dempe and Zemkoho (2013) and the references therein for a detailed discussion.

A common point between the implicit function and optimal value function reformulations is that they are both based on implicitly-defined functions, i.e., the lower-level optimal solution function for the former and the optimal value function for the latter. This means that, for each of these two reformulations, we typically might not be able to have a completely explicit expression of the problem. Hence, our focus in this paper is on applying the KKT reformulation to KKT reformulation to (BHO). However, throughout, we label it as mathematical program with equilibrium (MPEC) reformulation, to easily match the vocabulary of the necessary concepts to the existing literature.

To introduce such a reformulation, we start by noting that, for each split $i \in [k]$ and any $\gamma \geq 0$, the kernel matrix $Q^i(\gamma)$ with $Q^i(\gamma)_{rs}$ defined in (8) is positive semidefinite due to being a Gram matrix. Since the lower-level constraints of problem (BHO) w.r.t. $\alpha^{(i)}$ are convex (and, in particular, linear), it follows that each lower-level problem is convex. This implies that, for each split $i \in [k]$, $\alpha^{(i)} \in S_D^{(i)}(C, \gamma)$ if and only if there exist Lagrange multipliers $\varepsilon \in \mathbb{R}^{\hat{n}^{(i)}}$, $\sigma^{(i)} \in \mathbb{R}^{\hat{n}^{(i)}}$, and $u_i \in \mathbb{R}$ corresponding, respectively, to the lower bound, upper bound, and equality constraints such that the following system is feasible:

$$Q^i(\gamma)\alpha^{(i)} - \mathbf{e}^{(i)} - \varepsilon^{(i)} + \sigma^{(i)} - u_i\hat{y}^{(i)} = 0, \tag{12a}$$

$$0 \leq \alpha^{(i)} \perp \varepsilon^{(i)} \geq 0, \tag{12b}$$

$$0 \leq \sigma^{(i)} \perp C\mathbf{e}^{(i)} - \alpha^{(i)} \geq 0, \tag{12c}$$

$${\alpha^{(i)}}^\top \hat{y}^{(i)} = 0. \tag{12d}$$

By defining α, σ, and u as follows

$$\alpha := \begin{bmatrix} \alpha^{(1)} \\ \vdots \\ \alpha^{(k)} \end{bmatrix}, \ \varepsilon := \begin{bmatrix} \varepsilon^{(1)} \\ \vdots \\ \varepsilon^{(k)} \end{bmatrix}, \ \sigma := \begin{bmatrix} \sigma^{(1)} \\ \vdots \\ \sigma^{(k)} \end{bmatrix}, \text{ and } u := \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix}, \tag{13}$$

we deduce the following KKT reformulation of problem (BHO), which we refer to as the "*preliminary* mathematical programming with equilibrium constraints" reformulation:

$$\begin{array}{lll} \min\limits_{C,\gamma,\alpha,\varepsilon,\sigma,u} & \mathcal{F}(C,\gamma,\alpha) & \\ \text{s.t.} & C \geq 0, \gamma \geq 0, & \\ & Q^i(\gamma)\alpha^{(i)} - \mathbf{e}^{(i)} - \varepsilon^{(i)} + \sigma^{(i)} - u_i\hat{y}^{(i)} = 0, & i \in [k], \\ & 0 \leq \alpha^{(i)} \perp \varepsilon^{(i)} \geq 0, & i \in [k], \\ & 0 \leq \sigma^{(i)} \perp C\mathbf{e}^{(i)} - \alpha^{(i)} \geq 0, & i \in [k], \\ & {\alpha^{(i)}}^\top \hat{y}^{(i)} = 0, & i \in [k]. \end{array} \tag{pMPEC-BHO}$$

Let us note that the reformulation we applied is standard for deriving single-level model for bilevel hyperparameter optimization problems; see, e.g., Kunapuli et al. (2008a); Li et al. (2022a,b) and references therein. However, unlike in (Kunapuli et al., 2008a, Section 5) and (Kunapuli et al., 2008b, Section 3), where the MPEC reformulation for a bilevel hyperparameter optimization problem for a nonlinear SVM is obtained directly from the primal lower-level problem, to allow for an infinite-dimensional kernel we proceed here from the dual of the lower-level training problem (LLP^i) (defined for each split $i \in [k]$), which, as mentioned, makes the analysis substantially harder.

3.2 Relationship between (BHO) and (pMPEC-BHO)

To establish the relationship between problems (BHO) and (pMPEC-BHO), recall the index sets in (10). Note that the set $[\hat{n}^{(i)}] \backslash \mathbb{I}_=$ coincides with the index set of the *support-vectors* of the i-th lower-level problem (which are called this way as their dual variable $\alpha_j^{(i)}$ directly contributes to the construction of the equation of the discriminating hyperplane). Basic duality theory shows that: (i) every data point with index $j \in \mathbb{I}_=$ is correctly classified and ends up on or outside the separation margin; (ii) every data point with index $j \in [\hat{n}^{(i)}] \backslash (\mathbb{I}_= \cup \mathbb{I}_<)$ is correctly classified and ends up on the separation margin; (iii) every data point with index $j \in \mathbb{I}_<$ may either be misclassified (while being either inside or outside of the margin) or correctly classified but being within the margin. The two sets will be needed for some important constraint qualifications, including the lower-level constant rank constraint qualification (LCRCQ) for problem (LLP^i) that we introduce first.

Definition 5. *For a split $i \in [k]$, a point $(\bar{C}, \bar{\gamma}, \bar{\alpha}^{(i)})$ satisfies the lower-level constant rank constraint qualification (LCRCQ^i) of problem* (LLP^i) *if there exists an open neighborhood $\mathcal{N}^{(i)}$ of $(\bar{C}, \bar{\gamma}, \bar{\alpha}^{(i)})$ such that, for every pair of index sets I_1, I_2 with $I_1 \subseteq \mathbb{I}_=(\bar{\alpha}^{(i)})$ and $I_2 \subseteq \mathbb{I}_<(\bar{\alpha}^{(i)})$, the family of gradient vectors*

$$\left\{ e_j^{(i)} \mid j \in I_1 \right\} \cup \left\{ -e_j^{(i)} \mid j \in I_2 \right\} \cup \left\{ y^{(i)} \right\} \tag{14}$$

has the same rank (depending on I_1, I_2) for all $(C, \gamma, \alpha^{(i)}) \in \mathcal{N}^{(i)}$. The LCRCQ will be said to hold at $(\bar{C}, \bar{\gamma}, \bar{\alpha})$ if the LCRCQ^i holds at $(\bar{C}, \bar{\gamma}, \bar{\alpha}^{(i)})$ for each $i \in [k]$.

Crucially, we can show the following:

Proposition 6. *The LCRCQ holds at every feasible point $\bar{v}$ of* (pMPEC-BHO)*.*

Next, we introduce the set of Lagrange multipliers for the lower-level problem. For each lower-level problem $i \in [k]$ and $\alpha^{(i)} \in S_D^{(i)}(C, \gamma)$, let $\Lambda^i(C, \gamma, \alpha^{(i)})$ be the set of Lagrange multipliers $(\varepsilon^{(i)}, \sigma^{(i)}, u_i)$ of problem (LLP^i) satisfying (12). Subsequently, let $\Lambda(C, \gamma, \alpha) := \times_{i=1}^k \Lambda^i(C, \gamma, \alpha^{(i)})$, where α is defined as in (13). Based on these definitions, we can establish the following relationship between problems (BHO) and (pMPEC-BHO):

Theorem 7. *The following assertions are satisfied:*

(i) *Let (C, γ, α) be a global (resp., local) optimal solution of problem* (BHO)*. Then, for each $(\varepsilon, \sigma, u) \in \Lambda(C, \gamma, \alpha)$, the point $(C, \gamma, \alpha, \varepsilon, \sigma, u)$ is a global (resp., local) optimal solution of problem* (pMPEC-BHO)*.*

(ii) *Conversely, let $(C, \gamma, \alpha, \varepsilon, \sigma, u)$ be a global optimal solution (resp., local optimal solution for* all vertices *$(\varepsilon, \sigma, u) \in \Lambda(C, \gamma, \alpha)$) of problem* (pMPEC-BHO)*. Then, (C, γ, α) is a global (resp., local) optimal solution of problem* (BHO)*.*

Proof The proof can straightforwardly be derived from Dempe and Dutta (2012). ■

The first observation we can make from this result is that, from a global optimal solution perspective, problems (BHO) and (pMPEC-BHO) are globally equivalent without any assumptions. However, from a local optimal solution point of view, the tricky aspect is ensuring that a local optimal solution $(C, \gamma, \alpha, \varepsilon, \sigma, u)$ of problem (pMPEC-BHO) leads to a point (C, γ, α) which is locally optimal for problem (BHO). For the latter to happen, we need the following assumption included in part (ii) of Theorem 7.

Assumption 1. *For all vertices $(\varepsilon, \sigma, u) \in \Lambda(C, \gamma, \alpha)$, the point $(C, \gamma, \alpha, \varepsilon, \sigma, u)$ is a local optimal solution of problem* (pMPEC-BHO)*.*

This is crucial, as our analysis in this paper is essentially focused around local optimal solutions for problem (pMPEC-BHO). So, it is important to know whether such a local optimal solution has a chance to lead to a local optimal solution of our original problem (BHO). Clearly, based on Assumption 1, a critical thing to look at is whether there is a way to reduce the number of vertices $(\varepsilon, \sigma, u) \in \Lambda(C, \gamma, \alpha)$ that need to be checked. Note that that $\Lambda(C, \gamma, \alpha)$ will reduce to a singleton (ε, σ, u) defined as in (13) if the following lower-level linear independence constraint qualification is satisfied:

Definition 8. *For $i \in [k]$, the lower-level linear independence constraint qualification (LLICQ^i) for problem* (LLP^i) *is satisfied at the point $(C, \gamma, \alpha^{(i)})$ if the following family of vectors is linearly independent:*

$$\left\{\hat{y}^{(i)}\right\} \cup \left\{e_j^{(i)} \mid j \in \mathbb{I}_=(\alpha^{(i)})\right\} \cup \left\{-e_j \mid j \in \mathbb{I}_<(\alpha^{(i)})\right\}.$$

Here, $e_j^{(i)}$ is the j^{th} column of the identity matrix $\mathcal{I} \in \mathbb{R}^{\hat{n}^{(i)} \times \hat{n}^{(i)}}$. The LLICQ will be said to hold at (C, γ, α) with α defined as in (13) *if the LLICQ^i holds at $(C, \gamma, \alpha^{(i)})$ for all lower-level problems* (LLP^i) *of index $i \in [k]$.*

Proposition 9. *Let $C \geq 0$. If $C = 0$, then the LLICQ^i fails at any $(C, \gamma, \alpha^{(i)})$ that satisfies the constraints of the lower-level problem* (LLP^i) *for any $i \in [k]$. Otherwise (i.e., if $C > 0$), for any $(C, \gamma, \alpha^{(i)})$ that satisfies the constraints of* (LLP^i) *for any $i \in [k]$, it holds that:*

(i) If $\mathbb{I}_=(\alpha^{(i)}) = [\hat{n}^{(i)}]$, the LLICQ^i fails at $(C, \gamma, \alpha^{(i)})$.

(ii) If $\mathbb{I}_=(\alpha^{(i)}) \cup \mathbb{I}_<(\alpha^{(i)}) = [\hat{n}^{(i)}]$, the LLICQ^i fails at $(C, \gamma, \alpha^{(i)})$.

(iii) If $\mathbb{I}_=(\alpha^{(i)}) \cup \mathbb{I}_<(\alpha^{(i)}) \neq [\hat{n}^{(i)}]$, the LLICQ^i holds at $(C, \gamma, \alpha^{(i)})$.

The proposition implies that if the condition

$$\mathbb{I}_=(\alpha^{(i)}) \cup \mathbb{I}_<(\alpha^{(i)}) \neq [\hat{n}^{(i)}] \ \text{ for all } \ i \in [k] \tag{15}$$

holds for (C, γ, α) such that $\alpha^{(i)} \in S_D^{(i)}(C, \gamma)$ for all $i \in [k]$, then $\Lambda(C, \gamma, \alpha) = \{(\varepsilon, \sigma, u)\}$, and problems (BHO) and (pMPEC-BHO) are globally and locally equivalent.

As mentioned before in Remark 4, under the Basic Settings, the case $C = 0$ is unlikely to hold in any (globally) optimal solution to (3).

3.3 MPEC preliminaries

By eliminating $\varepsilon^{(i)}$ from (12a), we obtain

$$\varepsilon^{(i)} = Q^i(\gamma)\alpha^{(i)} - \mathbf{e}^{(i)} + \sigma^{(i)} + u_i \hat{y}^{(i)} := \theta^{(i)}(v) \ \text{ for } \ i \in [k].$$

Substituting it into (12b), we obtain the following reduced form of problem (pMPEC-BHO):

$$\begin{aligned} \min_{v \in \mathbb{R}^m} \quad & f(v) := \mathcal{F}(C, \gamma, \alpha) \\ \text{s.t.} \quad & g(v) \leq 0, \ \ h(v) = 0, \\ & G(v) \geq 0, \ \ H(v) \geq 0, \ \ G(v)^\top H(v) = 0. \end{aligned} \tag{MPEC-BHO}$$

Here, the variable $v \in \mathbb{R}^m$, while the functions $g : \mathbb{R} \to \mathbb{R}$, $h : \mathbb{R}^m \to \mathbb{R}^k$, $G : \mathbb{R}^m \to \mathbb{R}^r$, and $H : \mathbb{R}^m \to \mathbb{R}^r$ are respectively given by

$$v := \begin{bmatrix} C \\ \gamma \\ \alpha \\ \sigma \\ u \end{bmatrix}, \; g(v) := -\gamma, \; h(v) := \widehat{Y}\alpha, \; \text{ and } \; \begin{cases} G(v) := \begin{bmatrix} G^1(v) \\ G^2(v) \end{bmatrix}, \\ H(v) := \begin{bmatrix} H^1(v) \\ H^2(v) \end{bmatrix}, \end{cases} \tag{16}$$

where

$$G^1(v) := \theta(v) = \begin{bmatrix} \theta^{(1)}(v) \\ \vdots \\ \theta^{(k)}(v) \end{bmatrix}, \; G^2(v) := C\mathbf{e} - \alpha, \; H^1(v) := \alpha, \; H^2(v) := \sigma,$$

$$\mathbf{e} := \begin{bmatrix} \mathbf{e}^{(1)} \\ \vdots \\ \mathbf{e}^{(k)} \end{bmatrix}, \quad \widehat{Y} := \begin{bmatrix} (\hat{y}^1)^\top \\ \vdots \\ (\hat{y}^k)^\top \end{bmatrix}, \quad \text{and } \; \hat{y}^i := \begin{bmatrix} 0_{\hat{n}^{(1)}} \\ \vdots \\ 0_{\hat{n}^{(i-1)}} \\ \hat{y}^{(i)} \\ 0_{\hat{n}^{(i+1)}} \\ \vdots \\ 0_{\hat{n}^{(k)}} \end{bmatrix} \; \text{for } i \in [k]. \tag{17}$$

Above, $m := 2+k+2n$, $r := 2n$, and, as before, where $n = \sum_{i=1}^{k} \hat{n}^{(i)}$. Compared to problem (BHO), it is worth noting that in problem (MPEC-BHO), the nonnegativity of C, which is explicit in (BHO), is implied by $C\mathbf{e} - \alpha \geq 0$ and $\alpha \geq 0$ and, thus, is not imposed explicitly.

In the remainder of this section, we introduce some basic MPEC theoretical concepts on which we will focus our attention in the subsequent sections. Start by recalling that for a given optimization problem, concepts such as constraint qualifications, and necessary and sufficient optimality conditions are not only crucial for their theoretical analysis, but also for the development of various numerical methods. This is no different in the context of problem (MPEC-BHO). However, for this problem in MPEC form, standard constraint qualification such as the Mangasarian–Fromovitz constraint qualification (MFCQ) is known to automatically fail for any of its feasible points; see, e.g., Flegel (2005); Ye et al. (1997). Hence, to address this issue, specifically tailored constraint qualifications have been introduced in the literature to derive optimality conditions and other relevant properties.

To describe some of these specific types of constraint qualifications, we now introduce the following decomposition of the index sets involved in the complementarity constraints featured in the feasible set of problem (MPEC-BHO). Letting $\bar{v} \in \mathbb{R}^m$ be a feasible point of problem (MPEC-BHO), for the inequalities defined by G and H we define:

$$\begin{aligned} I_{0+}(\bar{v}) &:= \{i \in [r] : \; G_i(\bar{v}) = 0, \;\; H_i(\bar{v}) > 0\}, \\ I_{+0}(\bar{v}) &:= \{i \in [r] : \; G_i(\bar{v}) > 0, \;\; H_i(\bar{v}) = 0\}, \\ I_{00}(\bar{v}) &:= \{i \in [r] : \; G_i(\bar{v}) = 0, \;\; H_i(\bar{v}) = 0\}. \end{aligned} \tag{18}$$

Considering the inequality defined by g, we define the set of active indices as

$$I_g(\bar{v}) := \{i \in [1] : \; g_i(\bar{v}) = 0\}. \tag{19}$$

For the sake of notation, in the sequel instead of $I_{0+}(\bar{v})$ we will simply write I_{0+}, and proceed similarly for $I_{+0}(\bar{v})$, $I_{00}(\bar{v})$, and $I_g(\bar{v})$.

Next, we introduce tractable versions of the Mangasarian-Fromovitz and linear independence constraint qualifications tailored to (MPEC-BHO).

Definition 10. *Let $\bar{v} \in \mathbb{R}^m$ be a feasible point of problem* (MPEC-BHO) *and consider the following famility of vectors:*

$$\begin{aligned} \{\nabla G_i(\bar{v}):\ i \in I_{0+} \cup I_{00}\} \cup \{\nabla H_i(\bar{v}):\ i \in I_{+0} \cup I_{00}\} \\ \cup \{\nabla g_i(\bar{v}):\ i \in I_g\} \cup \{\nabla h_i(\bar{v}):\ i \in [k]\}. \end{aligned} \tag{20}$$

The point $\bar{v}$ will be said to satisfy:

(a) *the* MPEC Mangasarian-Fromovitz constraint qualification *(MPEC-MFCQ) if the family of vectors in* (20) *is positively linearly independent;*

(b) *the* MPEC linear independence constraint qualification *(MPEC-LICQ) if the family of vectors in* (20) *is linearly independent.*

We remark that one can easily verify that, if a point satisfies the MPEC-LICQ, then it automatically satisfies the MPEC-MFCQ as well.

The literature offers a multitude of stationarity concepts which are suitably defined to handle problems with complementarity conditions among their constraints and that can be seen as analogues of the KKT conditions for problems that do not feature constraints of MPEC type. Here, we introduce the three main stationarity concepts for (MPEC-BHO). To do so, we will rely on the following Lagrangian function of (MPEC-BHO) and defined for the point $v \in \mathbb{R}^m$ and the Lagrange multipliers $\lambda \in \mathbb{R}$, $\mu \in \mathbb{R}^k$, and $\eta, \zeta \in \mathbb{R}^r$:

$$\mathbb{L}(v, \lambda, \mu, \eta, \zeta) := f(v) + \lambda g(v) + \mu^\top h(v) - \eta^\top G(v) - \zeta^\top H(v).$$

Definition 11. *A feasible point $\bar{v} \in \mathbb{R}^m$ of problem* (MPEC-BHO) *will be said to be*

(a) strongly stationary (S-stationary) *if there exist Lagrange multipliers $\bar{\lambda} \in \mathbb{R}$, $\bar{\mu} \in \mathbb{R}^k$, $\bar{\eta} \in \mathbb{R}^r$, and $\bar{\zeta} \in \mathbb{R}^r$ such that*

$$\nabla_v \mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}) = 0, \tag{21}$$

$$\forall i \in I_g:\ \ \bar{\lambda}_i \geq 0,\ \ \forall i \in \{1\} \setminus I_g:\ \ \bar{\lambda}_i = 0, \tag{22}$$

$$\forall j \in I_{+0}:\ \ \bar{\eta}_j = 0,\ \ \forall j \in I_{0+}:\ \ \bar{\zeta}_j = 0, \tag{23}$$

$$\forall j \in I_{00}:\ \ \bar{\eta}_j \leq 0,\ \ \bar{\zeta}_j \leq 0, \tag{24}$$

(b) Mordukhovich stationary (M-stationary) *if we can find Lagrange multipliers $\bar{\lambda} \in \mathbb{R}$, $\bar{\mu} \in \mathbb{R}^k$, $\bar{\eta} \in \mathbb{R}^r$, and $\bar{\zeta} \in \mathbb{R}^r$ such that* (21)–(23) *and*

$$\forall j \in I_{00}:\ \ \left(\bar{\eta}_j \bar{\zeta}_j = 0\right)\ \ \vee\ \ \left(\bar{\eta}_j < 0,\ \ \bar{\zeta}_j < 0\right), \tag{25}$$

(c) Clarke stationary (C-stationary) *if we can find Lagrange Multipliers $\bar{\lambda} \in \mathbb{R}$, $\bar{\mu} \in \mathbb{R}^k$, $\bar{\eta} \in \mathbb{R}^r$, and $\bar{\zeta} \in \mathbb{R}^r$ such that* (21)–(23) *hold together with*

$$\forall j \in I_{00}:\ \ \bar{\eta}_j \bar{\zeta}_j \geq 0. \tag{26}$$

It is well-known that if a local optimal solution $\bar{v}$ to an MPEC satisfies the MPEC-MFCQ, then this point is M-stationary. However, if the stronger MPEC-LICQ holds at $\bar{v}$, then this point is S-stationary. A point that is S-stationary is also M-stationary, and the latter implies that C-stationarity holds.

It is also important to note that the S-stationarity concept is equivalent to the KKT conditions of problem (MPEC-BHO) seen from the perspective of a standard optimization problem with $G(v)^\top H(v) = 0$ treated as a usual equality constraint; see, e.g., Flegel (2005). With this in mind, it clearly makes sense to think that an S-stationary point that satisfies some strong second order conditions would be locally optimal for problem (MPEC-BHO). This is indeed the case under the following MPEC-tailored strong second order sufficient condition (MPEC-SSOSC):

Definition 12. *Let $\bar{v}$ be a S-stationary point of problem* (MPEC-BHO) *with multiplier vector $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$ is said to satisfy the MPEC-SSOSC if it holds that*

$$\forall d \in \mathcal{C}(\bar{v}) \setminus \{0\} : \quad d^\top \nabla^2_{vv} \mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}) d > 0 \tag{27}$$

with $\mathcal{C}(\bar{v})$ defined as follows (*with* $\operatorname{supp}(a) := \{i \in [n] : \ a_i \neq 0\}$ *for* $a \in \mathbb{R}^n$) :

$$\mathcal{C}(\bar{v}) := \left\{ d \in \mathbb{R}^m : \begin{array}{rl} \nabla g_i(\bar{v})^\top d = 0 & \forall i \in \operatorname{supp}(\bar{\lambda}) \\ \nabla h_i(\bar{v})^\top d = 0 & \forall i \in [k] \\ \nabla G_i(\bar{v})^\top d = 0 & \forall i \in \operatorname{supp}(\bar{\eta}) \\ \nabla H_i(\bar{v})^\top d = 0 & \forall i \in \operatorname{supp}(\bar{\zeta}) \end{array} \right\}. \tag{28}$$

Note, for example, that if $\bar{v}$ is a S-stationary point of (MPEC-BHO) that satisfies the MPEC-SSOSC together with the MPEC-LICQ, then there exists a neighborhood $\mathcal{U}(\bar{v})$ of $\bar{v}$ such that $\bar{v}$ is the only M-stationary point among all the feasible points of (MPEC-BHO) in $\mathcal{U}(\bar{v})$; see, e.g., (Kanzow and Schwartz, 2013, Theorem 4.11) and references therein.

4 MPEC–MFCQ

In this section, we analyze the MPEC-MFCQ in the context of (MPEC-BHO). To proceed, we introduce the following index sets associated to the complementarity conditions involved in the feasible of problem (MPEC-BHO), which will play an important role in the analysis.

Definition 13. *Letting $\bar{v}$ be a feasible point of* (MPEC-BHO), *we define the index sets*

$$\begin{aligned} \Lambda_1(\bar{v}) &:= \{i \in [n] : \ \bar{\alpha}_i = 0, \ \ \theta(\bar{v})_i = 0, \ \ \bar{\sigma}_i = 0\}, & (29) \\ \Lambda_2(\bar{v}) &:= \{i \in [n] : \ \bar{\alpha}_i = 0, \ \ \theta(\bar{v})_i > 0, \ \ \bar{\sigma}_i = 0\}, & (30) \\ \Lambda_3(\bar{v}) &:= \{i \in [n] : \ 0 < \bar{\alpha}_i \leq \bar{C}, \ \ \theta(\bar{v})_i = 0, \ \ \bar{\sigma}_i = 0\}, & (31) \\ \Lambda_4(\bar{v}) &:= \{i \in [n] : \ \bar{\alpha}_i = \bar{C}, \ \ \theta(\bar{v})_i = 0, \ \ \bar{\sigma}_i > 0\}. & (32) \end{aligned}$$

$\Lambda_3(\bar{v})$ is further partitioned as follows:

$$\begin{aligned} \Lambda_3^+(\bar{v}) &:= \{i \in [n] : 0 < \bar{\alpha}_i < \bar{C}, \ \ \theta(\bar{v})_i = 0, \ \ \bar{\sigma}_i = 0\}, & (33) \\ \Lambda_3^c(\bar{v}) &:= \{i \in [n] : \bar{\alpha}_i = \bar{C}, \ \ \theta(\bar{v})_i = 0, \ \ \bar{\sigma}_i = 0\}. & (34) \end{aligned}$$

Definition 14. *Let $\bar{v}$ be a feasible point of* (MPEC-BHO)*. For each lower-level problem* (LLP^i) *for $i \in [k]$, $\Lambda_3^{+(i)}(\bar{v})$ and $\Lambda_3^{c(i)}(\bar{v})$ correspond to the sets $\Lambda_3^+(\bar{v})$ and $\Lambda_3^c(\bar{v})$, respectively, and for $r \in [4]$, $\Lambda_r^{(i)}(\bar{v})$ denotes the set corresponding to $\Lambda_r(\bar{v})$.*

Definition 15. *We define the following partition of the index sets associated to the complementarity conditions involved in the feasible set of problem* (MPEC-BHO)*:*

$$
\begin{aligned}
J_{H^1}(\bar{v}) &:= \{i \in [n] \; : \; \bar{\alpha}_i = 0, \;\; \theta(\bar{v})_i > 0\}, && \text{(35a)}\\
J_{H^2}(\bar{v}) &:= \{i \in [n] \; : \; \bar{\sigma}_i = 0, \;\; \bar{C} - \bar{\alpha}_i > 0\}, && \text{(35b)}\\
J_{G^1}(\bar{v}) &:= \{i \in [n] \; : \; \bar{\alpha}_i > 0, \;\; \theta(\bar{v})_i = 0\}, && \text{(35c)}\\
J_{G^2}(\bar{v}) &:= \{i \in [n] \; : \; \bar{\sigma}_i > 0, \;\; \bar{C} - \bar{\alpha}_i = 0\}, && \text{(35d)}\\
J_{GH^1}(\bar{v}) &:= \{i \in [n] \; : \; \bar{\alpha}_i = 0, \;\; \theta(\bar{v})_i = 0\}, && \text{(35e)}\\
J_{GH^2}(\bar{v}) &:= \{i \in [n] \; : \; \bar{\sigma}_i = 0, \;\; \bar{C} - \bar{\alpha}_i = 0\}. && \text{(35f)}
\end{aligned}
$$

For the ease of notation, in the sequel, we will simply write Λ_i for each $i \in [4]$ rather than $\Lambda_i(\bar{v})$, and proceed similarly for $\Lambda_3^+(\bar{v})$ and $\Lambda_3^c(\bar{v})$, and also for $J_{H^i}(\bar{v})$, $J_{G^i}(\bar{v})$, and $J_{GH^i}(\bar{v})$ for $i = 1, 2$. We observe that we have the equalities $J^i = J_{H^i} \cup J_{G^i} \cup J_{GH^i}$ for $i = 1$, 2, and it is clear that $J^1 = J^2 = [n]$.

Next, we give some relationships between the index sets in (35) and the index sets described in (29)–(32); also see Figure 1 for an illustration.

Proposition 16. *The index sets in* (35) *and* (29) – (32) *satisfy the following relationships:*

(a) $J_{H^1} = \Lambda_2, \;\; J_{G^1} = \Lambda_3 \cup \Lambda_4, \;\; J_{GH^1} = \Lambda_1$*;*

(b) $J_{H^2} = \Lambda_1 \cup \Lambda_2 \cup \Lambda_3^+, \;\; J_{G^2} = \Lambda_4, \;\; J_{GH^2} = \Lambda_3^c$.

The proof is the same as in (Li et al., 2022a, Proposition 5), and is therefore omitted.

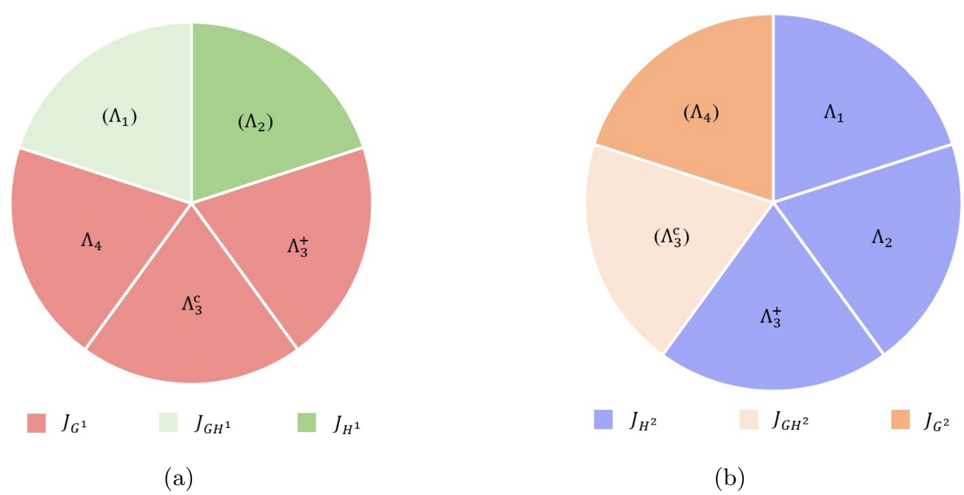


Figure 1: Index sets for the complementarity constraints in Proposition 16.

Due to our Basic Settings (i), we have the following properties.

Proposition 17. *For each lower-level problem $i \in [k]$, let $\alpha^{(i)}$ be the solution. There exists at least an index $j \in [\hat{n}^{(i)}]$ such that*

$$\alpha_j^{(i)} = 0 \text{ and } \hat{y}_j^{(i)} = 1. \tag{36}$$

Similarly, for each lower-level problem $i \in [k]$, we can find at least one $j \in [\hat{n}^{(i)}]$ such that

$$\alpha_j^{(i)} = 0 \text{ and } \hat{y}_j^{(i)} = -1. \tag{37}$$

Proposition 18. (*a*) *The set of gradient vectors in* (20) *at a feasible point $\bar{v}$ of problem* (MPEC-BHO) *can be written in matrix form as follows:*

$$\Gamma := \begin{bmatrix}
\mathbf{0}_{(J_{G^1},L_1)} & \nabla_\gamma\theta(\bar{v})_{J_{G^1}} & Q(\bar{\gamma})_{(J_{G^1},\cdot\,)} & \mathcal{I}_{(J_{G^1},\cdot)} & P_{(J_{G^1},\cdot)} \\
\mathbf{0}_{(J_{GH^1},L_1)} & \nabla_\gamma\theta(\bar{v})_{J_{GH^1}} & Q(\bar{\gamma})_{(J_{GH^1},\cdot\,)} & \mathcal{I}_{(J_{GH^1},\cdot)} & P_{(J_{GH^1},\cdot)} \\
\mathbf{0}_{(J_{H^1},L_1)} & \mathbf{0}_{(J_{H^1},L_2)} & \mathcal{I}_{(J_{H^1},\cdot)} & \mathbf{0}_{(J_{H^1},L_4)} & \mathbf{0}_{(J_{H^1},L_5)} \\
\mathbf{0}_{(J_{GH^1},L_1)} & \mathbf{0}_{(J_{GH^1},L_2)} & \mathcal{I}_{(J_{GH^1},\cdot)} & \mathbf{0}_{(J_{GH^1},L_4)} & \mathbf{0}_{(J_{GH^1},L_5)} \\
\mathbf{e}_{(J_{G^2},L_1)} & \mathbf{0}_{(J_{G^2},L_2)} & -\mathcal{I}_{(J_{G^2},\cdot)} & \mathbf{0}_{(J_{G^2},L_4)} & \mathbf{0}_{(J_{G^2},L_5)} \\
\mathbf{e}_{(J_{GH^2},L_1)} & \mathbf{0}_{(J_{GH^2},L_2)} & -\mathcal{I}_{(J_{GH^2},\cdot)} & \mathbf{0}_{(J_{GH^2},L_4)} & \mathbf{0}_{(J_{GH^2},L_5)} \\
\mathbf{0}_{(J_{H^2},L_1)} & \mathbf{0}_{(J_{H^2},L_2)} & \mathbf{0}_{(J_{H^2},L_3)} & \mathcal{I}_{(J_{H^2},\cdot)} & \mathbf{0}_{(J_{H^2},L_5)} \\
\mathbf{0}_{(J_{GH^2},L_1)} & \mathbf{0}_{(J_{GH^2},L_2)} & \mathbf{0}_{(J_{GH^2},L_3)} & \mathcal{I}_{(J_{GH^2},\cdot)} & \mathbf{0}_{(J_{GH^2},L_5)} \\
\mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \widehat{Y} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \\
\mathbf{0}_{(I_g,L_1)} & -\mathbf{e}_{(I_g,L_2)} & \mathbf{0}_{(I_g,L_3)} & \mathbf{0}_{(I_g,L_4)} & \mathbf{0}_{(I_g,L_5)}
\end{bmatrix}, \tag{38}$$

where L_q, with $q \in [5]$, are the index sets of the columns that correspond, respectively, to the variables C, γ, α, σ, and u; $P = \widehat{Y}^\top$ with $\widehat{Y}$ given in (17)*; and*

$$Q(\gamma) := \begin{bmatrix}
Q^1(\gamma) & \cdots & \mathbf{0}_{\hat{n}^{(1)}\times\hat{n}^{(k)}} \\
\vdots & \ddots & \vdots \\
\mathbf{0}_{\hat{n}^{(k)}\times\hat{n}^{(1)}} & \cdots & Q^k(\gamma)
\end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{39}$$

(*b*) *Thanks to Proposition* 16, Γ *in* (38) *takes the following form:*

$$\Gamma = \begin{bmatrix}
\mathbf{0}_{(\Lambda_3^+,L_1)} & \nabla_\gamma\theta(\bar{v})_{\Lambda_3^+} & Q(\bar{\gamma})_{(\Lambda_3^+,\cdot)} & \mathcal{I}_{(\Lambda_3^+,\cdot)} & P_{(\Lambda_3^+,\cdot)} \\
\mathbf{0}_{(\Lambda_3^c,L_1)} & \nabla_\gamma\theta(\bar{v})_{\Lambda_3^c} & Q(\bar{\gamma})_{(\Lambda_3^c,\cdot)} & \mathcal{I}_{(\Lambda_3^c,\cdot)} & P_{(\Lambda_3^c,\cdot)} \\
\mathbf{0}_{(\Lambda_4,L_1)} & \nabla_{\bar{\gamma}}\theta(\bar{v})_{\Lambda_4} & Q(\gamma)_{(\Lambda_4,\cdot)} & \mathcal{I}_{(\Lambda_4,\cdot)} & P_{(\Lambda_4,\cdot)} \\
\mathbf{0}_{(\Lambda_1,L_1)} & \nabla_\gamma\theta(\bar{v})_{\Lambda_1} & Q(\bar{\gamma})_{(\Lambda_1,\cdot)} & \mathcal{I}_{(\Lambda_1,\cdot)} & P_{(\Lambda_1,\cdot)} \\
\mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_4)} & \mathbf{0}_{(\Lambda_2,L_5)} \\
\mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_4)} & \mathbf{0}_{(\Lambda_1,L_5)} \\
\mathbf{e}_{(\Lambda_4,L_1)} & \mathbf{0}_{(\Lambda_4,L_2)} & -\mathcal{I}_{(\Lambda_4,\cdot)} & \mathbf{0}_{(\Lambda_4,L_4)} & \mathbf{0}_{(\Lambda_4,L_5)} \\
\mathbf{e}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & -\mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_4)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\
\mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathbf{0}_{(\Lambda_1,L_3)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_5)} \\
\mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathbf{0}_{(\Lambda_2,L_3)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_5)} \\
\mathbf{0}_{(\Lambda_3^+,L_1)} & \mathbf{0}_{(\Lambda_3^+,L_2)} & \mathbf{0}_{(\Lambda_3^+,L_3)} & \mathcal{I}_{(\Lambda_3^+,\cdot)} & \mathbf{0}_{(\Lambda_3^+,L_5)} \\
\mathbf{0}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & \mathbf{0}_{(\Lambda_3^c,L_3)} & \mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\
\mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \widehat{Y} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \\
\mathbf{0}_{(I_g,L_1)} & -\mathbf{e}_{(I_g,L_2)} & \mathbf{0}_{(I_g,L_3)} & \mathbf{0}_{(I_g,L_4)} & \mathbf{0}_{(I_g,L_5)}
\end{bmatrix}. \tag{40}$$

Remark 19. *Note that, if some index set such as, e.g., Λ_3^+, is empty, then by construction, the corresponding row block does not appear in* (40).

We are now ready to present our main theorem about the MPEC-MFCQ.

Theorem 20. *The MPEC-MFCQ holds at any point $\bar{v} := (\bar{C}, \bar{\gamma}, \bar{\alpha}, \bar{\sigma}, \bar{u})$ that is feasible to problem* (MPEC-BHO).

Proof Assume there exists a nonnegative vector $\rho = (\rho_1, \cdots, \rho_{10})$ such that $\rho^\top \Gamma = 0$. Here, $\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6,\ \rho_7, \rho_8,\ \rho_9,\ \rho_{10}$ correspond to the blocks of J_{G^1}, J_{GH^1}, J_{H^1}, J_{GH^1}, J_{G^2}, J_{GH^2}, J_{H^2}, J_{GH^2}, $[k]$, I_g in Γ. By (38), we have $0 = \rho^\top \Gamma := [S_1\ \ S_2\ \ S_3\ \ S_4\ \ S_5]$. Hence,

$$
\begin{aligned}
S_1 &= \rho_5^\top \mathbf{e}_{(J_{G^2}, L_1)} + \rho_6^\top \mathbf{e}_{(J_{GH^2}, L_1)} = 0, && (41)\\
S_2 &= \rho_1^\top \nabla_\gamma \theta(\bar{v})_{J_{G^1}} + \rho_2^\top \nabla_\gamma \theta(\bar{v})_{J_{GH^1}} - \rho_{10} \mathbf{e}_{(I_g, L_2)} = 0, && (42)\\
S_3 &= \rho_1^\top (Q(\gamma))_{(J_{G^1}, \cdot)} + \rho_2^\top Q(\bar{\gamma})_{(J_{GH^1}, \cdot)} + \rho_3^\top \mathcal{I}_{(J_{H^1}, \cdot)} + \rho_4^\top \mathcal{I}_{(J_{GH^1}, \cdot)} \\
&\quad - \rho_5^\top \mathcal{I}_{(J_{G^2}, \cdot)} - \rho_6^\top \mathcal{I}_{(J_{GH^2}, \cdot)} + \rho_9^\top \widehat{Y} = 0, && (43)\\
S_4 &= \rho_1^\top \mathcal{I}_{(J_{G^1}, \cdot)} + \rho_2^\top \mathcal{I}_{(J_{GH^1}, \cdot)} + \rho_7^\top \mathcal{I}_{(J_{H^2}\ \cdot)} + \rho_8^\top \mathcal{I}_{(J_{GH^2}, \cdot)} = 0, && (44)\\
S_5 &= \rho_1^\top P_{(J_{G^1}, \cdot)} + \rho_2^\top P_{(J_{GH^1}, \cdot)} = 0. && (45)
\end{aligned}
$$

With the nonnegativity of ρ, (41) implies that $\rho_5 = 0, \rho_6 = 0$. Similarly, (44) implies that $\rho_1 = 0,\ \rho_2 = 0, \rho_7 = 0,\ \rho_8 = 0$. Therefore, by (42), it holds that $\rho_{10} = 0$. (43) reduces to

$$\rho_3^\top \mathcal{I}_{(J_{H^1}, \cdot)} + \rho_4^\top \mathcal{I}_{(J_{GH^1}, \cdot)} + \rho_9^\top \widehat{Y} = 0.$$

By Lemma 39 in the Appendix, we obtain that $\rho_3 = 0,\ \rho_4 = 0$, and $\rho_9 = 0$. ■

The result of Theorem 20 complements what was shown in Li et al. (2022a), i.e., that the MPEC-MFCQ automatically holds for the version of problem (3) where the kernel is linear and the lower-level training problem is formulated in the primal space, and therefore showing that the result holds as well for the case of an RBF kernel with the lower-level problem stated in the dual space.

5 MPEC–LICQ

In this section, we analyze the MPEC-LICQ for each feasible point of problem (MPEC-BHO). The following further observations for the lower-level problem, derived from Basic Settings (ii), will play an important role in the analysis.

Proposition 21. *For each lower-level problem* (LLP^i) *for $i \in [k]$, there are at least two positive support vectors and two negative support vectors; i.e., there exist distinct indices j_1, j_2, j_3, $j_4 \in [\hat{n}^{(i)}]$ such that*

$$\alpha_{j_1}^{(i)} > 0,\ \ \alpha_{j_2}^{(i)} > 0\ \ \textit{with}\ \ \hat{y}_{j_1}^{(i)} = 1,\ \hat{y}_{j_2}^{(i)} = 1$$

and

$$\alpha_{j_3}^{(i)} > 0,\ \ \alpha_{j_4}^{(i)} > 0\ \ \textit{with}\ \ \hat{y}_{j_3}^{(i)} = -1,\ \hat{y}_{j_4}^{(i)} = -1.$$

Thanks to this result, we can establish the first main result of this section.

Theorem 22. *Let $\bar{v}$ be a feasible point of problem* (MPEC-BHO). *Then the MPEC-LICQ fails at $\bar{v}$ if one of the following conditions holds:*

$$| J_{GH^1} | + | J_{GH^2} | > 2 \text{ and } I_g = \emptyset; \tag{46}$$

$$| J_{GH^1} | + | J_{GH^2} | \geq 2 \text{ and } I_g \neq \emptyset. \tag{47}$$

Based on this result, we can see that the behaviour of the MPEC-LICQ is closely related to the number of elements in the index sets I_{GH^1} and I_{GH^2}. Hence, next, we first consider the very special case $\gamma > 0$, and I_{GH^1} and I_{GH^2} are both empty. To this end, we make the following assumptions.

Assumption 2. *$I_g = \emptyset$, $\Lambda_3^+ \neq \emptyset$, and $Q(\gamma)_{(\Lambda_3^+, \Lambda_3^+)}$ positive definite.*

Assumption 3. *For a feasible point $\bar{v}$ of problem* (MPEC-BHO)*, assume that LLICQ^i holds at $\bar{\alpha}^{(i)}$ for each $i \in [k]$. Or equivalently, for each lower-level problem $i \in [k]$ with $\Lambda_3^{+i} \neq \emptyset$, there exists $j \in [\hat{n}^{(i)}]$ satisfying*

$$0 < \bar{\alpha}_j^{(i)} < \bar{C}, \quad (\theta(\bar{v}))_j^{(i)} = 0, \quad \bar{\sigma}_j^{(i)} = 0 \text{ for } i \in [k]. \tag{48}$$

This assumption is the same as the condition ensuring the LLICQ that we have in (15). Indeed, note the relationships

$$\mathbb{I}_=(\alpha^{(i)}) = \Lambda_1^{(i)} \cup \Lambda_2^{(i)} \text{ and } \mathbb{I}_<^C(\alpha^{(i)}) = \Lambda_4^{(i)} \cup \Lambda_3^{c(i)} \text{ for } i \in [k].$$

Condition (15) reduces to $\lambda_3^{+(i)} \neq \emptyset$ for $i \in [k]$, which coincides with (48) in Assumption 3. If Assumption 3 does not hold as the LLICQ^i fails for some lower-level problem $i \in [k]$, we can still show that MPEC-LICQ still holds if LLICQ fails at a single lower-level problem. To show this, let us first define

$$K := \{i \in [k] : \ \Lambda_3^{+(i)} = \emptyset\} \text{ and } K^c := [k] \backslash K. \tag{49}$$

Furthermore, define $s := (s_1, \cdots, s_k)^\top \in \mathbb{R}^k$, where

$$s_i := \sum_{j \in J_{G^2} \cup J_{GH^2}} y_j \quad \text{for } i \in [k].$$

Relying on these definitions, we introduce the following assumption:

Assumption 4. *For a feasible point $\bar{v}$ of problem* (MPEC-BHO)*, assume that there exists a lower-level problem $i \in [k]$ where the LLICQ^i fails. Further assume that*

$$s_i \neq 0 \text{ for } i \in K. \tag{50}$$

Without (50), Assumption 4 would be the opposite part of Assumption 3. Note that the two assumptions cannot simultaneously hold; however, they may fail at the same time.

Theorem 23. *For a feasible point $\bar{v}$ of problem* (MPEC-BHO) *satisfying Assumption* 2 *and* $| J_{GH^1} | + | J_{GH^2} |= 0$*, it holds that:*

(i) *If $J_{G^1} = J_{G^2}$, then the MPEC-LICQ fails at $\bar{v}$;*

(ii) *If $J_{G^2} = \emptyset$, the MPEC-LICQ holds at $\bar{v}$ if and only if Assumption* 3 *holds;*

(iii) *If $J_{G^2} \subset J_{G^1}$ and $J_{G^2} \neq \emptyset$, the MPEC-LICQ holds at $\bar{v}$ if and only if either Assumption* 3 *or Assumption* 4 *hold.*

Note that $| J_{GH^1} | + | J_{GH^2} |= 0$ basically implies that the strict complementarity condition holds in the KKT system of each lower-level problem.

To extend the result to the case where $| J_{GH^1} | + | J_{GH^2} | \neq 0$, we introduce the following notation. Let $A^1 \in \mathbb{R}^{k\times|\Lambda_3^+|}$, $A^2 \in \mathbb{R}^{|\Lambda_3^c|\times|\Lambda_3^+|}$, $A^3 \in \mathbb{R}^{|\Lambda_4|\times|\Lambda_3^+|}$ and $A^4 \in \mathbb{R}^{|\Lambda_1|\times|\Lambda_3^+|}$ be matrices satisfying the following conditions:

$$A^1 Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^+)} = \widehat{Y}_{([k],\Lambda_3^+)}, \quad A^2 Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^+)} = Q(\bar{\gamma})_{(\Lambda_3^c,\Lambda_3^+)}, \tag{51}$$

$$A^3 Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^+)} = Q(\bar{\gamma})_{(\Lambda_4,\Lambda_3^+)}, \quad A^4 Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^+)} = Q(\bar{\gamma})_{(\Lambda_1,\Lambda_3^+)}. \tag{52}$$

Thanks to Assumptions 2 and 3, $Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^+)}$ is positive definite and, therefore, the above matrices A^i for $i = 1, \ldots, 4$ are unique. Next, we introduce the following quantities:

$$\left\{ \begin{array}{lcl} a^1 & := & \Big[Q(\bar{\gamma})_{(\Lambda_3^c,\, \Lambda_4 \cup \Lambda_3^c)} - B^1 \widehat{Y}_{([k],\, \Lambda_4\cup\Lambda_3^c)} \\ & & \qquad\qquad + \left(B^1 A^1 - A^2\right) Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_4\cup\Lambda_3^c)} \Big] \mathbf{e}_{|\Lambda_4\cup\Lambda_3^c|}, \\ a^2 & := & \Big[Q(\bar{\gamma})_{(\Lambda_1,\, \Lambda_4 \cup \Lambda_3^c)} - B^2 \widehat{Y}_{([k],\, \Lambda_4\cup\Lambda_3^c)} \\ & & \qquad\qquad + \left(B^2 A^1 - A^4\right) Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_4\cup\Lambda_3^c)} \Big] \mathbf{e}_{|\Lambda_4\cup\Lambda_3^c|}, \\ b^1 & := & \nabla_\gamma \theta(v)_{\Lambda_3^c} + (B^1 A^1 - A^2) \nabla_\gamma \theta(v)_{\Lambda_3^+}, \\ b^2 & := & \nabla_\gamma \theta(v)_{\Lambda_1} + (B^2 A^1 - A^4) \nabla_\gamma \theta(v)_{\Lambda_3^+}, \\ b^3 & := & b^1 + B^3 \left(A^1 \nabla_\gamma \theta(v)_{\Lambda_3^+} \right)_{K^c}, \\ b^4 & := & b^2 + B^4 \left(A^1 \nabla_\gamma \theta(v)_{\Lambda_3^+} \right)_{K^c}, \\ U^1 & := & P_{(\Lambda_3^c,\, \cdot)} - A^2 P_{(\Lambda_3^+,\, \cdot)}, \\ U^2 & := & P_{(\Lambda_1,\, \cdot)} - A^4 P_{(\Lambda_3^+,\, \cdot)}, \\ Z & := & -A^1 P_{(\Lambda_3^+,\, \cdot)}. \end{array} \right. \tag{53}$$

Let now $B^1 \in \mathbb{R}^{|\Lambda_3^c|\times k}$, $B^2 \in \mathbb{R}^{|\Lambda_1|\times k}$, $B^3 \in \mathbb{R}^{|\Lambda_3^c|\times|K^c|}$, and $B^4 \in \mathbb{R}^{|\Lambda_1|\times|K^c|}$ be matrices satisfying the following conditions:

$$B^1 Z = U^1, \tag{54}$$

$$B^2 Z = U^2, \tag{55}$$

$$B^3 Z_{(K^c,\, K^c)} = U^1_{(\cdot,\, K^c)}, \tag{56}$$

$$B^4 Z_{(K^c,\, K^c)} = U^2_{(\cdot,\, K^c)}. \tag{57}$$

Note that, if Z and $Z_{(K^c,K^c)}$ are nonsingular, B^1, B^2, B^3, and B^4 are uniquely defined.

Theorem 24. *Let $\bar{v}$ be a feasible point of problem* (MPEC-BHO) *satisfying Assumption* 2 *and the condition* $| J_{GH^1} | + | J_{GH^2} | \in \{1, 2\}$.

(i) *If Assumption 3 is satisfied:*

(a) If $| J_{GH^1} | + | J_{GH^2} | = 2$ *holds and the matrix* $M^0 := \begin{bmatrix} a^1 & b^1 \\ a^2 & b^2 \end{bmatrix}$ *is nonsingular, then the MPEC-LICQ holds.*

(b) If $| J_{GH^1} | + | J_{GH^2} | = 1$ *holds:*

i. If $| J_{GH^1} | = 1$ *and* $(a^2, b^2) \neq 0$*, then the MPEC-LICQ holds.*

ii. If $| J_{GH^2} | = 1$ *and* $(a^1, b^1) \neq 0$*, then the MPEC-LICQ holds.*

(ii) *If Assumption* 4 *is satisfied:*

- *If* $| J_{GH^1} | + | J_{GH^2} | = 2$ *and the matrix* $M^1 := \begin{bmatrix} b^3 & U^1_{(\cdot,\, K)} \\ b^4 & U^2_{(\cdot,\, K)} \end{bmatrix}$ *is nonsingular, then the MPEC-LICQ holds.*
- *If* $| J_{GH^1} | + | J_{GH^2} | = 1$*:*
  - *If* $| J_{GH^1} | = 1$ *and* $\left(b^4,\, U^2_{(\cdot,\, K)}\right) \neq 0$*, then the MPEC-LICQ holds.*
  - *If* $| J_{GH^2} | = 1$ *and* $\left(b^3,\, U^1_{(\cdot,\, K)}\right) \neq 0$*, then the MPEC-LICQ holds.*

(iii) *If Assumption* 3 *and Assumption* 4 *fail, the MPEC-LICQ fails.*

To analyze the case where $I_g \neq \emptyset$ and $| J_{GH^1} | + | J_{GH^2} | \in \{0, 1\}$, we need the following further assumption:

Assumption 5. *Assume that* $| \Lambda_3^{(i)} \cup \Lambda_1^{(i)} | \leq 1$ *for all* $i \in [k]$.

Based on Proposition 3 in Li et al. (2022a), Assumption 5 basically means that for each lower-level problem, the number of correctly classified training data which lies on the boundary of margin does not exceed one.

Theorem 25. *Let $\bar{v}$ be a feasible point of problem* (MPEC-BHO) *satisfying* $I_g \neq \emptyset$ *and* $| J_{GH^2} | + | J_{GH^1} | \in \{0, 1\}$.

(i) *If Assumption* 5 *holds, the MPEC-LICQ holds at $\bar{v}$ if and only if at least one among Assumption* 3 *or* 4 *holds.*

(ii) *If Assumption* 5 *fails, then the MPEC-LICQ fails at $\bar{v}$.*

The above results are summarized in Figure 2, where Assumption C1 in Case 2.3.1 refers to the requirement that one of the following conditions holds:

- $| J_{GH^1} | + | J_{GH^2} | = 2$ and the matrix M^0 is nonsingular,
- $| J_{GH^1} | + | J_{GH^2} | = 1$, $| J_{GH^1} | = 1$ and $(a^2, b^2) \neq 0$,
- $| J_{GH^1} | + | J_{GH_2} | = 1$, $| J_{GH^2} | = 1$ and $(a^1, b^1) \neq 0$,

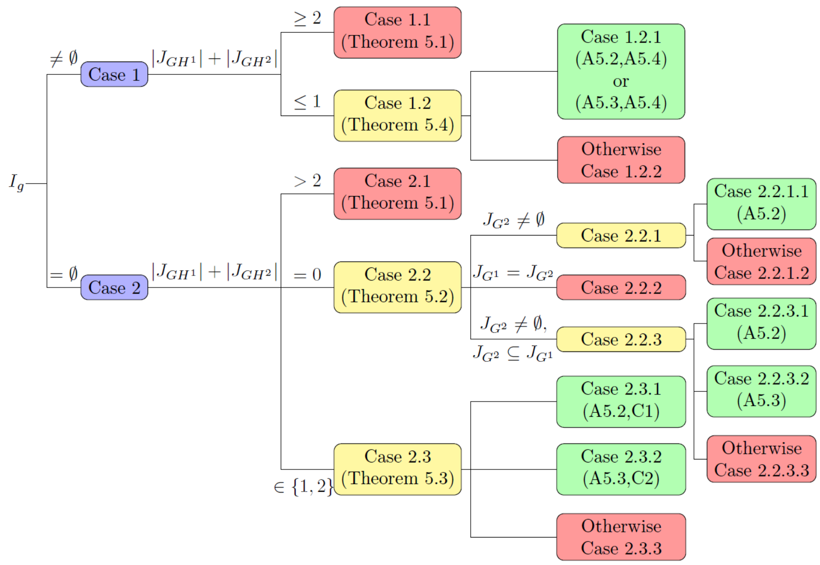


Figure 2: Summary of the behaviour of MPEC-LICQ for problem (MPEC-BHO), where A5.i stands for Assumption 5.i with $i \in \{2, 3, 4\}$. The red (resp. green) color indicates when the MPEC-SSOSC fails (resp. holds). As for the yellow color, it represents a transitional leave.

while Assumption C2 in Case 2.3.2 means that one of the following conditions holds:

- $\mid J_{GH^1} \mid + \mid J_{GH_2} \mid = 2$ and the matrix M^1 is nonsingular,
- $\mid J_{GH^1} \mid + \mid J_{GH^2} \mid = 1$, $\mid J_{GH^1} \mid = 1$ and $\left(b^4, U^2_{(\cdot,\, K)}\right) \neq 0$,
- $\mid J_{GH^1} \mid + \mid J_{GH^2} \mid = 1$, $\mid J_{GH^2} \mid = 1$ and $\left(b^3, U^1_{(\cdot,\, K)}\right) \neq 0$.

Note that in practice, Case 2-1 takes place in most situations as it is often the case that strict complementarity does not hold for some lower-level problems. In such a case, the number of indices in $\mid J_{GH^1} \mid + \mid J_{GH^2} \mid$ easily exceeds two. In other words, for most feasible points, MPEC-LICQ is likely to fail.

6 MPEC–SSOSC

Our main aim in this section is to study the MPEC-SSOSC in the context of problem (MPEC-BHO), with the intention of identifying situation where the condition fails or holds.

6.1 No differentiability requirement for leader's objective function

To proceed with the analysis, let $\bar{v}$ be the feasible point of (MPEC-BHO) and let $\bar{\eta} \in \mathbb{R}^{2n}$ and $\bar{\zeta} \in \mathbb{R}^{2n}$ be the Lagrange multipliers associated to the constraint functions G and H, respectively. Consider now the partition

$$\bar{\eta} := \begin{bmatrix} \bar{\eta}^1 \\ \bar{\eta}^2 \end{bmatrix}, \ \bar{\zeta} := \begin{bmatrix} \bar{\zeta}^1 \\ \bar{\zeta}^2 \end{bmatrix} \text{ with } \bar{\eta}^i := \begin{bmatrix} \bar{\eta}^i_{\Lambda_1} \\ \bar{\eta}^i_{\Lambda_2} \\ \bar{\eta}^i_{\Lambda_3} \\ \bar{\eta}^i_{\Lambda_4} \end{bmatrix}, \ \bar{\zeta}^i := \begin{bmatrix} \bar{\zeta}^i_{\Lambda_1} \\ \bar{\zeta}^i_{\Lambda_2} \\ \bar{\zeta}^i_{\Lambda_3} \\ \bar{\zeta}^i_{\Lambda_4} \end{bmatrix} \text{ for } i = 1,\ 2, \tag{58}$$

where the sets Λ_i, $i = 1, \ldots, 4$ are defined in (29)–(32). The following holds:

Proposition 26. *If $\bar{v}$ is a weak stationary point for problem* (MPEC-BHO) *with Lagrange multiplier vector $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$, then we have*

$$\bar{\eta}^1_{\Lambda_2} = 0,\ \bar{\eta}^2_{\Lambda_1} = 0,\ \bar{\eta}^2_{\Lambda_2} = 0,\ \bar{\eta}^2_{\Lambda_3^+} = 0,\ \bar{\zeta}^1_{\Lambda_3} = 0,\ \bar{\zeta}^1_{\Lambda_4} = 0, \quad \text{and } \bar{\zeta}^2_{\Lambda_4} = 0. \tag{59}$$

Remark 27. *Proposition* 26 *implies that for a weak stationary point $\bar{v}$ of problem* (MPEC-BHO) *with multipliers $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$ the following holds:*

$$\operatorname{supp}(\bar{\eta}^1) \subseteq (\Lambda_1 \cup \Lambda_3) \ \text{ and } \ \operatorname{supp}(\bar{\eta}^2) \subseteq (\Lambda_3^c \cup \Lambda_4); \tag{60}$$

$$\operatorname{supp}(\bar{\zeta}^1) \subseteq (\Lambda_1 \cup \Lambda_2) \ \text{ and } \ \operatorname{supp}(\bar{\zeta}^2) \subseteq (\Lambda_1 \cup \Lambda_3). \tag{61}$$

Denote

$$\begin{aligned}
\Lambda_1^{\eta^1} &:= \{j \in \Lambda_1 \ : \ \bar{\eta}^1 \neq 0\}, & \Lambda_3^{\eta^1} &:= \{j \in \Lambda_3 \ : \ \bar{\eta}^1 \neq 0\}, \\
(\Lambda_3^c)^{\eta^2} &:= \{j \in \Lambda_3^c \ : \ \bar{\eta}^2 \neq 0\}, & \Lambda_4^{\eta^2} &:= \{j \in \Lambda_4 \ : \ \bar{\eta}^2 \neq 0\}, \\
\Lambda_1^{\zeta^1} &:= \{j \in \Lambda_1 \ : \ \bar{\zeta}^1 \neq 0\}, & \Lambda_2^{\zeta^1} &:= \{j \in \Lambda_2 \ : \ \bar{\zeta}^1 \neq 0\}, \\
\Lambda_1^{\zeta^2} &:= \{j \in \Lambda_1 \ : \ \bar{\zeta}^2 \neq 0\}, & \Lambda_3^{\zeta^2} &:= \{j \in \Lambda_3 \ : \ \bar{\zeta}^2 \neq 0\},
\end{aligned}$$

and consider the following set that will play a critical role in the analysis:

$$\Lambda^0 := (\Lambda_1 \cup \Lambda_3) \setminus \left(\Lambda_1^{\eta^1} \cup \Lambda_1^{\zeta^2} \cup \Lambda_3^{\eta^1} \cup \Lambda_3^{\zeta^2} \right). \tag{62}$$

Theorem 28. *The MPEC-SSOSC fails at any weakly stationary point $\bar{v}$ of problem* (MPEC-BHO) *with multiplier vector $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$ if one of the following conditions holds:*

(i) $\Lambda_2 \cup \Lambda_4 \neq \emptyset$,

(ii) $\Lambda_2 \cup \Lambda_4 = \emptyset$ *and* $\Lambda^0 \neq \emptyset$,

where Λ_i for $i = 2, 4$ are defined as in (29) *and* (32).

Note that Theorem 28 does not require any second order information of f. The reason is that if (i) or (ii) holds, the set $(\bar{v})$ in (28) reduces to $\{0\}$, which means that the MPEC-SSOSC holds automatically. Based on this result, we next look closely at what happens under the following assumptions.

Assumption 6. *Let $\bar{v}$ be any weakly stationary point $\bar{v}$ of problem* (MPEC-BHO) *with multipliers $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$. At $\bar{v}$, it holds that $\Lambda_2 \cup \Lambda_4 = \emptyset$ and $\Lambda_0 = \emptyset$.*

Next, we discuss the special case where strict complementarity holds for the complementary constraints in (MPEC-BHO). In other words, we make the following assumption.

Assumption 7. *Assume that, for a weakly stationary point $\bar{v}$ of problem* (MPEC-BHO) *with multiplier vector $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$, conditions* (60) *and* (61) *hold as equalities, i.e.,*

$$\Lambda_1 = \Lambda_1^{\eta^1} = \Lambda_1^{\zeta^1} = \Lambda_1^{\zeta^2},\ \Lambda_3 = \Lambda_3^{\eta^1} = \Lambda_3^{\zeta^2},\ \Lambda_2 = \Lambda_2^{\eta^1},\ \Lambda_3^c = (\Lambda_3^c)^{\eta^2},\ \textit{and}\ \Lambda_4 = \Lambda_4^{\eta^2}.$$

To proceed with the next assumption, we select for each lower-level problem $i \in [k]$, the smallest index in $\Lambda_1^{(i)}$, which we denote by j_i. The corresponding index in Λ_1 is denoted by $\bar{j}_i$. Based on this, let

$$J^0 := \bigcup_{i\in[k]} \{\bar{j}_i\} \ \text{ and } \ \bar{J}^0 := \Lambda_1 \backslash J^0. \tag{63}$$

Let $\widetilde{A}_1 \in \mathbb{R}^{|\Lambda_1\cup\Lambda_3^c|\times|\Lambda_3^+|}$ and $\widetilde{A}_2 \in \mathbb{R}^{k\times|\Lambda_3^+|}$ be matrices of suitable sizes such that

$$Q(\bar{\gamma})_{(\Lambda_1\cup\Lambda_3^c,\, \Lambda_3^+)} = \widetilde{A}_1 Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_3^+)}, \widehat{Y}_{([k],\, \Lambda_3^+)} = \widetilde{A}_2 Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_3^+)}. \tag{64}$$

Note that, if $Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_3^+)}$ is positive definite, $\widetilde{A}^1$ and $\widetilde{A}^2$ are unique. Therefore, below, we first make the following assumption.

Assumption 8. *Let $\Lambda_3^+ \neq \emptyset$ and $Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^+)}$ be positive definite.*

Assumption 9. *Let J^0 and $\bar{J}^0$ be defined as in* (63)*. We assume rank$(\widetilde{M}) = 2$, where*

$$\widetilde{M} := \begin{bmatrix} \bar{a}_{\bar{J}^0} - \bar{A}^1 \bar{a}_{J^0} & \bar{b}_{\bar{J}^0} - \bar{A}^1 \bar{b}_{J^0} \\ \check{y}^1 - \bar{A}^2 \bar{a}_{J^0} & \check{y}^2 - \bar{A}^2 \bar{b}_{J^0} \end{bmatrix}$$

with $\bar{A}^1 \in \mathbb{R}^{|\bar{J}_0|\times k}$ and $\bar{A}^2 \in \mathbb{R}^{k\times k}$ being matrices such that

$$\bar{A}^1 \bar{P}_{(J^0,\, \cdot)} = \bar{P}_{(\bar{J}^0,\, \cdot)} \quad \textit{and} \quad \bar{A}^2 \bar{P}_{(J^0,\, \cdot)} = -\widetilde{A}_2 P_{(\Lambda_3^+,\, \cdot)} \tag{65}$$

respectively, while $\bar{a}$, $\bar{b}$, $\bar{P}$, $\check{y}^1$, and $\check{y}^2$ are defined as follows:

$$\begin{cases} \bar{a} := \left(Q(\bar{\gamma})_{(\Lambda_1\cup\Lambda_3^c,\, \Lambda_3^c)} - \widetilde{A}_1 Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_3^c)}\right) \mathbf{e}_{|\Lambda_3^c|}, \\ \bar{b} := \nabla_\gamma \theta(\bar{v})_{\Lambda_1\cup\Lambda_3^c} - \widetilde{A}_1 \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+}, \\ \bar{P} := P_{(\Lambda_1\cup\Lambda_3^c,\, \cdot)} - \widetilde{A}_1 P_{(\Lambda_3^+,\, \cdot)}, \\ \check{y}^1 := \left(\widehat{Y}_{([k],\, \Lambda_3^c)} - \widetilde{A}_2 Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_3^c)}\right) \mathbf{e}_{|\Lambda_3^c|}, \\ \check{y}^2 := -\widetilde{A}_2 \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+}. \end{cases} \tag{66}$$

Assumption 10. *Let J^0 and $\bar{J}^0$ be defined as in* (63)*. rank$(\widehat{M}) = |\Lambda_3^+| + 1$, where*

$$\widehat{M} := \begin{bmatrix} Q(\gamma)_{(\bar{J}^0,\, \Lambda_3^c)} \mathbf{e}_{|\Lambda_3^c|} - \bar{A}^3 Q(\gamma)_{(J^0,\, \Lambda_3^c)} \mathbf{e}_{|\Lambda_3^c|} & Q(\gamma)_{(\bar{J}^0,\, \Lambda_3^+)} - \bar{A}^3 Q(\gamma)_{(J^0,\, \Lambda_3^+)} \\ \widehat{Y}_{([k],\, \Lambda_3^c)} \mathbf{e}_{|\Lambda_3^c|} & \widehat{Y}_{([k],\, \Lambda_3^+)} \end{bmatrix}$$

with $\bar{A}^3 \in \mathbb{R}^{(|\Lambda_1|-k)\times k}$ being the unique matrix such that

$$\bar{A}^3 P_{(J^0,\, \cdot)} = P_{(\bar{J}^0,\, \cdot)}. \tag{67}$$

Assumption 9 and Assumption 10 will lead to the fact that $\mathcal{C}(\bar{v}) = \{0\}$ under proper situations, which leads to the following result.

Theorem 29. *Let $\bar{v}$ be a strongly stationary point of problem* (MPEC-BHO) *with multipliers vector $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$ satisfying Assumptions* 6, 7, 8. *Then, it holds that*

(i) *If $I_g = \emptyset$ and Assumption* 9 *hold at $\bar{v}$, then the MPEC-SSOSC holds at $\bar{v}$;*

(ii) *If $I_g \neq \emptyset$, $\bar{\lambda} > 0$, and Assumption* 10 *hold, then the MPEC-SSOSC holds at $\bar{v}$.*

6.2 Differentiable leader's objective function

The working assumption from now on is that the upper-level objective function f of the problem be at least twice continuously differentiable. There are multiple ways to achieve this; one is to apply the same trick in (5) to eliminate the max operator from the upper-level objective function. Secondly, a smooth loss function (e.g., the logistic loss) can be used instead of the hinge loss used in (3). One could also consider a smooth approximation of the max operator, such as the well-known smooth-max operator which we will use for the computational experiments in Section 7.

Note that the upper-level objective function f does not depend on σ or u. Hence, we automatically have $\nabla_\sigma f(v) = 0$ and $\nabla_u f(v) = 0$. It is also clear that if we were to apply the same trick in (5) to eliminate the max operator from the upper-level objective function, then the function f will be independent from C. Of course, note that if proceed with this trick, the main change in our analysis of the previous two sections will be the structure of the function g. However, most of the results could be derived similarly. This will be carefully analyze in a separate work. Next, we precisely analyze what happens in the latter scenario; i.e., we make the following assumption:

Assumption 11. *Let $\nabla_C f(v) = 0$.*

Assumption 12. *Let $\bar{v}$ be a strongly stationary point of* (MPEC-BHO) *with $\bar{\gamma} > 0$ and* $\mathrm{rank}(\widetilde{M}) < 2$, *and $d^\gamma (\dot{V}^f)^\top \dot{D}^f \dot{V}^f d^\gamma > 0$ for all $d^\gamma \neq 0$. Here,*

$$\dot{V}^f := \begin{bmatrix} 1 \\ \tau \\ k\mathbf{e}_{|\Lambda_3^c|} \end{bmatrix} \quad \text{and} \quad \dot{D}^f := \begin{bmatrix} \dot{D}^f_{11} & \dot{D}^f_{12} \\ (\dot{D}^f_{11})^\top & \dot{D}^f_{22} \end{bmatrix}, \tag{68}$$

while (recall $\bar{\eta}$ defined in (58)*)*

$$\left\{ \begin{array}{rcl} \tau & := & \left(Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_3^+)}\right)^{-1} \left(\beta Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_3^c)|\Lambda_3^c|} - \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+} \right. \\ & & \left. - P_{(\Lambda_3^+,\, \cdot)} (\bar{P}_{(J^0,\, \cdot)})^{-1} (\bar{a}_{J^0}\beta - \bar{b}_{J^0})\right), \\ \dot{D}^f_{11} & := & \nabla^2_{\gamma\gamma} f(\bar{v}) + \sum_{i \in \Lambda_3} \bar{\eta}^1_i \nabla^2_{\gamma\gamma} \theta_i(\bar{v}), \\ \dot{D}^f_{12} & := & \nabla^2_{\gamma\alpha_{\Lambda_3}} f(\bar{v}) + \sum_{i \in \Lambda_3} \bar{\eta}^1_i \nabla^2_{\gamma\alpha_{\Lambda_3}} \theta_i(\bar{v}), \\ \dot{D}^f_{22} & := & \nabla^2_{\alpha_{\Lambda_3}\alpha_{\Lambda_3}} f(\bar{v}). \end{array} \right. \tag{69}$$

Assumption 13. *Let $\bar{v}$ be a strongly stationary point of problem* (MPEC-BHO) *with $\bar{\gamma} = 0$ and $\bar{\lambda} > 0$, let rank$(\widehat{M}) <| \Lambda_3^+ | +1$ and*

$$\hat{d}^\top (V^f)^\top D^f V^f \hat{d} > 0 \quad \textit{for all} \quad \hat{d} := \begin{bmatrix} d^C \\ d^\alpha_{\Lambda_3^+} \end{bmatrix} \neq 0 \ \textit{ with } \ d^C,\ d^\alpha_{\Lambda_3^+},\ d^u \ \textit{ satisfying}$$

$$\begin{bmatrix} Q(\gamma)_{([n],\,\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} & Q(\gamma)_{([n],\,\Lambda_3^+)} & P_{([n],\,\cdot)} \\ \widehat{Y}_{([k],\,\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} & \widehat{Y}_{([k],\,\Lambda_3^+)} & \mathbf{0}_{([k],L_5)} \end{bmatrix} \begin{bmatrix} d^C \\ d^\alpha_{\Lambda_3^+} \\ d^u \end{bmatrix} = 0. \tag{70}$$

Here, we have

$$V^f := \begin{bmatrix} 0 & \mathbf{I}_{(|\Lambda_3^+|,\,|\Lambda_3^+|)} \\ \mathbf{e}_{|\Lambda_3^c|} & 0 \end{bmatrix} \quad \textit{and} \quad D^f := \nabla^2_{\alpha_{\Lambda_3}\alpha_{\Lambda_3}} f(\bar{v}). \tag{71}$$

Theorem 30. *Let $\bar{v}$ be a strongly stationary point of* (MPEC-BHO) *with multipliers vector $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$ satisfying Assumptions* 6, 7, 8, *and* 11.

(i) *If $I_g = \emptyset$ and Assumption* 12 *hold at $\bar{v}$, then MPEC-SSOSC holds;*

(ii) *If $I_g \neq \emptyset$, $\bar{\lambda} > 0$, and Assumption* 13 *hold, MPEC-SSOSC holds.*

6.3 Summary of main results on the MPEC-SSOSC

The above results about MPEC-SSOSC are summarized in Figure 3.

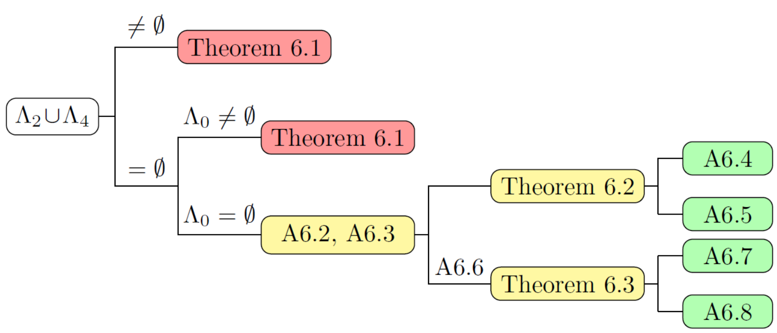


Figure 3: Partition of different cases for MPEC-SSOSC, where A6.i stands for Assumption 6.i with $i \in \{2, \ldots, 8\}$. The red (resp. green) color indicates when MPEC-SSOSC fails (resp. holds). As for the yellow color, it represents a transitional leave.

7 Applications to Scholtes' Relaxation Method

In this section, we introduce the Scholtes relaxation method, and show how the results from Sections 4–6 can be used to derive its convergence results.

Recall that the basic idea of the Scholtes relaxation method Scholtes (2001) is as follows. Let $\{t_j\} \downarrow 0$ be a sequence of relaxation parameters. At each iteration, we replace problem (MPEC-BHO) by the relaxed mathematical program

$$\begin{array}{ll} \min & f(v) \\ \text{s.t.} & g(v) \leq 0, \;\; h(v) = 0, \\ & G(v) \geq 0, \;\; H(v) \geq 0, \;\; G_i(v)^\top H_i(v) \leq t_j, \;\; i \in [2n], \end{array} \tag{NLP-t_j}$$

which is clearly a *nonlinear program* (NLP–for short), even in the case where our original bilevel program is linear. (NLP-t_j) is then solved and t_j decreased at each iteration. The process is repeated until a certain stopping criterion is satisfied. Clearly, (NLP-t_j) parameterized in t_j for $j = 1, 2, \ldots$, has a larger feasible set than (MPEC-BHO). Hence, the decrease in t_j leads to a feasible set which gets progressively closer to that of (MPEC-BHO).

Throughout this section, we assume that the function f is continuously differentiable. Otherwise, a smoothing approximation can be used; see Subsection 7.2 for a possible choice of a smoothing function. The details of the Scholtes-based global relaxation method (GRM) are shown in Algorithm 1.

Algorithm 1 The Global Relaxation Method (GRM) $(v_0, t_0, \sigma, t_{min})$

Require a starting vector v_0, an initial relaxation parameter t_0, and parameters $\hat{\rho} \in (0, 1), t_{\min} > 0$.

Set $j := 0$.

do

  Compute an approximate solution v^{j+1} of problem (NLP-t_j) using v^j as starting point.

  Let $t_{j+1} \leftarrow \hat{\rho} \cdot t_j$ and $j \leftarrow j + 1$.

while $t_j > t_{\min}$

Return the final iterate $v_{opt} := v^j$, the corresponding function value $f(v_{opt})$, and the maximum constraint violation $\text{Vio}(v_{opt})$.

7.1 Convergence results

In this subsection, we provide some convergence results for Algorithm 1, which can be established thanks to our results from Sections 4–6.

Corollary 31. *Let $\{t_j\} \downarrow 0$ and v^j be a KKT point of* (NLP-t_j) *with $v^j \to \bar{v}$, where $\bar{v}$ is feasible point of* (MPEC-BHO). *Then $\bar{v}$ is a C-stationary point of* (MPEC-BHO).

Proof Considering the fact that MPEC-MFCQ automatically holds at any feasible point of (MPEC-BHO) (cf. Theorem 20), we get the result; see, e.g., Hoheisel et al. (2013). ■

This result can be strengthened if we impose assumptions from Section 5 ensuring the fulfilment of the stronger MPEC-LICQ.

Theorem 32. *Let $\{t_j\} \downarrow 0$ and $\{(v^j, \lambda^j, \mu^j, \eta^j, \zeta^j, \delta^j)\}$ be a sequence of KKT points of* (NLP-t_j) *with $v^j \to \bar{v}$. Furthermore, let all the assumptions resulting from one of the green leaves in Figure 2 be satisfied. Then $\bar{v}$ M-stationary for problem* (MPEC-BHO).

Proof See Subsection F for the proof. ■

This is quite an interesting result, as it demonstrates that, unlike in most of the literature, where it is common to get only C-stationarity points, the stronger M-stationarity could be obtained from the Scholtes algorithm.

The next results shows that, under stronger assumptions, we can guarantee a unique M-stationary point in some neighborhood of the point of interest.

Corollary 33. *Let $\bar{v}$ be a S-stationary point of problem* (MPEC-BHO) *with multipliers* $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$ *that satisfies assumptions (i) and (ii) below:*

(i) All the assumptions resulting from one of the green leaves in Figure 2 are satisfied.

(ii) All the assumptions resulting from one of one of the green leaves in Figure 3 hold.

Then, there exists a neighborhood of $\bar{v}$, where $\bar{v}$ is the only M-stationary point.

Proof Recall that assumption (i) ensures the fulfillment of the MPEC-LICQ based on Theorem 23, Theorem 24, or Theorem 25. Similarly, any assumption in (ii) leads to the satisfaction of the MPEC-SSOSC according to Theorem 29 or Theorem 30. Therefore, the result follows by applying (Kanzow and Schwartz, 2013, Theorem 4.11). ■

Finally, we state the following results, which also requires the MPEC-SSOC and MPEC-LICQ to ensure that, for a given S-stationary point $\bar{v}$, we can find a sequence of local optimal solutions of the relaxed problems (NLP-t_j) that converges to this point. Note that, here, $\mathbb{X}(t_j)$ denotes the feasible set of the relaxed problem (NLP-t_j).

Corollary 34. *Let $\bar{v}$ be a S-stationary point of* (MPEC-BHO)*, with multiplier vector* $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$*, that satisfies Assumption (ii) in Corollary 33. Then there exists a neighborhood $U(\bar{v})$ of $\bar{v}$ such that for every sequence $t_j \downarrow 0$ such that the relaxed problems* (NLP-t_j) *have at least one local minimum $v^j \in U(\bar{v}) \cap \mathbb{X}(t_j)$ for all j sufficiently large. If additionally, Assumption (i) in Corollary 33 holds at $\bar{v}$, then $v^j \to \bar{v}$.*

Proof Given that Assumption (ii) in Corollary 33 ensures that MPEC-SSOSC holds, and similarly, as we have the satisfaction of the MPEC-LICQ under Assumption (i) in Corollary 33, the result directly follows by applying (Kanzow and Schwartz, 2013, Theorem 4.12). ■

7.2 Illustrative numerical examples

In this section we will provide illustrative examples of our bilevel hyperparameter tuning model in action on real data sets from the Cleveland Heart Disease, Glass Classification data set, Pima Indians Diabetes and the Sonar, Mines vs. Rocks data sets from the UCI Machine Learning Repository. In this paper we perform k-fold using $k = 3$ as is suggested in similar works on bilevel optimization for SVM hyperparameter tuning Kunapuli et al. (2008b); Bennett et al. (2006, 2008). In these experiments we use the following smooth approximation of the function $\max(0, x)$:

$$\mathrm{SmoothMax}\,(x) := x + \sqrt{x^2 + \zeta},$$

Table 1: Solutions found by each hyperparameter tuning method. Here, PID stands for Pima Indians Diabetes.

Data Set	Tuning Method	Objective Value	Time(s)	γ	C
Cleveland	Bilevel	**1.0979**	5311.0	0.0147	9.6651
	Grid Search	1.1110	5245.0	0.0100	857.70
Glass	Bilevel	**1.5215**	2252.3	0.0989	11.195
	Grid Search	1.5666	4203.4	0.0631	21.544
PID	Bilevel	**2.0420**	6530.5	0.0211	5.7434
	Grid Search	2.0851	4552.6	0.0016	73.564
Sonar	Bilevel	**1.8078**	5779.6	0.0066	11.691
	Grid Search	1.8252	5211.5	0.0100	251.19

where $\zeta > 0$ is a small perturbation. As for the function Ω_C, for $C > 0$, we adopt the following smooth quadratic function on $(0,\ C)$:

$$\Omega_C(\alpha) := 1 - \left(\frac{2\alpha}{C} - 1\right)^2 + \tau \ \text{ for all } \ \alpha \in (0,\ C),$$

where τ is a small perturbation parameter used to in order to avoid dividing by 0 (see the proof of Proposition 2 in Section A of the appendix).

In order to keep the size of the bievel program manageable, we sample $n = 100$ pairs of data points and labels. This means that each fold will contain either 33 or 34 pairs of data points and labels. To ensure that class imbalance did not play a role in our analysis our 100 point data sets contain 50 data point, label pairs from each class. As scaling can effect the performance of RBF kernel SVMs, we apply standard scaling to each variable.

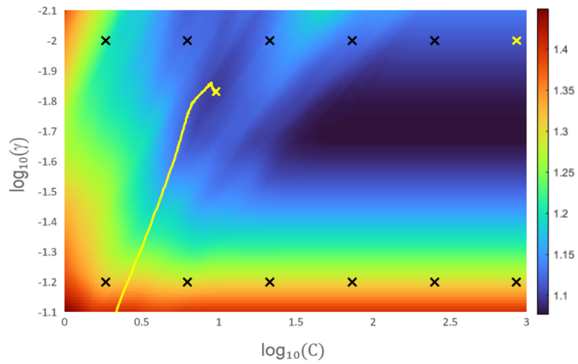
We compare our bilevel hyperparameter tuning algorithm with grid search. This grid search will be performed by iteratively selecting values or γ and C from a grid of hyperparameter combinations, solving the lower-level training problem (LLP^i) using `fmincon` in MATLAB and then using the output values for α together with the selected values for γ and C to compute the value of the objective function of problem (NLP-t_j).

The hyperparameter grid used here has the shape 16×16 and is comprised of a logarithmic range of values of

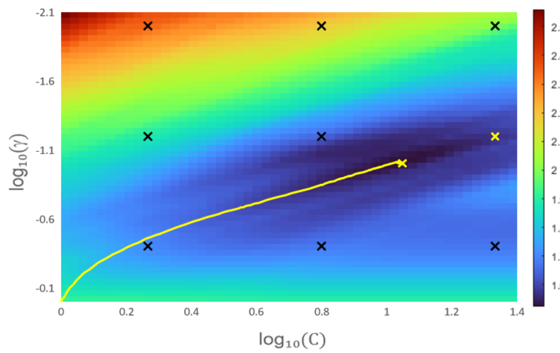
$$C \in \{10^{-4+8(i-1)/16-1} \ \forall i \in [16]\} \ \text{ and } \ \gamma \in \{10^{-6+12(i-1)/16-1} \ \forall i \in [16]\}.$$

The size of this grid was chosen as it resulted in the most similar run-times to those of our bilevel algorithm across the data sets in our experiments. The comparison of our bilevel algorithm and grid search on the aforementioned data sets can be found in Table 1.

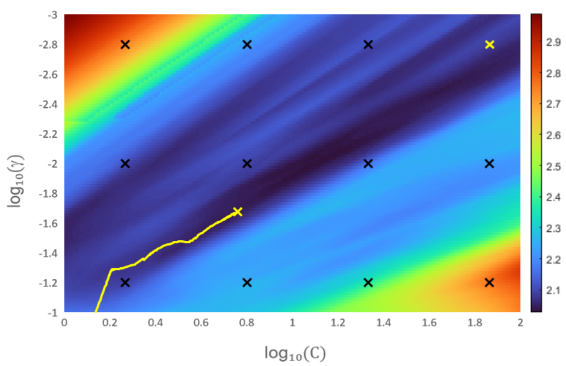
Figure 4 shows how our bilevel hyperparameter tuning algorithm navigates the hyperparameter space for two of the data sets. The path taken by our bilevel hyperparameter tuning algorithm is shown in yellow with the yellow cross being it's finial solution. The black crosses represent the points tested by the grid search described previously. Of course, most of the points evaluated by this grid search do not lie in the region shown in these plots. The heatmap shown in the background was generated by conducting a very fine grid



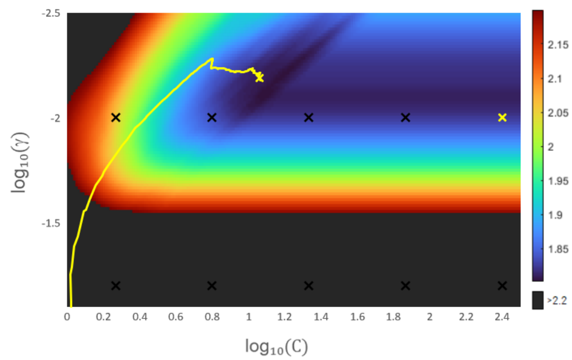
(a) Cleveland



(b) Glass



(c) PID



(d) Sonar

Figure 4: Performance comparison for grid search and our method for tuning hyperparameters for the Cleveland, Glass, PID (Pima Indian Diabetes), and Sonar datasets.

search of this local region, the run time of which was orders or magnitude greater than that of either the previous grid search or our bilevel algorithm. As can be seen, in each instance the bilevel model seems to converge to a local minimum. As expected, we see that this approach is able to find solutions in-between the points tested by grid search.

The Glass data set illustrates the exact behaviour which makes the bilevel approach conceptually preferable to grid search. As shown in Table 1, the best hyperparameter combination found by grid search for this data set is $\gamma = 10^{-1.2}$ and $C = 10^{1.\dot{3}}$. As can be seen in Figure 4, this is the closest point tested by the grid search to the apparent local minimum which the bilevel algorithm has correctly located. In other words, there exist a local minimum between the points tested by grid search and the bilevel algorithm was able to navigate this space to approximately locate this local minimum.

Similar conclusions can be drawn for PID and Sonar data sets.

8 Conclusion and Future Work

We have proposed a bilevel optimization model, that is based on cross-validation principle, to calculate hyperparameters for SVM with nonlinear kernel. Then considering the MPEC/KKT reformulation of the problem, key concepts (namely, the MPEC-MFCQ, MPEC-LICQ, and MPEC-SSOSC) are studied and can conditions ensuring that they hold or fail are established. Overall, only the MPEC-MFCQ holds automatically. For the other ones, the required assumptions are provided, and summarized in Figures 2 and 3.

Despite the usefulness of these results, as illustrated in the context of the Scholtes algorithm (see Section 7), many open questions remain, and will be explored in future works. For instance, the hinge loss used in the leader's objective function (2) is essentially for illustrative purposes, and did not affect much of our analysis. Our results can easily be adapted to other loss functions, including the *counting loss* (to minimize the number of missclassified points) commonly used in techniques such as grid search and many bilevel hyperparameter optimization for machine learning papers (see Li et al. (2022a) and references therein). Furthermore, recall that we can easily get a smooth functions in the leader's objective function, even in the context of problem (BHO), where the trick used in the lower-level problem (see (5)) can be applied to eliminate the *max* operator in the leader's objective function. However, in the latter case, such a transformation could lead to a problem with larger size.

Also, in this paper we have used the Gaussian kernel (7) just for illustrative purposes, as it is also one of the most commonly used in the machine learning literature. However, it should be possible, to extend our analysis to multiple other typers of kernel functions. Moreover, we have considered just a single regularization hyperparameter and corresponding ℓ_2 regularization term. In a future work, we will study the problem with multiple hyperparameters, possibly associated regularization functions that are nonsmooth.

There are also multiple other types of MPEC-tailored second order sufficient conditions that could be studied in the context of problem (MPEC-BHO) (see, e.g., Guo et al. (2013) for an overview of such conditions). Their analysis and consequences will be evaluated in the future. Various other types of methods (different from tested in Section 7 of this paper) requiring the theory provided in this paper could be studied; see, e.g., Hoheisel et al. (2013) for a sample of such methods.

Acknowledgments and Disclosure of Funding

SC's work is supported by The Turing Institute under the EPSRC grants EP/N510129/1 and EP/W037211/1); AD's work is jointly funded by DAS and EPSRC via the studentship EP/R513325/1; QL's research is supported by the National Science Foundation of China (NSFC) with Reference 12071032 and 12271526; AZ's work is supported by EPSRC grant EP/V049038/1 and the Turing Institute under the EPSRC grant EP/N510129/1.

Appendix A. Proofs for Section 2

Proof of Proposition 1

$C = 0$ implies $\omega^{(i)*} = 0$ for all $i \in [k]$. Thus, for each $i \in [k]$ and $j \in [\bar{n}^{(i)}]$, we have

$$\Xi_{\phi_\gamma}\left(\bar{y}_j^{(i)}, \omega^{(i)}, \bar{X}_j^{(i)}, b^{(i)}\right) = \max\left\{0, 1 - \bar{y}_j^{(i)}\left((\omega^{(i)})^\top \phi_\gamma(\bar{X}_j^{(i)}) + b^{(i)}\right)\right\} = \max\left\{0, 1 - \bar{y}_j^{(i)} b^{(i)}\right\}.$$

The problem decomposes into k sub-problems, one per split of index $i \in [k]$, which reads

$$\min_{\gamma \geq 0,\, b^{(i)}} \quad F(0, \gamma, 0, b^{(i)}) := \frac{1}{k}\frac{1}{\bar{n}^{(i)}} \sum_{j \in [\bar{n}^{(i)}]} \max\left\{0, 1 - \bar{y}_j^{(i)} b^{(i)}\right\}. \tag{72}$$

We rewrite the summation as the following univariate function:

$$\mathscr{S}(b^{(i)}) := \underbrace{\sum_{j \in A^{(i)}} \max\left\{0, 1 - b^{(i)}\right\}}_{\mathscr{S}^1(b^{(i)})} + \underbrace{\sum_{j \in B^{(i)}} \max\left\{0, 1 + b^{(i)}\right\}}_{\mathscr{S}^2(b^{(i)})}.$$

Subsequently, we proceed with a distinction of cases on the value of $b^{(i)}$.

- If $b^{(i)} \leq -1$, $\mathscr{S}^1(b^{(i)}) = [A^{(i)}](1 - b^{(i)})$ (since $1 - b^{(i)} \geq 0$) and $\mathscr{S}^2(b^{(i)}) = 0$ (since $1 + b^{(i)} \leq 0$). Thus, $\mathscr{S}(b^{(i)}) = |A^{(i)}|(1 - b^{(i)})$. We have $\min_{b^{(i)}) \leq -1} \mathscr{S}(b^{(i)}) = 2|A^{(i)}|$, attained at $b^{(i)} = -1$.
- If $-1 \leq b^{(i)} \leq 1$, we have the equalities $\mathscr{S}^1(b^{(i)}) = |A^{(i)}|(1 - b^{(i)})$ (since $1 - b^{(i)} \geq 0$) and $\mathscr{S}^2(b^{(i)}) = |B^{(i)}(1 + b^{(i)})$ (since $1 + b^{(i)} \geq 0$). Thus, it holds that $\mathscr{S}(b^{(i)}) = |A^{(i)}|(1 - b^{(i)}) + |B^{(i)}|(1 + b^{(i)}) = |A^{(i)}| + |B^{(i)}| + (|B^{(i)}| - |A^{(i)}|)b^{(i)}$. Hence,

$$\min_{-1 \leq b^{(i)} \leq 1} \mathscr{S}(b^{(i)}) = \begin{cases} |A^{(i)}| + |B^{(i)}| & \text{if } |A^{(i)}| = |B^{(i)}|, \text{ attained at any } b^{(i)} \in [-1, 1], \\ 2|A^{(i)}| & \text{if } |B^{(i)}| > |A^{(i)}|, \text{ attained at } b^{(i)} = -1, \\ 2|B^{(i)}| & \text{if } |B^{(i)}| < |A^{(i)}|, \text{ attained at } b^{(i)} = 1. \end{cases}$$

- If $b^{(i)} \geq 1$, $\mathscr{S}^1(b^{(i)}) = 0$ (since $1 - b^{(i)} \leq 0$) and $\mathscr{S}^2(b^{(i)}) = |B^{(i)}|(1 + b^{(i)})$ (since $1 + b^{(i)} \geq 0$). Thus, $\mathscr{S}(b^{(i)}) = |B^{(i)}|(1 + b^{(i)})$. We have $\min_{b^{(i)}) \geq 1} \mathscr{S}(b^{(i)}) = 2|B^{(i)}|$, attained at $b^{(i)} = 1$.

Thus, we deduce that

$$\min_{b^{(i)}} \mathscr{S}(b^{(i)}) = \begin{cases} 2|A^{(i)}| = 2|B^{(i)}| & \text{if } |A^{(i)}| = |B^{(i)}|, @ \text{ any } b^{(i)} \in [-1, 1], \\ 2|B^{(i)}| & \text{if } |B^{(i)}| < |A^{(i)}|, @\ b^{(i)} = 1, \\ 2|A^{(i)}| & \text{if } |B^{(i)}| > |A^{(i)}|, @\ b^{(i)} = -1, \end{cases}$$

where @ stands for "attained at". □

Proof of Proposition 2

Let (C, ω, b) be a feasible point of problem (4). Note that if $C = 0$, the result obviously follows from Proposition 1. Otherwise, let $C > 0$. Then consider a fold $i \in [k]$ and introduce the Lagrangian function of the corresponding problem (5),

$$\begin{aligned} L_C^{(i)}\left(\vartheta^{(i)}\right) \;\; := \;\; & \tfrac{1}{2}\|\omega^{(i)}\|^2 \; + \; C \textstyle\sum_{j\in[\hat{n}^{(i)}]} \xi_j^{(i)} \\ & + \textstyle\sum_{j\in[\hat{n}^{(i)}]} \alpha_j^{(i)} \left[1 - \xi_j^{(i)} - \hat{y}_j^{(i)} \left(\left(\omega^{(i)}\right)^\top \phi_\gamma\left(\hat{X}_j^{(i)}\right) + b^{(i)}\right)\right] \\ & - \textstyle\sum_{j\in[\hat{n}]} \eta_j^{(i)} \xi_j^{(i)}, \end{aligned}$$

where $\vartheta^{(i)} := \left(w^{(i)}, b^{(i)}, \xi^{(i)}, \alpha^{(i)}, \eta^{(i)}\right)$ with $\alpha^{(i)}$ and $\eta^{(i)}$ being the Lagrange multipliers associated to the two classes of constraints in problem (5). Based on the above expression, a fixed fold $i \in [k]$, the Lagrangian dual of problem (5) can be written as

$$\max_{\alpha^{(i)},\, \eta^{(i)}\geq 0} \left\{ \min_{\omega^{(i)},\, b^{(i)},\, \xi^{(i)}\geq 0} L_C^{(i)}\left(\vartheta^{(i)}\right) \right\}. \tag{73}$$

For any $i \in [k]$ and fixed $\alpha^{(i)}$, $\eta^{(i)} \geq 0$, assume that the inner minimization problem in (73) has an optimal solution $\left(w^{(i)}, b^{(i)}, \xi^{(i)}\right)$, then since the constraints of (5) are all linear, then it follows that the point $\left(w^{(i)}, b^{(i)}, \xi^{(i)}\right)$ is a stationary point; hence, implying that

$$\omega^{(i)} - \sum_{j\in[\hat{n}^{(i)}]} \alpha_j^{(i)} \hat{y}_j^{(i)} \phi_\gamma(\hat{X}_j^{(i)}) = 0, \tag{74}$$

$$\forall j \in [\hat{n}^{(i)}]: \quad C - \alpha_j^{(i)} - \eta_j^{(i)} = 0. \tag{75}$$

Recall the upper-level objective function of problem (3)

$$F(C, \omega, b) := \frac{1}{k} \sum_{i\in[k]} \sum_{\ell\in[\bar{n}^{(i)}]} \frac{1}{\bar{n}^{(i)}} \Xi_{\phi_\gamma}\left(\bar{y}_\ell^{(i)},\, \omega^{(i)},\, \bar{X}_\ell^{(i)},\, b^{(i)}\right), \tag{76}$$

where $\omega := \left(\omega^{(1)}, \ldots, \omega^{(k)}\right)$, $b := \left(b^{(1)}, \ldots, b^{(k)}\right)$, and

$$\Xi_{\phi_\gamma}\left(\bar{y}_\ell^{(i)},\, \omega^{(i)},\, \bar{X}_\ell^{(i)},\, b^{(i)}\right) := \max\left\{0,\; 1 - \bar{y}_\ell^{(i)} \left(\left(\omega^{(i)}\right)^\top \phi_\gamma\left(\bar{X}_\ell^{(i)}\right) + b^{(i)}\right)\right\}.$$

Based on (74), we have

$$\begin{aligned} \left(\omega^{(i)}\right)^\top \phi_\gamma\left(\bar{X}_\ell^{(i)}\right) &= \sum_{j\in[\hat{n}^{(i)}]} \alpha_j^{(i)} \hat{y}_j^{(i)} \phi_\gamma\left(\hat{X}_j^{(i)}\right)^\top \phi_\gamma\left(\bar{X}_\ell^{(i)}\right) \\ &= \sum_{j\in[\hat{n}^{(i)}]} \alpha_j^{(i)} \hat{y}_j^{(i)} \exp\left(-\gamma \left\|\hat{X}_j^{(i)} - \bar{X}_\ell^{(i)}\right\|^2\right). \end{aligned} \tag{77}$$

Next, we consider the complementarity conditions associated to the constraints in (3):

$$0 \leq \alpha_j^{(i)} \cdot \left[1 - \xi_j^{(i)} - \hat{y}_j^{(i)} \left((\omega^{(i)})^\top \phi_\gamma(\hat{X}_j^{(i)}) + b^{(i)}\right)\right] \leq 0, \;\; j \in [\hat{n}^{(i)}], \tag{78}$$

$$0 \leq \xi_j^{(i)} \cdot \eta_j^{(i)} \geq 0, \;\; j \in [\hat{n}^{(i)}]. \tag{79}$$

For any $i \in [k]$ and $j \in [\hat{n}^{(i)}]$ such that $\alpha_j^{(i)} \in (0,\ C)$, we have from (75) that $\eta_j^{(i)} > 0$ and therefore, from (79), it follows that $\xi_j^{(i)} = 0$, and hence, from (78),

$$b^{(i)} = \hat{y}_j^{(i)} - \left(\omega^{(i)}\right)^\top \phi_\gamma\left(\hat{X}_j^{(i)}\right),\ \ \forall i \in [k],\ \ \forall j \in \left[\hat{n}^{(i)}\right]\ \text{ s.t. }\ \alpha_j^{(i)} \in (0,\ C)\,.$$

Considering the expression of $\omega^{(i)}$ from (74), and similarly to (77), it follows that for all $i \in [k]$ and $j \in \left[\hat{n}^{(i)}\right]$ such that $\alpha_j^{(i)} \in (0,\ C)$, we have

$$b^{(i)} = \hat{y}_j^{(i)} - \sum_{t \in [\hat{n}^{(i)}]} \alpha_t^{(i)} \hat{y}_t^{(i)} \exp\left(-\gamma \left\| \hat{X}_t^{(i)} - \hat{X}_j^{(i)} \right\|^2\right).$$

To proceed with the final step, take $i \in [k]$ and define the sets

$$\begin{aligned} \bar{I}^{(i)} &:= \left\{ j \in [\hat{n}^{(i)}] :\ 0 < \alpha_j^{(i)} < C \right\}, \\ \hat{I}^{(i)} &:= \left\{ j \in [\hat{n}^{(i)}] :\ \alpha_j^{(i)} = 0 \ \text{ or } \ \alpha_j^{(i)} = C \right\}. \end{aligned}$$

The considering any function $\Omega_C : \mathbb{R} \to \mathbb{R}$ satisfying $\Omega_C(\zeta) = 0$ for $\zeta \leq 0$ and $\zeta \geq C$ and $\Omega_C(\zeta) > 0$ all $\zeta \in (0, C)$, it holds that

$$\begin{aligned} &\frac{1}{\sum\limits_{j \in [\hat{n}^{(i)}]} \Omega_C\left(\alpha_j^{(i)}\right)} \sum_{j \in [\hat{n}^{(i)}]} \Omega_C\left(\alpha_j^{(i)}\right)\left(\hat{y}_j^{(i)} - \mathbb{H}^{(i)}\left(\hat{X}_j^{(i)}, \gamma\right)\right) \\ &= \frac{1}{\sum\limits_{j \in \bar{I}^{(i)}} \Omega_C\left(\alpha_j^{(i)}\right)} \sum_{j \in \bar{I}^{(i)}} \Omega_C\left(\alpha_j^{(i)}\right)\left(\hat{y}_j^{(i)} - \mathbb{H}^{(i)}\left(\hat{X}_j^{(i)}, \gamma\right)\right) \\ &\qquad + \frac{1}{\sum\limits_{j \in \bar{I}^{(i)}} \Omega_C\left(\alpha_j^{(i)}\right)} \sum_{j \in \hat{I}^{(i)}} \Omega_C\left(\alpha_j^{(i)}\right)\left(\hat{y}_j^{(i)} - \mathbb{H}^{(i)}\left(\hat{X}_j^{(i)}, \gamma\right)\right) \\ &= \frac{1}{\sum\limits_{j \in \bar{I}^{(i)}} \Omega_C\left(\alpha_j^{(i)}\right)} \sum_{j \in \bar{I}^{(i)}} \Omega_C\left(\alpha_j^{(i)}\right) b^{(i)} \\ &= b^{(i)} \end{aligned}$$

with $\mathbb{H}^{(i)}\left(\hat{X}_j^{(i)}, \gamma\right) := \sum\limits_{t \in [\hat{n}^{(i)}]} \alpha_t^{(i)} \hat{y}_t^{(i)} \exp\left(-\gamma \left\| \hat{X}_t^{(i)} - \hat{X}_j^{(i)} \right\|^2\right)$. Replacing this expression of $b^{(i)}$ for $i \in [k]$, together with (77) in (76), we get the expression in Proposition 2. $\square$

Proof of Proposition 3

Recall the primal form of lower bound problem in (5). The KKT conditions for problem (5) are written as follows:

$$w^{(i)} - \sum_{j \in [\hat{n}^{(i)}]} \alpha_j^{(i)} \hat{y}_j^{(i)} \omega^{(i)\top} \phi(\hat{X}_j^{(i)}) = 0, \tag{80}$$

$$C - \underline{\alpha}_j^{(i)} - \alpha_j^{(i)} = 0,\ j \in [\hat{n}^{(i)}], \tag{81}$$

$$\sum_{j \in [\hat{n}^{(i)}]} \xi_j^{(i)} \hat{y}_j^{(i)} = 0,\ j \in [\hat{n}^{(i)}], \tag{82}$$

$$\underline{\alpha}_j^{(i)} \geq 0,\ \xi_j^{(j)} \geq 0,\ \alpha_j^{(i)} \xi_j^{(i)} = 0,\ j \in [\hat{n}^{(i)}], \tag{83}$$

$$\alpha_j^{(i)} \geq 0,\ \xi_j^{(j)} \geq 1 - \hat{y}_j^{(i)}(\omega^{(i)\top} \phi(\hat{X}_j^{(i)}) + b^{(i)}),\ j \in [\hat{n}^{(i)}], \tag{84}$$

$$\alpha_j^{(i)} \left(\xi_j^{(i)} - 1 + \hat{y}_j^{(i)}(\omega^{(i)\top} \phi(\hat{X}_j^{(i)}) + b^{(i)}) \right) = 0,\ j \in [\hat{n}^{(i)}]. \tag{85}$$

Here $\alpha^{(i)}$ and $\underline{\alpha}^{(i)}$ are the Lagrange multipliers corresponding to the first and second part of inequality constraints in (5). Note that $\alpha^{(i)}$ is also the solution of (LLP^i). By assumption, it holds that $[\hat{n}^{(i)}] = \mathbb{I}_{=}(\alpha^{(i)}) \cup \mathbb{I}_{<}(\alpha^{(i)})$. Hence for each $j \in \mathbb{I}_{=}(\alpha^{(i)})$, it holds that $\alpha_j^{(i)} = 0$ by definition, and $\underline{\alpha}_j^{(i)} = C$ by (81), which gives $0 = \xi_j^{(i)} \geq 1 - \hat{y}_j^{(i)}(\omega^{(i)\top} \phi(\hat{X}_j^{(i)}) + b^{(i)}$ by (83), (84) and (85). Therefore, we obtain that (note $\left(\hat{y}_j^{(i)}\right)^2 = 1$)

$$b^{(i)} \geq \hat{y}_j^{(i)} - \omega^{(i)\top} \phi(\hat{X}_j^{(i)}),\ j \in \mathbb{I}_{=}(\alpha^{(i)}). \tag{86}$$

Similarly, for each $j \in \mathbb{I}_{<}(\alpha^{(i)})$, it holds that $\alpha_j^{(i)} = C$ by definition, and $\underline{\alpha}_j^{(i)} = 0$ by (81), which gives $\xi_j^{(i)} = 1 - \hat{y}_j^{(i)}(\omega^{(i)\top} \phi(\hat{X}_j^{(i)}) + b^{(i)} \geq 0$ by (83), (84) and (85). Therefore,

$$b^{(i)} \leq \hat{y}_j^{(i)} - \omega^{(i)\top} \phi(\hat{X}_j^{(i)}),\ j \in \mathbb{I}_{<}(\alpha^{(i)}), \tag{87}$$

which gives $b^{(i)} \in [b_{min}^{(i)}, b_{max}^{(i)}]$ by the definition of $b_{min}^{(i)},\ b_{max}^{(i)}$ in (11) and the definition of $\mathbb{H}^{(i)}(X, \gamma)$ in Proposition 2.

Next, we will show that the lower-level function is a constant function with respect to $b^{(i)}$. Indeed, the optimal lower-level objective function can be written as follows (the first part $\|w^{(i)}\|^2$ is not related to $b^{(i)}$ and therefore ignored)

$$
\begin{aligned}
\mathbb{S}(b^{(i)}) &:= C \sum_{j\in[\hat{n}^{(i)}]} \xi_j^{(i)} \\
&= C \sum_{j\in\mathbb{I}_<(\alpha^{(i)})} \xi_j^{(i)} \\
&= C \sum_{j\in\mathbb{I}_<(\alpha^{(i)})} \left(1-\hat{y}_j^{(i)}\left(\omega^{(i)\top}\phi(\hat{X}_j^{(i)}+b^{(i)}\right)\right) \\
&= -C \sum_{j\in\mathbb{I}_<(\alpha^{(i)})} \hat{y}_j^{(i)} b^{(i)} + C \sum_{j\in\mathbb{I}_<(\alpha^{(i)})} \left(1-\hat{y}_j^{(i)}\omega^{(i)\top}\phi(\hat{X}_j^{(i)}\right) \\
&= C \sum_{j\in\mathbb{I}_<(\alpha^{(i)})} \left(1-\hat{y}_j^{(i)}\omega^{(i)\top}\phi(\hat{X}_j^{(i)}\right)
\end{aligned}
$$

with the second equality based on (10), while the last equality is based on (82) and (10). Therefore, after obtaining $\alpha^{(i)}$ by solving the dual form of lower-level problem, i.e., (LLP^i), any $b^{(i)} \in [b_{min}^{(i)}, b_{max}^{(i)}]$ is an optimal solution of the primal problem (5). □

Appendix B. Proofs for Section 3

Proof of Proposition 6

It is obvious that the set of vectors in (14) only depends on I_1 and I_2, and does not depend on v. Therefore, for each $i \in [k]$, the vectors in (14) have the same rank (depending on I_1, I_2) for all vectors $(C, \gamma, \alpha^{(i)}) \in \mathcal{N}^{(i)}$, $i \in [k]$. That is, LCRCQ^i holds at $(\bar{C}, \bar{\gamma}, \bar{\alpha}^i)$, $i \in [k]$. Therefore, LCRCQ holds at $\bar{v}$. □

Proof for Theorem 7

First note that for each $i \in [k]$, the lower-level problem (LLP^i) is convex in the lower-level variable $\alpha^{(i)}$. Therefore, it suffices to observe that since all the lower-level constraints are linear w.r.t. $\alpha^{(i)}$ for all $i \in [k]$ and $C \geq 0$, it holds that $\Lambda(C, \gamma, \alpha) \neq \emptyset$ for all vectors $(C, \gamma, \alpha) \in \mathbb{R}_+^2 \times \mathbb{R}^n$ such that $\alpha^{(i)} \in S_D^{(i)}(C, \gamma)$ for $i \in [k]$. Hence, the proof of the result directly follows from Dempe and Dutta (2012). □

Proof of Proposition 9

If $C = 0$, then $\alpha^{(i)} = 0$ is the only feasible point of problem (LLP^i). In this case, the constraint for lower-level problem become

$$\alpha^{(i)} = 0,\ (\alpha^{(i)})^\top \hat{y}^{(i)} = 0$$

and therefore lead to the columns in $\mathcal{I} \in \mathbb{R}^{\hat{n}^{(i)} \times \hat{n}^{(i)}}$ and $\hat{y}^{(i)}$ are linearly dependent. This implies that LLICQ^i fails.

Next, for (ii), note that if $\mathbb{I}_=(\alpha^{(i)}) \cup \mathbb{I}_<(\alpha^{(i)}) = [\hat{n}^{(i)}]$, the set of vectors

$$\{e_j^{(i)} \mid j \in \mathbb{I}_=(\alpha^{(i)})\} \cup \{-e_j \mid j \in \mathbb{I}_<(\alpha^{(i)})\}$$

forms a basis of $\mathbb{R}^{\hat{n}^{(i)}}$. As a result, the set of vectors in (10) are linearly dependent. Therefore, LLICQ fails at $\alpha^{(i)}$. This gives (ii). A special case of (ii) is $\mathbb{I}_{=}(\alpha^{(i)}) = [\hat{n}^{(i)}]$ and $\mathbb{I}_{<}(\alpha^{(i)}) = \emptyset$. In this case, $\alpha^{(i)} = 0$. One can verify that $\alpha^{(i)} = 0$ is a feasible point of (LLP^i). Consequently, LLICQ fails at $\alpha^{(i)} = 0$, which gives (i).

If $\mathbb{I}_{=}(\alpha^{(i)}) \cup \mathbb{I}_{<}(\alpha^{(i)}) \neq [\hat{n}^{(i)}]$, one can see that the set of vectors in (10) are linearly independent, which implies that LLICQ holds at $\alpha^{(i)}$. □

Appendix C. Proofs for Section 4

We first introduce the set

$$(\Lambda_3^+)^c(\bar{v}) := [n] \backslash \Lambda_3^+ = (\Lambda_1 \cup \Lambda_2 \cup \Lambda_3^c \cup \Lambda_4)\,. \tag{88}$$

Proof of Proposition 18

(a) Based on Definition 10, we can write the set of gradient vectors in (20) at a feasible point $\bar{v}$ in the rows of the matrix Γ as

$$\Gamma = \begin{bmatrix} (\nabla G^1(\bar{v})_{J_{G^1}})^\top \\ (\nabla G^1(\bar{v})_{J_{GH^1}})^\top \\ (\nabla H^1(\bar{v})_{J_{H^1}})^\top \\ (\nabla H^1(\bar{v})_{J_{GH^1}})^\top \\ (\nabla G^2(\bar{v})_{J_{G^2}})^\top \\ (\nabla G^2(\bar{v})_{J_{GH^2}})^\top \\ (\nabla H^2(\bar{v})_{J_{H^2}})^\top \\ (\nabla (H^2(\bar{v})_{J_{GH^2}})^\top \\ (\nabla h(\bar{v}))^\top \\ (\nabla g_{I_g}(\bar{v}))^\top \end{bmatrix}. \tag{89}$$

Now, we can easily show that the matrix Γ in (89) is equivalent to the more specific form in (38). To proceed, first note that by (13), it holds that

$$H(v) = L^H v + b^H, \;\; h(v) = L^h v, \;\; g(v) = L^g(v),$$

where (noting that $m = 2n$)

$$L^H := \begin{bmatrix} \mathbf{0}_{n\times 1} & \mathbf{0}_{n\times 1} & \mathcal{I}_{n\times n} & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times k} \\ \mathbf{0}_{n\times 1} & \mathbf{0}_{n\times 1} & \mathbf{0}_{n\times n} & \mathcal{I}_{n\times n} & \mathbf{0}_{n\times k} \end{bmatrix} \in \mathbb{R}^{2n\times m}, \; b^H := \begin{bmatrix} -\mathbf{e}_n \\ \mathbf{0}_{n\times 1} \end{bmatrix} \in \mathbb{R}^{2n},$$

$$L^h := \begin{bmatrix} \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \widehat{Y} & \mathbf{0}_{k\times n} & \mathbf{0}_{k\times k} \end{bmatrix} \in \mathbb{R}^{k\times m},$$

$$L^g := \begin{bmatrix} \mathbf{0}_{I_g\times 1} & -\mathbf{e}_{I_g\times 1} & \mathbf{0}_{I_g\times n} & \mathbf{0}_{I_g\times n} & \mathbf{0}_{I_g\times k} \end{bmatrix} \in \mathbb{R}^{|I_g|\times m}.$$

Note that both $H(v)$ and $h(v)$ are linear. Consequently, it is easy to see that

$$(\nabla H(v))^\top = L^H \;\text{ and }\; (\nabla h(v))^\top = L^h.$$

However, $G(v)$ is nonlinear due to the kernel operator, and based on (16), it holds that

$$G^1(v) \quad = \quad Q(\gamma)\alpha - \mathbf{e}_n + \sigma + Pu \ \text{ with } \ P = \left[\begin{array}{ccc} \hat{y}^{(1)} & \cdots & \mathbf{0}_{\hat{n}^{(1)}\times 1} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{\hat{n}^{(k)}\times 1} & \cdots & \hat{y}^{(k)} \end{array}\right] \in \mathbb{R}^{n\times k},$$

and $Q(\gamma) \in \mathbb{R}^{n\times n}$ is defined by (39). Therefore, we have $G(v) = L^G(\gamma)v$, where

$$L^G(\gamma) := \left[\begin{array}{ccccc} \mathbf{0}_{n\times 1} & \mathbf{0}_{n\times 1} & Q(\gamma) & \mathcal{I}_{n\times n} & P \\ \mathbf{e}_{n\times 1} & \mathbf{0}_{n\times 1} & -\mathcal{I}_{n\times n} & \prime_{n\times n} & \mathbf{0}_{n\times k} \end{array}\right] \in \mathbb{R}^{2n\times m}.$$

For $\nabla G(v)$, it holds that

$$(\nabla G^2(v))^\top = \left[\begin{array}{ccccc} \mathbf{e}_{n\times 1} & \mathbf{0}_{n\times 1} & -\mathcal{I}_{n\times n} & \mathbf{0}_{n\times 0} & \mathbf{0}_{n\times k} \end{array}\right],$$

$$(\nabla G^1(v))^\top = \left[\begin{array}{ccccc} \mathbf{0}_{n\times 1} & \nabla_\gamma \theta(v) & Q(\gamma) & \mathcal{I}_{n\times n} & P \end{array}\right],$$

where $\nabla_\gamma \theta(v)$ is given as follows. Assume that i corresponds to the l_i-th sample in the training set of the $n^{(s_i)}$-th lower-level problem. That is, $i = \sum_{\ell=1}^{s_i-1} \hat{n}^{(\ell)} + l_i$. Hence,

$$(\nabla_\gamma \theta(v))_i = -\sum_{j=1}^{\hat{n}^{(s_i)}} y_{l_i} y_j \left\| X_{l_i}^{(s_i)} - X_j^{(s_i)} \right\|_2^2 \exp\left(-\gamma \left\| X_{l_i}^{(s_i)} - X_j^{(s_i)} \right\|_2^2\right) \alpha_j^{s_i}.$$

Finally, it is easy to verify that $P = \widehat{Y}^\top$. Hence, the form of the matrix Γ in (38).

(b) follows easily by the results in Proposition 16 and (a). □

Lemma 35. *The row vectors in the following matrix Γ^1 are positively linearly independent, where Γ^1 is the submatrix of Γ in* (38) *(in green) defined by*

$$\Gamma^1 := \left[\begin{array}{c} \mathcal{I}_{(J_{H^1},\cdot)} \\ \mathcal{I}_{(J_{GH^1},\cdot)} \\ \widehat{Y} \end{array}\right].$$

Proof Assume there exists nonzero nonnegative $\eta = (\eta_{J_{H^1}}, \eta_{J_{GH^1}}, \eta_5)$ such that $\eta^\top \Gamma^1 = 0$. Here, $\eta_{J_{H^1}}, \eta_{J_{GH^1}}, \eta_5$ correspond to the blocks of J_{H^1}, J_{GH^1}, $[k]$ in Γ^1. Then we have

$$0 = \eta^\top \Gamma^1 = \eta_{J_{H^1}}^\top \mathcal{I}_{(J_{H^1},\cdot)} + \eta_{J_{GH^1}}^\top \mathcal{I}_{(J_{GH^1},\cdot)} + \eta_5^\top \widehat{Y}. \tag{90}$$

First, we analyze the connections of J_{H^1} and J_{GH^1}. By Proposition 16, we have

$$J_{H^1} = \Lambda_2, J_{GH^1} = \Lambda_1.$$

(90) can be written as

$$\begin{aligned} 0 \quad &= \quad \eta_{J_{H^1}}^\top \mathcal{I}_{(\Lambda_2,\cdot)} + \eta_{J_{GH^1}}^\top \mathcal{I}_{(\Lambda_1,\cdot)} + \eta_5^\top \widehat{Y} \\ &= \quad \left[\begin{array}{c} \left(\eta_{J_{GH^1}}\right) \\ \left(\eta_{J_{H^1}}\right) \end{array}\right]^\top \mathcal{I}_{(\Lambda_1 \bigcup \Lambda_2,\cdot)} + \left((\eta_5)_1 \hat{y}^{(1)}, \cdots, (\eta_5)_k \hat{y}^{(k)}\right) \\ &:= \quad \hat{\eta}^\top \mathcal{I}_{(\Lambda_1 \bigcup \Lambda_2,\cdot)} + \left((\eta_5)_1 \hat{y}^{(1)}, \cdots, (\eta_5)_k \hat{y}^{(k)}\right). \end{aligned}$$

Denote $\hat{y} = (\hat{y}^{(1)}; \cdots; \hat{y}^{(k)}) \in \mathbb{R}^n$ be the labels of test data for all lower-level problems and

$$\widehat{\Lambda}_+^{(i)} := \left(\Lambda_1^{(i)} \bigcup \Lambda_2^{(i)}\right) \bigcap \left\{j \in [\hat{n}^{(i)}] \ : \ \hat{y}_j^{(i)} = 1\right\},$$

$$\widehat{\Lambda}_+ := \left(\Lambda_1 \bigcup \Lambda_2\right) \bigcap \{j \in [n] \ : \ \hat{y}_j = 1\}.$$

Let $\widehat{\Lambda}_c = [n] \backslash \widehat{\Lambda}_+$. By Proposition 17 and Definition 13, we have

$$\left(\Lambda_1^{(i)} \bigcup \Lambda_2^{(i)}\right) \bigcap \left\{j \in [\hat{n}^{(i)}] \ : \ y_j^{(i)} = 1\right\} \neq \emptyset$$

and $\mid \left(\Lambda_1^{(i)} \bigcup \Lambda_2^{(i)}\right) \mid \geq 2$, for each fold $i \in [k]$. It follows that

$$\widehat{\Lambda}_+ \neq \emptyset, \text{and } \Lambda_1 \bigcup \Lambda_2 = \widehat{\Lambda}_+ \bigcup \left(\widehat{\Lambda}_c \bigcap \left(\Lambda_1 \bigcup \Lambda_2\right)\right) := \widehat{\Lambda}_+ \bigcup \widehat{\Lambda}_0,$$

which gives

$$\hat{\eta}^\top \mathcal{I}_{(\Lambda_1 \bigcup \Lambda_2, \cdot)} = \hat{\eta}_{\widehat{\Lambda}_+}^\top \mathcal{I}_{(\widehat{\Lambda}_+, \cdot)} + \hat{\eta}_{\widehat{\Lambda}_0}^\top \mathcal{I}_{(\widehat{\Lambda}_0, \cdot)}.$$

Similarly, it holds that

$$\begin{aligned}
\left((\eta_5)_1 \hat{y}^{(1)}, \cdots, (\eta_5)_k \hat{y}^{(k)}\right) \quad &= \quad \left((\eta_5)_1 \hat{y}^{(1)}_{\hat{\Lambda}_+^{(1)}}, \mathbf{0}_{\hat{\Lambda}_c^{(1)}}, \cdots, (\eta_5)_k \hat{y}^{(k)}_{\hat{\Lambda}_+^{(k)}}, \mathbf{0}_{\hat{\Lambda}_c^{(k)}}\right) \\
&\quad + \left(\mathbf{0}_{\hat{\Lambda}_+^{(1)}}, (\eta_5)_1 \hat{y}^{(1)}_{\hat{\Lambda}_c^{(1)}}, \cdots, \mathbf{0}_{\hat{\Lambda}_+^{(k)}}, (\eta_5)_k \hat{y}^{(k)}_{\hat{\Lambda}_c^{(k)}}\right) \\
&:= \quad \hat{y}_{\hat{\Lambda}_+}(\eta_5) + \hat{y}_{\hat{\Lambda}_c}(\eta_5).
\end{aligned}$$

Therefore, we have

$$0 = \hat{\eta}_{\hat{\Lambda}_+}^\top \mathcal{I}_{(\widehat{\Lambda}_+, \cdot)} + \hat{y}_{\hat{\Lambda}_+}(\eta_5) + \hat{y}_{\hat{\Lambda}_c}(\eta_5) + \hat{\eta}_{\widehat{\Lambda}_0}^\top \mathcal{I}_{(\widehat{\Lambda}_0, \cdot)}.$$

This implies that

$$\hat{\eta}_{\hat{\Lambda}_+}^\top \mathcal{I}_{(\widehat{\Lambda}_+, \cdot)} + \hat{y}_{\hat{\Lambda}_+}(\eta_5) = 0 \ \text{ and } \ \hat{y}_{\hat{\Lambda}_c}(\eta_5) + \hat{\eta}_{\widehat{\Lambda}_0}^\top \mathcal{I}_{(\widehat{\Lambda}_0, \cdot)} = 0. \tag{91}$$

By the definition of $\widehat{\Lambda}_+$ and the nonnegativity of $\hat{\eta}_{\widehat{\Lambda}_+}$ as well as η_5, and noting that $\hat{y}^{(i)}_{\widehat{\Lambda}_+^{(i)}} = 1$, the first equation in (91) gives $\hat{\eta}_{\hat{\Lambda}_+} = 0$ and $\eta_5 = 0$. The second equation in (91) reduces to the following $\hat{\eta}_{\widehat{\Lambda}_0}^\top \mathcal{I}_{(\widehat{\Lambda}_0, \cdot)} = 0$, implying that $\hat{\eta}_{\widehat{\Lambda}_0} = 0$. Overall, we obtain that $\eta = 0$. In other words, the row vectors in Γ^1 are positively linearly independent. Hence, the proof. ■

Appendix D. Proofs for Section 5

Proof of Theorem 22

(i) Note that the number of columns in Γ is $R_1 = 2 + 2n + k$, while the number of rows in Γ can be obtained as

$$\begin{aligned} R_2 &= \mid J_{G^1} \mid + \mid J_{H^1} \mid +2 \mid J_{GH^1} \mid + \mid J_{G^2} \mid + \mid J_{H^2} \mid +2 \mid J_{GH^2} \mid +k+ \mid I_g \mid \\ &= 2n + k + \mid I_g \mid + \mid J_{GH^1} \mid + \mid J_{GH^2} \mid . \end{aligned}$$

The rank of Γ is denoted by $rank(\Gamma)$, which satisfies

$$rank(\Gamma) \le \min\{R_1, R_2\} = \min\{2 + 2n + k, 2n + k + \mid J_{GH^1} \mid + \mid J_{GH^2} \mid + \mid I_g \mid\}.$$

Note that $\mid I_g \mid$ is either 0 or 1. If (46) holds, it holds that

$$2n + k + \mid J_{GH^1} \mid + \mid J_{GH^2} \mid + \mid I_g \mid > 2 + 2n + k.$$

As a result, it holds that $rank(\Gamma) \le 2 + 2n + k$, meaning that the row vectors in Γ are linearly independent. In other words, the MPEC-LICQ fails at $\bar{v}$.

(ii) Since $\mid I_g \mid = 0$, therefore, by similar argument as in (i), we have

$$2n + k + \mid J_{GH^1} \mid + \mid J_{GH^2} \mid + \mid I_g \mid > 2 + 2n + k,$$

implying that $rank(\Gamma) \le 2 + 2n + k$. Therefore, the row vectors in Γ are linearly dependent. This implies that the MPEC-LICQ fails at $\bar{v}$. $\square$

Proof of Lemma 38

By multiplying the first (resp. second, thrid, fourth) row block from the left by $-Y_{([k],\Lambda_2}$ (resp. $-Y_{([k],\,\Lambda_1}$, $Y_{([k],\,\Lambda_4}$, $Y_{([k],\,\Lambda_3^c}$), and add it to the fifth row block, we reach the matrix

$$\mathcal{P}^1 = \begin{bmatrix} \mathbf{0}_{(\Lambda_2,\, L_1)} & \mathcal{I}_{(\Lambda_2,\, \cdot)} \\ \mathbf{0}_{(\Lambda_1,\, L_1)} & \mathcal{I}_{(\Lambda_1,\, \cdot)} \\ \mathbf{e}_{(\Lambda_4,\, L_1)} & -\mathcal{I}_{(\Lambda_4,\, \cdot)} \\ \mathbf{e}_{(\Lambda_3^c,\, L_1)} & -\mathcal{I}_{(\Lambda_3^c,\, \cdot)} \\ \widehat{Y}_{([k],\, \Lambda_4\cup\Lambda_2^c)}\mathbf{e}_{\Lambda_4\cup\Lambda_3^c} & \mathcal{Y}^1 \end{bmatrix} \quad \text{with } \mathcal{Y}^1 := \left[\mathbf{0}_{([k],\, \Lambda_2\cup\Lambda_1\cup\Lambda_4\cup\Lambda_3^c)},\ \widehat{Y}_{([k],\, \Lambda_3^+)}\right].$$

For $\rho = (\rho_1, \cdots, \rho_5)$, let $\rho_1, \cdots,\ \rho_5$ correspond to the first row block to the fifth row block in $\mathcal{P}^1$. If ρ satisfies $\rho^\top \mathcal{P}^1 = 0$, it holds that

$$\rho_5^\top Y_{([k],\Lambda_4\cup\Lambda_2^c)}\mathbf{e}_{\Lambda_4\cup\Lambda_3^c} = 0, \tag{92}$$

$$\rho_1^\top \mathcal{I}_{(\Lambda_2,\cdot)} + \rho_2^\top \mathcal{I}_{(\Lambda_1,\cdot)} - \rho_3^\top \mathcal{I}_{(\Lambda_4,\cdot)} - \rho_4^\top \mathcal{I}_{(\Lambda_3^c,\cdot)} + \rho_5^\top \mathcal{Y}^1 = 0. \tag{93}$$

(i) If Assumption 3 holds, it holds that $rank(\widehat{Y}_{([k],\Lambda_3^+)}) = k$. Therefore, (93) gives that $\rho_j = 0$ for $j = 1, \ldots, 5$. In other words, the row vectors in $\mathcal{P}^1$ are linearly independent. Therefore, the row vectors in $\mathcal{P}$ are linearly independent.

(ii) By (93), it holds that $\rho_j = 0$ for $j = 1, \ldots, 4$. Hence, (93) reduces to $\rho_5^\top \mathcal{Y}^1 = 0$. If Assumption 4 holds, $rank(\widehat{Y}_{([k],\Lambda_3^+)}) = k - 1$. Therefore, $\widehat{Y}_{([k],\Lambda_3^+)}$ takes the following form (assuming that $K = \{1\}$):

$$\widehat{Y}_{([k],\Lambda_3^+)} = \begin{bmatrix} 0 & \mathbf{0}_{(1,\Lambda_3^{+2})} & \cdots & \mathbf{0}_{1,\Lambda_3^{+k}} \\ 0 & (\hat{y}^{(2)}_{\Lambda_3^{+2}})^\top & \cdots & \mathbf{0}_{1,\Lambda_3^{+k}} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \mathbf{0}_{(1,\Lambda_3^{+2})} & \cdots & (\hat{y}^{(k)}_{\Lambda_3^{+k}})^\top \end{bmatrix}.$$

Therefore, $\rho_5^\top \mathcal{Y}^1 = 0$ gives $(\rho_5)_j = 0,\ j = 2, \ldots, k$. On the other hand, by the definition of $s_i,\ i \in K$, (92) reduces to $(\rho^5)_j s_j = 0,\ j \in [k]$. Hence, $(rho_5)_1 = 0$. In other words, the row vectors in both $\mathcal{P}^1$ and $\mathcal{P}$ are linearly independent.

(iii) If Assumptions 3 and 4 fail, the row vectors in $\mathcal{P}$ are linearly dependent, given that the row vectors of

$$\begin{bmatrix} \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_4)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_4)} & \mathbf{0}_{(\Lambda_1,L_5)} \\ \mathbf{e}_{(\Lambda_4,L_1)} & \mathbf{0}_{(\Lambda_4,L_2)} & -\mathcal{I}_{(\Lambda_4,\cdot)} & \mathbf{0}_{(\Lambda_4,L_4)} & \mathbf{0}_{(\Lambda_4,L_5)} \\ \mathbf{e}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & -\mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_4)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\ \mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \widehat{Y} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \end{bmatrix}$$

obtained from Γ (40) are linearly dependent. This gives (iii). □

We need the following proposition and lemmas in order to prove Theorem 23.

Proposition 36. *The following holds for the lower-level problems:*

(i) *There are at least two lower-level subproblems; i.e., $k \geq 2$.*

(ii) *By Proposition 17 as well as Proposition 16, it holds that*

$$\mid J^{(i)}_{H^1} \cup J^{(i)}_{GH^1} \mid \geq 2,\ i \in [k].$$

Here $J^{(i)}_{H^1}$ (resp. $J^{(i)}_{GH^1}$) corresponds to J_{H^1} (resp. J_{GH^1}) in the i-th lower-level problem, $i \in [k]$. That is, for each lower-level problem, there exists at least one positive sample data which is an unsupported vector, and at least one negative sample data which is an unsupported vector.

(iii) $\mid J_{G^j} \mid + \mid J_{H^j} \mid + \mid J_{GH^j} \mid = n$ *for* $j = 1,\ 2$.

(iv) If for each lower-level problem, strict complementarity conditions hold for the corresponding KKT systems, then it holds that

$$J_{GH^j} = \emptyset,\ \mid J_{G^j} \mid + \mid J_{H^j} \mid = n,\ \text{ and } J_{H^j} = [n] \backslash J_{G^j} \text{ for } j = 1,\ 2.$$

(v) *For each* (LLP^i), $i \in [k]$, *Proposition* 21 *implies that* $\mid J^{(i)}_{G^1} \mid \geq 4,\ i \in [k]$ *where* $J^{(i)}_{G^1}$ *corresponds to* J_{G^1} *in the i-th lower-level problem,* $i \in [k]$. *Therefore,* $\mid J_{G^1} \mid \geq 4k$.

Lemma 37. *Let $\bar{v}$ be feasible to (MPEC-BHO) with corresponding Γ defined in Proposition 18. Assume that we do some basic row transformations to Γ and obtain $\overline{\Gamma}$. Then,*

(i) $rank(\Gamma) = rank(\overline{\Gamma})$,

(ii) *MPEC-LICQ holds at $\bar{v}$ if and only if the row vectors in $\overline{\Gamma}$ are linearly independent.*

Proof As the basic row transformations will not change the rank of a matrix, we obtain (i). For (ii), by the definition of MPEC-LICQ, this condition is satisfied at $\bar{v}$ if and only if the row vectors in Γ are linearly independent. Due to (i), the row vectors in Γ are linearly independent if and only if the row vectors in $\overline{\Gamma}$ are linearly independent. Therefore, (ii) holds. ■

To proceed next, let

$$\mathcal{P} := \begin{bmatrix} \mathbf{0}_{(\Lambda_2, L_1)} & \mathcal{I}_{(\Lambda_2, \cdot)} \\ \mathbf{0}_{(\Lambda_1, L_1)} & \mathcal{I}_{(\Lambda_1, \cdot)} \\ \mathbf{e}_{(\Lambda_4, L_1)} & -\mathcal{I}_{(\Lambda_4, \cdot)} \\ \mathbf{e}_{(\Lambda_3^c, L_1)} & -\mathcal{I}_{(\Lambda_3^c, \cdot)} \\ \mathbf{0}_{([k], L_1)} & \widehat{Y} \end{bmatrix}. \tag{94}$$

Lemma 38. *Let $\bar{v}$ be a feasible point of problem* (MPEC-BHO) *with $\bar{C} > 0$.*

(i) *If Assumption* 3 *holds, the row vectors in $\mathcal{P}$ are linearly independent.*

(ii) *If Assumption* 4 *holds, the row vectors in $\mathcal{P}$ are linearly independent.*

(iii) *If Assumptions* 3 *and* 4 *fail, the rows vector in $\mathcal{P}$ are linearly dependent.*

At a feasible point $\bar{v}$ of problem (MPEC-BHO), denote

$$\left\{ \begin{array}{lll} \dot{a}^1 & := & Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_4 \cup \Lambda_3^c)} \mathbf{e}_{|\Lambda_4 \cup \Lambda_3^c|}, \quad \dot{a}^2 := Q(\bar{\gamma})_{(\Lambda_3^c,\, \Lambda_4 \cup \Lambda_3^c)} \mathbf{e}_{|\Lambda_4 \cup \Lambda_3^c|}, \\ \dot{a}^3 & := & Q(\bar{\gamma})_{(\Lambda_4,\, \Lambda_4 \cup \Lambda_3^c)} \mathbf{e}_{|\Lambda_4 \cup \Lambda_3^c|}, \quad \dot{a}^4 := Q(\bar{\gamma})_{(\Lambda_1,\, \Lambda_4 \cup \Lambda_3^c)} \mathbf{e}_{|\Lambda_4 \cup \Lambda_3^c|}, \\ \dot{a}^5 & := & \dot{a}^2 - A^2 \dot{a}^1, \quad \dot{a}^6 := \dot{a}^3 - A^3 \dot{a}^1, \quad \dot{a}^7 = \dot{a}^4 - A^4 \dot{a}^1, \\ \dot{b}^1 & := & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^c} - A^2 \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+}, \\ \dot{b}^2 & := & \nabla_\gamma \theta(\bar{v})_{\Lambda_4} - A^3 \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+}, \\ \dot{b}^3 & := & \nabla_\gamma \theta(\bar{v})_{\Lambda_1} - A^4 \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+}, \\ \dot{U}^1 & := & P_{(\Lambda_4, \cdot)} - A^3 P_{(\Lambda_3^+, \cdot)}, \\ \dot{y}^1 & := & \widehat{Y}_{([k], \Lambda_4 \cup \Lambda_3^c)} \mathbf{e}_{|\Lambda_4 \cup \Lambda_3^c|}, \quad \dot{y}^2 := \dot{y}^1 - A^1 \dot{a}^1, \\ \dot{y}^3 & := & -A^1 \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+}. \end{array} \right. \tag{95}$$

Lemma 39. *Let $\bar{v}$ be a feasible point of problem (MPEC-BHO) that satisfies Assumption 2. $\mid J_{GH^2} \mid + \mid J_{GH^1} \mid \in \{0,1,2\}$. Define $\widehat{\Gamma}$ and $\overline{\Gamma}$ respectively as follows:*

$$\widehat{\Gamma} := \begin{bmatrix} \dot{a}^1 & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+} & \hat{Q} & \mathbf{0}_{(\Lambda_3^+, L_4)} & P_{(\Lambda_3^+, \cdot)} \\ \dot{a}^2 & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^c} & \hat{Q}^1 & \mathbf{0}_{(\Lambda_3^c, L_4)} & P_{(\Lambda_3^c, \cdot)} \\ \dot{a}^3 & \nabla_\gamma \theta(\bar{v})_{\Lambda_4} & \hat{Q}^2 & \mathcal{I}_{(\Lambda_4, \cdot)} & P_{(\Lambda_4, \cdot)} \\ a^4 & \nabla_\gamma \theta(\bar{v})_{\Lambda_1} & \hat{Q}^3 & \mathcal{I}_{(\Lambda_1, \cdot)} & P_{(\Lambda_1, \cdot)} \\ \mathbf{0}_{(\Lambda_2, L_1)} & \mathbf{0}_{(\Lambda_2, L_2)} & \mathcal{I}_{(\Lambda_2, \cdot)} & \mathbf{0}_{(\Lambda_2, L_4)} & \mathbf{0}_{(\Lambda_2, L_5)} \\ \mathbf{0}_{(\Lambda_1, L_1)} & \mathbf{0}_{(\Lambda_1, L_2)} & \mathcal{I}_{(\Lambda_1, \cdot)} & \mathbf{0}_{(\Lambda_1, L_4)} & \mathbf{0}_{(\Lambda_1, L_5)} \\ \mathbf{e}_{(\Lambda_4, L_1)} & \mathbf{0}_{(\Lambda_4, L_2)} & -\mathcal{I}_{(\Lambda_4, \cdot)} & \mathbf{0}_{(\Lambda_4, L_4)} & \mathbf{0}_{(\Lambda_4, L_5)} \\ \mathbf{e}_{(\Lambda_3^c, L_1)} & \mathbf{0}_{(\Lambda_3^c, L_2)} & -\mathcal{I}_{(\Lambda_3^c, \cdot)} & \mathbf{0}_{(\Lambda_3^c, L_4)} & \mathbf{0}_{(\Lambda_3^c, L_5)} \\ \mathbf{0}_{(\Lambda_1, L_1)} & \mathbf{0}_{(\Lambda_1, L_2)} & \mathbf{0}_{(\Lambda_1, L_3)} & \mathcal{I}_{(\Lambda_1, \cdot)} & \mathbf{0}_{(\Lambda_1, L_5)} \\ \mathbf{0}_{(\Lambda_2, L_1)} & \mathbf{0}_{(\Lambda_2, L_2)} & \mathbf{0}_{(\Lambda_2, L_3)} & \mathcal{I}_{(\Lambda_2, \cdot)} & \mathbf{0}_{(\Lambda_2, L_5)} \\ \mathbf{0}_{(\Lambda_3^+, L_1)} & \mathbf{0}_{(\Lambda_3^+, L_2)} & \mathbf{0}_{(\Lambda_3^+, L_3)} & \mathcal{I}_{(\Lambda_3^+, \cdot)} & \mathbf{0}_{(\Lambda_3^+, L_5)} \\ \mathbf{0}_{(\Lambda_3^c, L_1)} & \mathbf{0}_{(\Lambda_3^c, L_2)} & \mathbf{0}_{(\Lambda_3^c, L_3)} & \mathcal{I}_{(\Lambda_3^c, \cdot)} & \mathbf{0}_{(\Lambda_3^c, L_5)} \\ \dot{y}^1 & \mathbf{0}_{([k], L_2)} & \hat{Y}^1 & \mathbf{0}_{([k], L_4)} & \mathbf{0}_{([k], L_5)} \end{bmatrix}, \tag{96}$$

$$\overline{\Gamma} := \begin{bmatrix} \dot{a}^1 & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+} & \hat{Q} & \mathbf{0}_{(\Lambda_3^+, L_4)} & P_{(\Lambda_3^+, \cdot)} \\ \dot{a}^5 & \dot{b}^1 & \mathbf{0}_{(\Lambda_3^c, L_3)} & \mathbf{0}_{(\Lambda_3^c, L_4)} & U^1 \\ \dot{a}^6 & \dot{b}^2 & \mathbf{0}_{(\Lambda_4, L_3)} & \mathcal{I}_{(\Lambda_4, \cdot)} & U^2 \\ \dot{a}^7 & \dot{b}^3 & \mathbf{0}_{(\Lambda_1, L_3)} & \mathbf{0}_{(\Lambda_1, L_4)} & U^3 \\ \mathbf{0}_{(\Lambda_2, L_1)} & \mathbf{0}_{(\Lambda_2, L_2)} & \mathcal{I}_{(\Lambda_2, \cdot)} & \mathbf{0}_{(\Lambda_2, L_4)} & \mathbf{0}_{(\Lambda_2, L_5)} \\ \mathbf{0}_{(\Lambda_1, L_1)} & \mathbf{0}_{(\Lambda_1, L_2)} & \mathcal{I}_{(\Lambda_1, \cdot)} & \mathbf{0}_{(\Lambda_1, L_4)} & \mathbf{0}_{(\Lambda_1, L_5)} \\ \mathbf{e}_{(\Lambda_4, L_1)} & \mathbf{0}_{(\Lambda_4, L_2)} & -\mathcal{I}_{(\Lambda_4, \cdot)} & \mathbf{0}_{(\Lambda_4, L_4)} & \mathbf{0}_{(\Lambda_4, L_5)} \\ \mathbf{e}_{(\Lambda_3^c, L_1)} & \mathbf{0}_{(\Lambda_3^c, L_2)} & -\mathcal{I}_{(\Lambda_3^c, \cdot)} & \mathbf{0}_{(\Lambda_3^c, L_4)} & \mathbf{0}_{(\Lambda_3^c, L_5)} \\ \mathbf{0}_{(\Lambda_1, L_1)} & \mathbf{0}_{(\Lambda_1, L_2)} & \mathbf{0}_{(\Lambda_1, L_3)} & \mathcal{I}_{(\Lambda_1, \cdot)} & \mathbf{0}_{(\Lambda_1, L_5)} \\ \mathbf{0}_{(\Lambda_2, L_1)} & \mathbf{0}_{(\Lambda_2, L_2)} & \mathbf{0}_{(\Lambda_2, L_3)} & \mathcal{I}_{(\Lambda_2, \cdot)} & \mathbf{0}_{(\Lambda_2, L_5)} \\ \mathbf{0}_{(\Lambda_3^+, L_1)} & \mathbf{0}_{(\Lambda_3^+, L_2)} & \mathbf{0}_{(\Lambda_3^+, L_3)} & \mathcal{I}_{(\Lambda_3^+, \cdot)} & \mathbf{0}_{(\Lambda_3^+, L_5)} \\ \mathbf{0}_{(\Lambda_3^c, L_1)} & \mathbf{0}_{(\Lambda_3^c, L_2)} & \mathbf{0}_{(\Lambda_3^c, L_3)} & \mathcal{I}_{(\Lambda_3^c, \cdot)} & \mathbf{0}_{(\Lambda_3^c, L_5)} \\ \dot{y}^2 & \dot{y}^3 & \mathbf{0}_{([k], L_3)} & \mathbf{0}_{([k], L_4)} & \dot{y}^4 \end{bmatrix}. \tag{97}$$

Here, $\hat{Q}$ is defined as in (104), $\hat{Y}^1$ is given in (106), $\dot{a}^i$ for $i = 1, \ldots, 7$, while $\dot{b}^i$, $\hat{Q}^i$, and $\dot{y}^i$ for $i = 1, 2, 3$ are defined in (95).

(a) *It holds that $rank(\Gamma) = rank(\widehat{\Gamma})$.*

(b) *If $\Lambda_3^+ \neq \emptyset$, we further have the following results:*

  (i) *$rank(\Gamma) = rank(\overline{\Gamma})$,*

  (ii) *For $\rho = (\rho_1, \cdots, \rho_{13})$, let $\rho_1, \cdots,\ \rho_{13}$ correspond to the first row block to the thirteenth row block in $\overline{\Gamma}$. If ρ satisfies $\rho^\top \overline{\Gamma} = 0$, the following holds:*

$$\rho_i = 0 \ \textit{for} \ i = 1, 3, 5, 6, 7, 8, 9, 10, 11, 12 \tag{98}$$

and

$$\begin{aligned}
\rho_2^\top \dot a^5 + \rho_4^\top \dot a^7 + \rho_{13}^\top \dot y^2 &= 0, && (99)\\
\rho_2^\top \dot b^1 + \rho_4^\top \dot b^3 + \rho_{13}^\top \dot y^3 &= 0, && (100)\\
\rho_2^\top U^1 + \rho_4^\top U^2 + \rho_{13}^\top Z &= 0. && (101)
\end{aligned}$$

Here U^1*,* U^2 *and* Z *are defined in* (53).

(iii) *MPEC-LICQ holds at* $\bar v$ *if and only if the row vectors in the following matrix*

$$M := \begin{bmatrix} \dot a^5 & \dot b^1 & U^1 \\ \dot a^7 & \dot b^3 & U^2 \\ \dot y^2 & \dot y^3 & Z \end{bmatrix} \tag{102}$$

are linearly independent.

(iv) *If* $\bar v$ *satisfies Assumption 3,* $\dot y^3$ *defined by* (95) *is negative definite.*

(v) *Following (iv), it holds that* $rank(M) = rank(M^0)$*, where* M^0 *is defined by*

$$M^0 := \begin{bmatrix} a^1 & b^1 & 0 \\ a^2 & b^2 & 0 \\ \dot y^2 & \dot y^3 & Z \end{bmatrix}. \tag{103}$$

Here a^1*,* a^2*,* b^1*, and* b^2 *are defined as in* (53).

Proof (a) By Proposition 16 (b) and $\gamma > 0$, Γ in (40) takes the following form:

$$\Gamma = \begin{bmatrix}
\mathbf{0}_{(\Lambda_3^+,L_1)} & \nabla_\gamma\theta(\bar v)_{\Lambda_3^+} & Q(\bar\gamma)_{(\Lambda_3^+,\cdot\ )} & \mathcal{I}_{(\Lambda_3^+,\cdot)} & P_{(\Lambda_3^+,\cdot)} \\
\mathbf{0}_{(\Lambda_3^c,L_1)} & \nabla_\gamma\theta(\bar v)_{\Lambda_3^c} & Q(\bar\gamma)_{(\Lambda_3^c,\cdot\ )} & \mathcal{I}_{(\Lambda_3^c,\cdot)} & P_{(\Lambda_3^c,\cdot)} \\
\mathbf{0}_{(\Lambda_4\ L_1)} & \nabla_\gamma\theta(\bar v)_{\Lambda_4} & Q(\bar\gamma)_{(\Lambda_4,\cdot\ )} & \mathcal{I}_{(\Lambda_4,\cdot)} & P_{(\Lambda_4,\cdot)} \\
\mathbf{0}_{(\Lambda_1,L_1)} & \nabla_\gamma\theta(\bar v)_{\Lambda_1} & Q(\bar\gamma)_{(\Lambda_1,\cdot\ )} & \mathcal{I}_{(\Lambda_1,\cdot)} & P_{(\Lambda_1,\cdot)} \\
\mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_4)} & \mathbf{0}_{(\Lambda_2,L_5)} \\
\mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_4)} & \mathbf{0}_{(\Lambda_1,L_5)} \\
\mathbf{e}_{(\Lambda_4,L_1)} & \mathbf{0}_{(\Lambda_4,L_2)} & -\mathcal{I}_{(\Lambda_4,\cdot)} & \mathbf{0}_{(\Lambda_4,L_4)} & \mathbf{0}_{(\Lambda_4,L_5)} \\
\mathbf{e}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & -\mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_4)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\
\mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathbf{0}_{(\Lambda_1,L_3)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_5)} \\
\mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathbf{0}_{(\Lambda_2,L_3)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_5)} \\
\mathbf{0}_{(\Lambda_3^+,L_1)} & \mathbf{0}_{(\Lambda_3^+,L_2)} & \mathbf{0}_{(\Lambda_3^+,L_3)} & \mathcal{I}_{(\Lambda_3^+,\cdot)} & \mathbf{0}_{(\Lambda_3^+,L_5)} \\
\mathbf{0}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & \mathbf{0}_{(\Lambda_3^c,L_3)} & \mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\
\mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \widehat{Y} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)}
\end{bmatrix}.$$

Substracting the eleventh row block from the first row block, substracting the twelveth row block from the second row block, and substracting the ninth row block from the fourth row block, we reach the following matrix:

$$\Gamma^0 = \begin{bmatrix} \mathbf{0}_{(\Lambda_3^+,L_1)} & \nabla_\gamma\theta(\bar{v})_{\Lambda_3^+} & Q(\bar{\gamma})_{(\Lambda_3^+,\cdot\,)} & \mathbf{0}_{(\Lambda_3^+,L_4)} & P_{(\Lambda_3^+,\cdot)} \\ \mathbf{0}_{(\Lambda_3^c,L_1)} & \nabla_\gamma\theta(\bar{v})_{\Lambda_3^c} & Q(\bar{\gamma})_{(\Lambda_3^c,\cdot\,)} & \mathbf{0}_{(\Lambda_3^c,L_4)} & P_{(\Lambda_3^c,\cdot)} \\ \mathbf{0}_{(\Lambda_4,L_1)} & \nabla_\gamma\theta(\bar{v})_{\Lambda_4} & Q(\bar{\gamma})_{(\Lambda_4,\cdot\,)} & \mathcal{I}_{(\Lambda_4,\cdot)} & P_{(\Lambda_4,\cdot)} \\ \mathbf{0}_{(\Lambda_1,L_1)} & \nabla_\gamma\theta(\bar{v})_{\Lambda_1} & Q(\bar{\gamma})_{(\Lambda_1,\cdot\,)} & \mathbf{0}_{(\Lambda_1^+,L_4)} & P_{(\Lambda_1,\cdot)} \\ \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_4)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_4)} & \mathbf{0}_{(\Lambda_1,L_5)} \\ \mathbf{e}_{(\Lambda_4,L_1)} & \mathbf{0}_{(\Lambda_4,L_2)} & -\mathcal{I}_{(\Lambda_4,\cdot)} & \mathbf{0}_{(\Lambda_4,L_4)} & \mathbf{0}_{(\Lambda_4,L_5)} \\ \mathbf{e}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & -\mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_4)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\ \mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathbf{0}_{(\Lambda_1,L_3)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_5)} \\ \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathbf{0}_{(\Lambda_2,L_3)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{0}_{(\Lambda_3^+,L_1)} & \mathbf{0}_{(\Lambda_3^+,L_2)} & \mathbf{0}_{(\Lambda_3^+,L_3)} & \mathcal{I}_{(\Lambda_3^+,\cdot)} & \mathbf{0}_{(\Lambda_3^+,L_5)} \\ \mathbf{0}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & \mathbf{0}_{(\Lambda_3^c,L_3)} & \mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\ \mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \hat{Y} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \end{bmatrix}.$$

Moreover, it holds that $[n] = \Lambda_1 \cup \Lambda_2 \cup \Lambda_3^+ \cup \Lambda_3^c \cup \Lambda_4$. Next, we conduct the following process.

Process $\mathcal{A}(1,5)$: Multiplying the fifth ($j = 5$) row block by $-Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_2)}$ from the left and adding it to the first ($i = 1$) row block, we can obtain the first row block with $Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_2)}$ replaced by zero. We denote this process as $\mathcal{A}(1,5)$.

Process $\mathcal{A}(1,6)$: This leads to the first row block with $Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_1)}$ replaced by zero.

Process $\mathcal{B}(1,7)$: Multiplying the fifth ($j = 7$) row block by $Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_4)}$ from the left and adding it to the first ($i = 1$) row block, we can obtain the first row block with $Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_4)}$ replaced by zero. Meanwhile, $\mathbf{0}_{(\Lambda_3^+,L_1)}$ is replaced by $Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_4)}\mathbf{e}_{|\Lambda_4|}$.

Process $\mathcal{B}(1,8)$: This leads to the first row block with $Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^c)}$ replaced by zero, and $Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_4)}\mathbf{e}_{|\Lambda_4|}$ replaced by $Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_4\cup\Lambda_3^c)}\mathbf{e}_{|\Lambda_4\cup\Lambda_3^c|}$, which is exactly $\dot{a}^1$ defined in (95).

Now the first row block in Γ^0 is replaced by the following row block

$$\begin{bmatrix} \dot{a}^1 & \nabla_\gamma\theta(\bar{v})_{\Lambda_3^+} & \hat{Q} & \mathbf{0}_{(\Lambda_3^+,L_4)} & P_{(\Lambda_3^+,\cdot)} \end{bmatrix}$$

with

$$\hat{Q} = \left[\mathbf{0}_{(\Lambda_3^+,\Lambda_3^+)}, Q(\bar{\gamma})_{(\Lambda_3^+,(\Lambda_3^+)^c)}\right]. \tag{104}$$

Carrying on to conduct process $\mathcal{A}(2,5)$, $\mathcal{A}(2,6), \mathcal{B}(2,7)$, $\mathcal{B}(2,8), \mathcal{A}(3,5), \mathcal{A}(3,6)$, $\mathcal{B}(3,7)$, $\mathcal{B}(3,8), \mathcal{A}(4,5), \mathcal{A}(4,6)$, $\mathcal{B}(4,7), \mathcal{B}(4,8), \mathcal{A}(13,5)$, $\mathcal{A}(13,6), \mathcal{B}(13,7)$, and $\mathcal{B}(13,8)$, we arrive at the following matrix:

$$\begin{bmatrix}
\dot{a}^1 & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+} & \hat{Q} & \mathbf{0}_{(\Lambda_3^+, L_4)} & P_{(\Lambda_3^+, \cdot)} \\
\dot{a}^2 & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^c} & \hat{Q}^1 & \mathbf{0}_{(\Lambda_3^c, L_4)} & P_{(\Lambda_3^c, \cdot)} \\
\dot{a}^3 & \nabla_\gamma \theta(\bar{v})_{\Lambda_4} & \hat{Q}^2 & \mathcal{I}_{(\Lambda_4, \cdot)} & P_{(\Lambda_4, \cdot)} \\
\dot{a}^4 & \nabla_\gamma \theta(\bar{v})_{\Lambda_1} & \hat{Q}^3 & \mathcal{I}_{(\Lambda_1, \cdot)} & P_{(\Lambda_1, \cdot)} \\
\mathbf{0}_{(\Lambda_2, L_1)} & \mathbf{0}_{(\Lambda_2, L_2)} & \mathcal{I}_{(\Lambda_2, \cdot)} & \mathbf{0}_{(\Lambda_2, L_4)} & \mathbf{0}_{(\Lambda_2, L_5)} \\
\mathbf{0}_{(\Lambda_1, L_1)} & \mathbf{0}_{(\Lambda_1, L_2)} & \mathcal{I}_{(\Lambda_1, \cdot)} & \mathbf{0}_{(\Lambda_1, L_4)} & \mathbf{0}_{(\Lambda_1, L_5)} \\
\mathbf{e}_{(\Lambda_4, L_1)} & \mathbf{0}_{(\Lambda_4, L_2)} & -\mathcal{I}_{(\Lambda_4, \cdot)} & \mathbf{0}_{(\Lambda_4, L_4)} & \mathbf{0}_{(\Lambda_4, L_5)} \\
\mathbf{e}_{(\Lambda_3^c, L_1)} & \mathbf{0}_{(\Lambda_3^c, L_2)} & -\mathcal{I}_{(\Lambda_3^c, \cdot)} & \mathbf{0}_{(\Lambda_3^c, L_4)} & \mathbf{0}_{(\Lambda_3^c, L_5)} \\
\mathbf{0}_{(\Lambda_1, L_1)} & \mathbf{0}_{(\Lambda_1, L_2)} & \mathbf{0}_{(\Lambda_1, L_3)} & \mathcal{I}_{(\Lambda_1, \cdot)} & \mathbf{0}_{(\Lambda_1, L_5)} \\
\mathbf{0}_{(\Lambda_2, L_1)} & \mathbf{0}_{(\Lambda_2, L_2)} & \mathbf{0}_{(\Lambda_2, L_3)} & \mathcal{I}_{(\Lambda_2, \cdot)} & \mathbf{0}_{(\Lambda_2, L_5)} \\
\mathbf{0}_{(\Lambda_3^+, L_1)} & \mathbf{0}_{(\Lambda_3^+, L_2)} & \mathbf{0}_{(\Lambda_3^+, L_3)} & \mathcal{I}_{(\Lambda_3^+, \cdot)} & \mathbf{0}_{(\Lambda_3^+, L_5)} \\
\mathbf{0}_{(\Lambda_3^c, L_1)} & \mathbf{0}_{(\Lambda_3^c, L_2)} & \mathbf{0}_{(\Lambda_3^c, L_3)} & \mathcal{I}_{(\Lambda_3^c, \cdot)} & \mathbf{0}_{(\Lambda_3^c, L_5)} \\
\dot{y}^1 & \mathbf{0}_{([k], L_2)} & \hat{Y}^1 & \mathbf{0}_{([k], L_4)} & \mathbf{0}_{([k], L_5)}
\end{bmatrix},$$

where $\dot{a}^i$ for $i = 2, 3, 4$ and $\dot{y}^1$ are defined as in (95), and

$$\hat{Q}^1 := \left[\, Q(\bar{\gamma})_{(\Lambda_3^c,\, \Lambda_3^+)},\ \mathbf{0}_{(\Lambda_3^c,\, (\Lambda_3^+)^c)} \right], \quad \hat{Q}^2 := \left[Q(\bar{\gamma})_{(\Lambda_4,\, \Lambda_3^+)},\ \mathbf{0}_{(\Lambda_4,\, (\Lambda_3^+)^c)} \right], \tag{105}$$

$$\hat{Q}^3 := \left[Q(\bar{\gamma})_{(\Lambda_1,\, \Lambda_3^+)},\ \mathbf{0}_{(\Lambda_1, (\Lambda_3^+)^c)} \right], \quad \hat{Y}^1 := \left[\widehat{Y}_{([k],\, \Lambda_3^+)},\ \mathbf{0}_{([k],\, (\Lambda_3^+)^c)} \right]. \tag{106}$$

Therefore, we obtain $\widehat{\Gamma}$ in (96). (a) is proved.

(b) If $\Lambda_3^+ \neq \emptyset$, then we can conduct further row transformation on $\widehat{\Gamma}$. Note that by assumption, $Q(\gamma)_{(\Lambda_3^+, \Lambda_3^+)}$ is positive definite. Therefore, the rows in $Q(\gamma)_{(\Lambda_3^+, \Lambda_3^+)}$ can fully express the rows in $\widehat{Y}_{([k], \Lambda_3^+)}$ and the rows in $Q(\gamma)_{(\Lambda_4, \Lambda_3^+)}$, $Q(\gamma)_{(\Lambda_1, \Lambda_3^+)}$ and $Q(\gamma)_{(\Lambda_3^c, \Lambda_3^+)}$. In other words, by conducting basic row transformation, we can make $\hat{Y}^1$, $\hat{Q}^1$, $\hat{Q}^2$ and $\hat{Q}^3$ replaced by zeros. Specifically, there exists $A^1 \in \mathbb{R}^{k \times |\Lambda_3^+|}$, $A^2 \in \mathbb{R}^{|\Lambda_3^c| \times |\Lambda_3^+|}$, $A^3 \in \mathbb{R}^{|\Lambda_4| \times |\Lambda_3^+|}$ and $A^4 \in \mathbb{R}^{|\Lambda_1| \times |\Lambda_3^+|}$ satisfy (51) and (52).

By multiplying the first row block by $-A^1$ ($-A^2$, $-A^3$, $-A^4$ respectively) from left and add them to the last (first, second, third, respectively) row block, we obtain the following matrix in which $\hat{Q}^1$, $\hat{Q}^2$, $\hat{Q}^3$, and $\hat{Y}^1$ are replaced by zeros:

$$\begin{bmatrix}
\dot{a}^1 & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+} & \hat{Q} & \mathbf{0}_{(\Lambda_3^+, L_4)} & P_{(\Lambda_3^+, \cdot)} \\
\dot{a}^5 & \dot{b}^1 & \mathbf{0}_{(\Lambda_3^c, L_3)} & \mathbf{0}_{(\Lambda_3^c, L_4)} & U^1 \\
\dot{a}^6 & \dot{b}^2 & \mathbf{0}_{(\Lambda_4, L_3)} & \mathcal{I}_{(\Lambda_4, \cdot)} & \dot{U}^1 \\
\dot{a}^7 & \dot{b}^3 & \mathbf{0}_{(\Lambda_1, L_3)} & \mathbf{0}_{(\Lambda_1, L_4)} & U^2 \\
\mathbf{0}_{(\Lambda_2, L_1)} & \mathbf{0}_{(\Lambda_2, L_2)} & \mathcal{I}_{(\Lambda_2, \cdot)} & \mathbf{0}_{(\Lambda_2, L_4)} & \mathbf{0}_{(\Lambda_2, L_5)} \\
\mathbf{0}_{(\Lambda_1, L_1)} & \mathbf{0}_{(\Lambda_1, L_2)} & \mathcal{I}_{(\Lambda_1, \cdot)} & \mathbf{0}_{(\Lambda_1, L_4)} & \mathbf{0}_{(\Lambda_1, L_5)} \\
\mathbf{e}_{(\Lambda_4, L_1)} & \mathbf{0}_{(\Lambda_4, L_2)} & -\mathcal{I}_{(\Lambda_4, \cdot)} & \mathbf{0}_{(\Lambda_4, L_4)} & \mathbf{0}_{(\Lambda_4, L_5)} \\
\mathbf{e}_{(\Lambda_3^c, L_1)} & \mathbf{0}_{(\Lambda_3^c, L_2)} & -\mathcal{I}_{(\Lambda_3^c, \cdot)} & \mathbf{0}_{(\Lambda_3^c, L_4)} & \mathbf{0}_{(\Lambda_3^c, L_5)} \\
\mathbf{0}_{(\Lambda_1, L_1)} & \mathbf{0}_{(\Lambda_1, L_2)} & \mathbf{0}_{(\Lambda_1, L_3)} & \mathcal{I}_{(\Lambda_1, \cdot)} & \mathbf{0}_{(\Lambda_1, L_5)} \\
\mathbf{0}_{(\Lambda_2, L_1)} & \mathbf{0}_{(\Lambda_2, L_2)} & \mathbf{0}_{(\Lambda_2, L_3)} & \mathcal{I}_{(\Lambda_2, \cdot)} & \mathbf{0}_{(\Lambda_2, L_5)} \\
\mathbf{0}_{(\Lambda_3^+, L_1)} & \mathbf{0}_{(\Lambda_3^+, L_2)} & \mathbf{0}_{(\Lambda_3^+, L_3)} & \mathcal{I}_{(\Lambda_3^+, \cdot)} & \mathbf{0}_{(\Lambda_3^+, L_5)} \\
\mathbf{0}_{(\Lambda_3^c, L_1)} & \mathbf{0}_{(\Lambda_3^c, L_2)} & \mathbf{0}_{(\Lambda_3^c, L_3)} & \mathcal{I}_{(\Lambda_3^c, \cdot)} & \mathbf{0}_{(\Lambda_3^c, L_5} \\
\dot{y}^2 & \dot{y}^3 & \mathbf{0}_{([k], L_3)} & \mathbf{0}_{([k], L_4)} & Z
\end{bmatrix},$$

which is exactly $\overline{\Gamma}$. Here, $\dot{a}^j$ for $j = 5, 6, 7$, $\dot{b}^j$ for $j = 1, 2, 3$, $\dot{y}^j$ for $j = 2, 3$, and $\dot{U}^1$ are given by (95), while U^1, U^2, and Z are defined in (53). Therefore, we have (i) from

$$rank(\Gamma) = rank(\Gamma^0) = rank(\widehat{\Gamma}) = rank(\overline{\Gamma}).$$

(ii) For $\rho = (\rho_1, \cdots, \rho_{13})$ such that $\rho^\top \overline{\Gamma} = 0$, $0 = \rho^\top \overline{\Gamma} := [S_1\ S_2\ S_3\ S_4\ S_5]$. Hence,

$$\begin{aligned}
S_1 &= \rho_1^\top \dot{a}^1 + \rho_2^\top \dot{a}^5 + \rho_3^\top \dot{a}^6 + \rho_4^\top \dot{a}^7 + \rho_7^\top \mathbf{e}_{(\Lambda_4, L_1)} + \rho_8^\top \mathbf{e}_{(\Lambda_3^c, L_1)} + \rho_{13}^\top \dot{y}^2 = 0, && (107)\\
S_2 &= \rho_1^\top \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+} + \rho_2^\top \dot{b}^1 + \rho_3^\top \dot{b}^2 + \rho_4^\top \dot{b}^3 + \rho_{13}^\top \dot{y}^3 = 0, && (108)\\
S_3 &= \rho_1^\top \hat{Q} + \rho_5^\top \mathcal{I}_{(\Lambda_2, \cdot)} + \rho_6^\top \mathcal{I}_{(\Lambda_1, \cdot)} - \rho_7^\top \mathcal{I}_{(\Lambda_4, \cdot)} - \rho_8^\top \mathcal{I}_{(\Lambda_3^c, \cdot)} = 0, && (109)\\
S_4 &= \rho_3^\top \mathcal{I}_{(\Lambda_4, \cdot)} + \rho_9^\top \mathcal{I}_{(\Lambda_1, \cdot)} + \rho_{10}^\top \mathcal{I}_{(\Lambda_2, \cdot)} + \rho_{11}^\top \mathcal{I}_{(\Lambda_3^+, \cdot)} + \rho_{12}^\top \mathcal{I}_{(\Lambda_3^c, \cdot)} = 0, && (110)\\
S_5 &= \rho_1^\top P_{(\Lambda_3^+, \cdot)} + \rho_2^\top U^1 + \rho_3^\top \dot{U}^1 + \rho_4^\top U^2 + \rho_{13}^\top Z = 0. && (111)
\end{aligned}$$

By (109) and (110), we have $\rho_i = 0$ for $i = 1, 3, 5, 6, 7, 8, 9, 10, 11, 12$. Hence, (107), (108) and (111) reduce to (99)-(101). This gives (ii).

(iii) Assume that $\rho = (\rho_1, \cdots, \rho_{13})$ satisfies $\rho^\top \overline{\Gamma} = 0$, by (ii), the row vectors in $\overline{\Gamma}$ are linearly dependent if and only if we can derive $\rho_i = 0$ for $i = 2, 4, 13$ from (99)-(101). Having $\rho_i = 0$ for $i = 2, 4, 13$ is equivalent to that the row vectors in M are linearly independent. Therefore, (iii) is proved.

(iv) If Assumption 3 holds, let us first show $rank(A^1) = [k]$. By Assumption 3,

$$\mid \Lambda_3^+ \mid = \sum_{i=1}^{k} \mid \Lambda_3^{+(i)} \mid \geq k. \tag{112}$$

Therefore, $\widehat{Y}_{([k], \Lambda_3^+)}$ takes the following form

$$\widehat{Y}_{([k], \Lambda_3^+)} = \begin{bmatrix} \left(\hat{y}^{(1)}_{\Lambda_3^{+(1)}}\right)^\top & \cdots & \mathbf{0}_{(1, \Lambda_3^{+(k)})} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{(1, \Lambda_3^{+(1)})} & \cdots & \left(\hat{y}^{(k)}_{(1, \Lambda_3^{+(k)})}\right)^\top \end{bmatrix},$$

implying that $rank\left(\widehat{Y}_{([k], \Lambda_3^+)}\right) = k$. On the other hand, it holds that

$$rank\left(\widehat{Y}_{([k], \Lambda_3^+)}\right) \leq \min\left\{rank(A^1), rank\left(Q(\gamma)_{(\Lambda_3^+, \Lambda_3^+)}\right)\right\} = \left\{rank(A^1), \mid \Lambda_3^+ \mid\right\}.$$

Together with (112), $rank(A^1) \geq k$. Recall that for $A^1 \in \mathbb{R}^{k \times |\Lambda_3^+|}$, we obtain $rank(A^1) = k$.

Next, we will show that $-Z$ is definite. To proceed, recall that $P = \widehat{Y}^\top$ in Proposition 18, together with (51), we have

$$Z = -A^1 P_{(\Lambda_3^+, \cdot)} = -A^1 \left(\widehat{Y}_{([k], \Lambda_3^+)}\right)^\top = -A^1 Q(\bar{\gamma})_{(\Lambda_3^+, \Lambda_3^+)} \left(A^1\right)^\top.$$

By the positive definiteness of $Q(\gamma)_{(\Lambda_3^+, \Lambda_3^+)}$ as well as $rank(A^1) = k$, $-Z$ is positive definite.

(v) Following (iv), let B^1, B^2 be the coefficient matrix such that (54) and (55) hold. By multiplying the third row block in M by $-B^1$ and adding it to the first row block, then multiplying the third row block in M by $-B^2$ and adding it to the second row block, we can transfer M to the following matrix:

$$\begin{bmatrix} \dot{a}^5 - B^1\dot{y}^2 & \dot{b}^1 - B^1\dot{y}^3 & 0 \\ \dot{a}^7 - B^2\dot{y}^2 & \dot{b}^3 - B^2\dot{y}^3 & 0 \\ \dot{y}^2 & \dot{y}^3 & Z \end{bmatrix}.$$

By the definition of $\dot{a}^i$ for $i = 5, 7$, $\dot{y}^i$ for $i = 2, 3$, and $\dot{b}^i$ for $i = 1, 3$, we reach M^0. By Lemma 37 (i), M and M^0 have the same rank. Hence, (iv) is proved. ■

Proof of Theorem 23

By Proposition 16 and $| J_{GH^1} | + | J_{GH^2} |= 0$, we have the following relationship for different index sets:

$$\begin{aligned} \Lambda_3^c = \emptyset,\ \Lambda_1 = \emptyset,\ J_{GH^1} = \emptyset,\ J_{GH^2} = \emptyset,\ \text{and}\ \Lambda_3 = \Lambda_3^+; \\ J_{H^1} = \Lambda_2,\ J_{G^1} = \Lambda_3^+ \cup \Lambda_4,\ J_{H^2} = \Lambda_2 \cup \Lambda_3^+,\ \text{and}\ J_{G^2} = \Lambda_4. \end{aligned} \tag{113}$$

As a result, it holds that

$$[n] = \Lambda_4 \cup \Lambda_2 \cup \Lambda_3 \ \text{and}\ (\Lambda_3^+)^c = \Lambda_4 \cup \Lambda_2. \tag{114}$$

By Lemma 39 (a), $\widehat{\Gamma}$ reduces the form

$$\widehat{\Gamma} = \begin{bmatrix} \dot{a}^1 & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+} & \hat{Q} & \mathbf{0}_{(\Lambda_3^+, L_4)} & P_{(\Lambda_3^+,\cdot)} \\ \dot{a}^3 & \nabla_\gamma \theta(\bar{v})_{\Lambda_4} & \hat{Q}^2 & \mathcal{I}_{(\Lambda_4,\cdot)} & P_{(\Lambda_u,\cdot)} \\ \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_4)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{e}_{(\Lambda_4,L_1)} & \mathbf{0}_{(\Lambda_4,L_2)} & -\mathcal{I}_{(\Lambda_4,\cdot)} & \mathbf{0}_{(\Lambda_4,L_4)} & \mathbf{0}_{(\Lambda_4,L_5)} \\ \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathbf{0}_{(\Lambda_2,L_3)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{0}_{(\Lambda_3^+,L_1)} & \mathbf{0}_{(\Lambda_3^+,L_2)} & \mathbf{0}_{(\Lambda_3^+,L_3)} & \mathcal{I}_{(\Lambda_3^+,\cdot)} & \mathbf{0}_{(\Lambda_3^+,L_5)} \\ \dot{y}^1 & \mathbf{0}_{([k],L_2)} & \hat{Y}^1 & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \end{bmatrix}.$$

(i) If $J_{G_1} = J_{G_2}$, together with (113) and (114), it holds that

$$\Lambda_3^+ = \emptyset,\ [n] = \Lambda_4 \cup \Lambda_2,\ \text{ and }\ (\Lambda_3^+)^c = [n].$$

$\Lambda_3^+ = \emptyset$ implies that Assumptions 3 and 4 fail. By Lemma 38, MPEC-LICQ fails at $\bar{v}$.

(ii) Due to $J_{G^2} = \emptyset$, (113) and (114) reduce to the following $\Lambda_4 = \emptyset$, $J_{H^1} = \Lambda_2$, $J_{G^1} = \Lambda_3^+$, $J_{H^2} = \Lambda_2 \cup \Lambda_3^+$, $J_{G^2} = \emptyset$, $[n] = \Lambda_2 \cup \Lambda_3^+$, $(\Lambda_3^+)^c = \Lambda_2$. By Lemma 36 (b) (v), $\Lambda_3^+ \neq \emptyset$. By Lemma 39 (b) (i), $\overline{\Gamma}$ takes the following form

$$\overline{\Gamma} = \begin{bmatrix} \dot{a}^1 & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+} & \hat{Q} & \mathbf{0}_{(\Lambda_3^+, L_4)} & P_{(\Lambda_3^+,\cdot)} \\ \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_4)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathbf{0}_{(\Lambda_2,L_3)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{0}_{(\Lambda_3^+,L_1)} & \mathbf{0}_{(\Lambda_3^+,L_2)} & \mathbf{0}_{(\Lambda_3^+,L_3)} & \mathcal{I}_{(\Lambda_3^+,\cdot)} & \mathbf{0}_{(\Lambda_3^+,L_5)} \\ \dot{y}^2 & \dot{y}^3 & \mathbf{0}_{([k],L_3)} & \mathbf{0}_{([k],L_4)} & Z \end{bmatrix}.$$

with $\dot{y}^2$, $\dot{a}^1$ in (95) and $\hat{Q}$ in (104) reduce to the following form

$$\dot{y}^2 = \mathbf{0}, \ \ \dot{a}^1 = \mathbf{0}, \ \text{ and } \ \hat{Q} = \left[Q(\bar{\gamma})_{(\Lambda_3^+,\, \Lambda_3^+)}, \ \mathbf{0}_{(\Lambda_3^+,\, \Lambda_2)} \right].$$

By Lemma 39 (b) (ii), ρ_i for $i = 2, 3, 4, 6, 7, 8, 9, 12$ do not appear in ρ. Hence, the system (98)-(100) reduces to

$$\rho_i = 0 \ \text{ for } \ i = 1, 5, 10, 11, \ \ \rho_{13}^\top \dot{y}^3 = 0, \ \text{ and } \ \rho_{13}^\top Z = 0.$$

If Assumption 3 holds, by Lemma 39 (b) (iv), $-Z$ is positive definite, we obtain that $\rho_{13} = 0$. It implies that the row vectors in $\overline{\Gamma}$ are linearly independent. By Lemma 37 (ii), MPEC-LICQ holds at $\bar{v}$.

If Assumption 3 fails, by the definition of K, $K \neq \emptyset$. Note that $\Lambda_4 = \emptyset$ and $\Lambda_3^c = \emptyset$, it holds that $S_i = 0$, $i \in K$. In other words, Assumption 4 fails as well. By Lemma 38 (iii), MPEC-LICQ fails at $\bar{v}$.

(iii) If $J_{G^2} \subset J_{G^1}$ and $J_{G^2} \neq \emptyset$, it holds that $\Lambda_u \neq \emptyset$ and $\Lambda_3^+ \neq \emptyset$. By Lemma 39 (b) (i), $\overline{\Gamma}$ reduces to the following form

$$\overline{\Gamma} = \begin{bmatrix} a^1 & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+} & \hat{Q} & \mathbf{0}_{(\Lambda_3^+, L_4)} & P_{(\Lambda_3^+, \cdot)} \\ \dot{a}^6 & \dot{b}^2 & \mathbf{0}_{(\Lambda_4, L_3)} & \mathcal{I}_{(\Lambda_4, \cdot)} & \dot{U}^1 \\ \mathbf{0}_{(\Lambda_2, L_1)} & \mathbf{0}_{(\Lambda_2, L_2)} & \mathcal{I}_{(\Lambda_2, \cdot)} & \mathbf{0}_{(\Lambda_2, L_4)} & \mathbf{0}_{(\Lambda_2, L_5)} \\ \mathbf{e}_{(\Lambda_4, L_1)} & \mathbf{0}_{(\Lambda_4, L_2)} & -\mathcal{I}_{(\Lambda_4, \cdot)} & \mathbf{0}_{(\Lambda_4, L_4)} & \mathbf{0}_{(\Lambda_4, L_5)} \\ \mathbf{0}_{(\Lambda_2, L_1)} & \mathbf{0}_{(\Lambda_2, L_2)} & \mathbf{0}_{(\Lambda_2, L_3)} & \mathcal{I}_{(\Lambda_2, \cdot)} & \mathbf{0}_{(\Lambda_2, L_5)} \\ \mathbf{0}_{(\Lambda_3^+, L_1)} & \mathbf{0}_{(\Lambda_3^+, L_2)} & \mathbf{0}_{(\Lambda_3^+, L_3)} & \mathcal{I}_{(\Lambda_3^+, \cdot)} & \mathbf{0}_{(\Lambda_3^+, L_5)} \\ \dot{y}^2 & \dot{y}^3 & \mathbf{0}_{([k], L_3)} & \mathbf{0}_{([k], L_4)} & Z \end{bmatrix}.$$

By Lemma 39 (b) (ii), ρ_i for $i = 2, 4, 6, 8, 9, 12$ do not appear in ρ. Hence, (98)–(100) reduces to $\rho_i = 0$ for $i = 1, 3, 5, 7, 10, 11$ and

$$\rho_{13}^\top \dot{y}^2 = 0, \ \rho_{13}^\top \dot{y}^3 = 0, \ \text{ and } \ \rho_{13}^\top Z = 0. \tag{115}$$

If Assumption 3 holds, by Lemma 39 (b) (iv), $-Z$ is positive definite. As a result, we have $\rho_{13} = 0$. Therefore, we obtain $\rho = 0$. It implies that the row vectors in $\overline{\Gamma}$ are linearly independent. By Lemma 37 (ii), MPEC-LICQ holds at $\bar{v}$.

If Assumption 4 holds, we will derive $\rho_{13} = 0$ by (115) under Assumption 4. Since $K \neq \emptyset$, without loss of generality, assume that $K = \{1, \cdots, l\}$. Therefore, for $i \in K$, it holds that $\Lambda_3^{+(i)} = \emptyset$. Note that from $J_{G^1}^{(i)} = \Lambda_3^{+(i)} \cup \Lambda_4^{(i)}$ and Lemma 36 (v), $J_{G^1}^{(i)} = \Lambda_4^{(i)}$, $\mid \Lambda_4^{(i)} \mid \geq 4$, $i \in K$. As a result, $J_{GH^2} = \Lambda_3^c = \emptyset$ and $\mid J_{G^2} \mid = \mid \Lambda_4 \mid \geq 4k$. It follows that

$$\widehat{Y}_{(i,\Lambda_4)} \mathbf{e}_{|\Lambda_4|} = (\hat{y}^{(i)}_{\Lambda_4^{(i)}})^\top \mathbf{e}_{|\Lambda_4^{(i)}|} = (\hat{y}^{(i)}_{J_{G^2}^{(i)}})^\top \mathbf{e}_{|J_{G^2}^{(i)}|} = s_i \ \text{ for } \ i \in K.$$

$\widehat{Y}_{([k],\Lambda_3^+)}$ reduces to the expression

$$\widehat{Y}_{([k],\Lambda_3^+)} = \begin{bmatrix} \mathbf{0}_{(1,\Lambda_3^{+(l+1)})} & \mathbf{0}_{(1,\Lambda_3^{+(l+2)})} & \cdots & \mathbf{0}_{(1,\Lambda_3^{+(k)})} \\ \cdots & \cdots & \ddots & \cdots \\ \mathbf{0}_{(1,(\Lambda_3^+)^{(l+1)})} & \mathbf{0}_{(1,(\Lambda_3^+)^{(l+2)})} & \cdots & \mathbf{0}_{(1,(\Lambda_3^+)^{(k)})} \\ \left(\hat{y}^{(l+1)}_{(1,\Lambda_3^{+(l+1)})}\right)^\top & \mathbf{0}_{(1,\Lambda_3^{+(l+2)})} & \cdots & \mathbf{0}_{(1,\Lambda_3^{+(k)})} \\ \cdots & \cdots & \ddots & \cdots \\ \mathbf{0}_{(1,\Lambda_3^{+(1)})} & \mathbf{0}_{(1,\Lambda_3^{+(2)})} & \cdots & \left(\hat{y}^{(k)}_{(1,\Lambda_3^{+(k)})}\right)^\top \end{bmatrix}. \tag{116}$$

Therefore, $A^1 \in \mathbb{R}^{k\times|\Lambda_3^+|}$ takes the partition

$$A^1 = \begin{bmatrix} \mathbf{0}_{(l,|\Lambda_3^+|)} \\ A_2^1 \end{bmatrix}. \tag{117}$$

Moreover, recall $\dot{y}^3$ in (95), it holds that

$$\dot{y}^3 = -A^1\nabla_\gamma\theta(\bar{v})_{\Lambda_3} = \begin{bmatrix} \mathbf{0}_{(K,1)} \\ -A_2^1\nabla_\gamma\theta(\bar{v})_{\Lambda_3} \end{bmatrix} := \begin{bmatrix} \mathbf{0}_{(K,1)} \\ \dot{y}^3_{K^c} \end{bmatrix}, \tag{118}$$

and

$$Z = -A^1P_{(\Lambda_3,\cdot)} = -\begin{bmatrix} \mathbf{0}_{(K,K)} & \mathbf{0}_{(K,K^c)} \\ \mathbf{0}^\top_{(K,K^c)} & A_2^1Q(\bar{\gamma})_{(\Lambda_3,\Lambda_3)}\left(A_2^1\right)^\top \end{bmatrix} := \begin{bmatrix} \mathbf{0}_{(K,K)} & \mathbf{0}_{(K,K^c)} \\ \mathbf{0}^\top_{(K,K^c)} & Z_{(K^c,K^c)} \end{bmatrix}. \tag{119}$$

Let ρ_{13} take the following partition

$$\rho_{13} = \begin{bmatrix} \rho_K \\ \rho_{K^c} \end{bmatrix}. \tag{120}$$

The equations in (115) reduces to the following equations

$$\rho_K^\top s_K + \rho_{K^c}^\top\dot{y}^2_{K^c} = 0, \rho_{K^c}^\top\dot{y}^3_{K^c} = 0,\ \rho_{K^c}^\top Z_{(K^c,K^c)} = 0. \tag{121}$$

Similar to the argument in Lemma 39 (b) (iv), we can obtain that $rank(A_2^1) =\mid K^c \mid$, and $rank(Z_{(K^c,K^c)}) =\mid K^c \mid$, implying that $-Z_{(K^c,K^c)}$ is positive definite. Therefore, we obtain $\rho_{K^c} = 0$, and (121) reduces to $\rho_K^\top s_K = 0$.

Note that $\mid K \mid\geq 1$. If $\mid K \mid= 1$ and $s_i \neq 0, i \in K$, it holds that $\rho_K = 0$. In this case, we obtain that $\rho = 0$. It implies that the row vectors in $\overline{\Gamma}$ are linearly independent. By Lemma 37 (ii), MPEC-LICQ holds at $\bar{v}$.

If both Assumptions 3 and 4 fail, by Lemma 38, MEPC-LICQ fails at $\bar{v}$. $\square$

Proof of Theorem 24

(i) By Lemma 39 (b) (iii) and (v), if Assumption 3 holds, we can easily obtain (i).

For (ii), similar to the proof in Theorem 23 (iii), since $K \neq \emptyset$, without loss of generality, assume that $K = \{1, \cdots, l\}$. Then for $i \in K$, it holds that $\Lambda_3^{+(i)} = \emptyset$. As a result,

$$\widehat{Y}_{(i,\Lambda_4)}\mathbf{e}_{|\Lambda_4|} + \widehat{Y}_{(i,\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} = \left(\hat{y}^{(i)}_{\Lambda_4^{(i)}}\right)^\top \mathbf{e}_{|\Lambda_4^{(i)}|} + \left(\hat{y}^{(i)}_{\Lambda_3^{+(i)}}\right)^\top \mathbf{e}_{|\Lambda_3^{+(i)}|} = s_i \text{ for } i \in K.$$

Moreover, $\widehat{Y}_{([k],\Lambda_3^+)}$ takes the form in (116). Therefore, we have $A^1 \in \mathbb{R}^{k\times|\Lambda_3^+|}$ takes the form in (117). By the definition of $\dot{y}^2$, it holds that

$$\begin{aligned} \dot{y}^2 &= \begin{bmatrix} \widehat{Y}_{(K,\Lambda_4)}\mathbf{e}_{|\Lambda_4|} + \widehat{Y}_{(K,\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} \\ \widehat{Y}_{(K^c,\Lambda_4)}\mathbf{e}_{|\Lambda_4|} + \widehat{Y}_{(K^c,\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} \end{bmatrix} - \begin{bmatrix} \mathbf{0}_{(K,1)} \\ A_2^1 Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^+)}\mathbf{e}_{|\Lambda_3^+|} \end{bmatrix} \\ &:= \begin{bmatrix} s_K \\ \dot{y}^2_{K^c} \end{bmatrix}. \end{aligned} \tag{122}$$

Similarly, we obtain the partition of $\dot{y}^3$ and Z, which are the same as in (118) and (119). Hence M reduces to the form

$$\dot{M}^1 = \begin{bmatrix} a^1 & b^1 & U^1_{(\cdot,K)} & U^1_{(\cdot,K^c)} \\ a^2 & b^2 & U^2_{(\cdot,K)} & U^2_{(\cdot,K^c)} \\ s_K & \mathbf{0}_{(K,1)} & \mathbf{0}_{(K,K)} & \mathbf{0}_{(K,K^c)} \\ y^2_{K^c} & y^3_{K^c} & \mathbf{0}_{(K^c,K)} & Z_{(K^c,K^c)} \end{bmatrix}.$$

Similar to the argument in Lemma 39 (b) (iv), we can obtain that $-Z_{(K^c,K^c)}$ is positive definite. Therefore, there exist B^3 and B^4 such that (54) holds. By multiplying $-B^3$ from the left to the fourth row block and add it to the first row block, multiplying $-B^4$ from the left to the fourth row block and add it to the second row block, we obtain

$$\dot{M}^2 := \begin{bmatrix} \dot{a}^8 & b^3 & U^1_{(\cdot,K)} & \mathbf{0}_{(\Lambda_3^c,K^c)} \\ \dot{a}^9 & b^4 & U^2_{(\cdot,K)} & \mathbf{0}_{(\Lambda_4,K^c)} \\ s_K & \mathbf{0}_{(K,1)} & \mathbf{0}_{(K,K)} & \mathbf{0}_{(K,K^c)} \\ \dot{y}^2_{K^c} & \dot{y}^3_{K^c} & \mathbf{0}_{(K^c,K)} & Z_{(K^c,K^c)} \end{bmatrix}.$$

Here b^3, b^4 are defined as in (53) and

$$\dot{a}^8 = a^1 - B^3\dot{y}^2_{K^c}, \dot{a}^9 = a^2 - B^4\dot{y}^2_{K^c}.$$

By Assumption 4, for $K = 1$ and $s_K \neq 0$, we can similarly do row transformation to $\dot{M}^2$ to make $\dot{a}^8, \dot{a}^9, \dot{y}^2_{K^c}$ zero. Then we obtain the following

$$\dot{M}^3 := \begin{bmatrix} \mathbf{0}_{(\Lambda_3^c,1)} & b^3 & U^1_{(\cdot,K)} & \mathbf{0}_{(\Lambda_3^c,K^c)} \\ \mathbf{0}_{(\Lambda_4,1)} & b^4 & U^2_{(\cdot,K)} & \mathbf{0}_{(\Lambda_4,K^c)} \\ s_K & \mathbf{0}_{(K,1)} & \mathbf{0}_{(K,K)} & \mathbf{0}_{(K,K^c)} \\ \mathbf{0}_{([k],1)} & \dot{y}^3_{K^c} & \mathbf{0}_{(K^c,K)} & Z_{(K^c,K^c)} \end{bmatrix}.$$

Next, for $\bar{\rho} = (\rho_1, \cdots, \rho_4)$, let $\rho_1, \cdots,\ \rho_4$ correspond to the first row block to the fourth row block in $\dot{M}^3$. If ρ satisfies $\rho^\top \dot{M}^3 = 0$, it holds that

$$\rho_3^\top s_K = 0, \rho_1^\top b^3 + \rho_2^\top b^4 + \rho_4^\top \dot{y}^3_{K^c} = 0,$$

$$\rho_1^\top U^1_{(\cdot,K)} + \rho_2^\top U^2_{(\cdot,K)} = 0, \rho_4^\top Z_{(K^c,K^c)} = 0.$$

We obtain that $\rho_3 = 0, \rho_4 = 0$ by $s_K \neq 0$ and the positive definiteness of $-Z_{(K^c,K^c)}$. The above equations reduce to the following

$$\rho_1^\top b^3 + \rho_2^\top b^4 = 0,\ \rho_1^\top U^1_{(\cdot,K)} + \rho_2^\top U^2_{(\cdot,K)} = 0.$$

Therefore, it can be easy to obtain the following results.

If $\mid J_{GH^1} \mid + \mid J_{GH^2} \mid = 2$ and the matrix M^1 defined in (ii) is nonsingular, MPEC-LICQ holds. If $\mid J_{GH^1} \mid = 1$, $\mid J_{GH^2} \mid = 0$, it holds that $\mid \Lambda_1 \mid = 1, \Lambda_3^c = \emptyset$. If $(b^4, U^2_{(\cdot,K)}) \neq 0$, MPEC-LICQ holds. If $\mid J_{GH^2} \mid = 1$, $\mid J_{GH^1} \mid = 0$, it holds that $\mid \Lambda_3^c \mid = 1$, $\Lambda_1 = \emptyset$. If $(b^3, U^1_{(\cdot,K)}) \neq 0$, MPEC-LICQ holds.

Otherwise, it can be seen that MPEC-LICQ fails. This concludes the proof. $\square$

Proof of Theorem 25

Since $\mid I_g \mid \neq \emptyset$, by Proposition 18 (b), we have Γ as in (40). By substracting the eleventh (resp. twelfth, ninth) row block from the first (resp. second, fourth) row block, by multiplying the last blcok by $-\nabla_\gamma \theta(\bar{v})_{(\Lambda_3^+,\cdot)}$ (resp. $-\nabla_\gamma \theta(\bar{v})_{(\Lambda_3^c,\cdot)}$, $-\nabla_\gamma \theta(\bar{v})_{(\Lambda_4,\cdot)}$, $-\nabla_\gamma \theta(\bar{v})_{(\Lambda_1,\cdot)}$) and adding it to the first (resp. second, third, fourth) block, we reach the matrix

$$\Gamma^0 = \begin{bmatrix}
\mathbf{0}_{(\Lambda_3^+,L_1)} & \mathbf{0}_{(\Lambda_3^+,L_2)} & Q(\bar{\gamma})_{(\Lambda_3^+,\cdot\ )} & \mathbf{0}_{(\Lambda_3^+,L_4)} & P_{(\Lambda_3^+,\cdot)} \\
\mathbf{0}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & Q(\bar{\gamma})_{(\Lambda_3^c,\cdot\ )} & \mathbf{0}_{(\Lambda_3^c,L_4)} & P_{(\Lambda_3^c,\cdot)} \\
\mathbf{0}_{(\Lambda_4,L_1)} & \mathbf{0}_{(\Lambda_4,L_2)} & Q(\bar{\gamma})_{(\Lambda_4,\cdot\ )} & \mathcal{I}_{(\Lambda_4,\cdot)} & P_{(\Lambda_4,\cdot)} \\
\mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & Q(\bar{\gamma})_{(\Lambda_1,\cdot\ )} & \mathbf{0}_{(\Lambda_1^+,L_4)} & P_{(\Lambda_1,\cdot)} \\
\mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_4)} & \mathbf{0}_{(\Lambda_2,L_5)} \\
\mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_4)} & \mathbf{0}_{(\Lambda_1,L_5)} \\
\mathbf{e}_{(\Lambda_4,L_1)} & \mathbf{0}_{(\Lambda_4,L_2)} & -\mathcal{I}_{(\Lambda_4,\cdot)} & \mathbf{0}_{(\Lambda_4,L_4)} & \mathbf{0}_{(\Lambda_4,L_5)} \\
\mathbf{e}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & -\mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_4)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\
\mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathbf{0}_{(\Lambda_1,L_3)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_5)} \\
\mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathbf{0}_{(\Lambda_2,L_3)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_5)} \\
\mathbf{0}_{(\Lambda_3^+,L_1)} & \mathbf{0}_{(\Lambda_3^+,L_2)} & \mathbf{0}_{(\Lambda_3^+,L_3)} & \mathcal{I}_{(\Lambda_3^+,\cdot)} & \mathbf{0}_{(\Lambda_3^+,L_5)} \\
\mathbf{0}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & \mathbf{0}_{(\Lambda_3^c,L_3)} & \mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\
\mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \widehat{Y} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \\
\mathbf{0}_{(I_g,L_1)} & -\mathbf{e}_{(I_g,L_2)} & \mathbf{0}_{(I_g,L_3)} & \mathbf{0}_{(I_g,L_4)} & \mathbf{0}_{(I_g,L_5)}
\end{bmatrix}.$$

We discuss the following two cases:

(i) If Assumption 5 holds, without loss of generality, assume that

$$\mid \Lambda_3^{(i)} \cup \Lambda_4^{(i)} \mid = 1 \text{ and } \Lambda_3^{(i)} \cup \Lambda_4^{(i)} = l_i \ \text{ for } \ i = 1, \cdots, l \text{ with } l \leq k.$$

The submatrix T defined by

$$T := [T_1,\ T_2] \quad \text{with } T_1 := \begin{bmatrix} Q(\bar{\gamma})_{(\Lambda_3^+,\,\cdot\,)} \\ Q(\bar{\gamma})_{(\Lambda_3^c,\cdot\,)} \\ Q(\bar{\gamma})_{(\Lambda_1,\cdot\,)} \end{bmatrix} \quad \text{and } T_2 := \begin{bmatrix} P_{(\Lambda_3^+,\cdot)} \\ P_{(\Lambda_3^c,\,\cdot)} \\ P_{(\Lambda_1,\,\cdot)} \end{bmatrix}$$

takes the following form:

$$T = \begin{bmatrix} \hat{y}_{l_1}^{(1)}(\hat{y}^{(1)})^\top & \mathbf{0}_{(1,\hat{n}^{(2)})} & \cdots & 0 & \hat{y}_{l_1}^{(1)} & 0 & 0 \\ \mathbf{0}_{(1,\hat{n}^{(1)})} & \hat{y}_{l_2}^{(2)}(\hat{y}^{(2)})^\top & \cdots & \mathbf{0}_{(1,\hat{n}^{(k)})} & 0 & \hat{y}_{l_2}^{(2)} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{(1,\hat{n}^{(1)})} & \mathbf{0}_{(1,\hat{n}^{(2)})} & \cdots & \hat{y}_{l_r}^{(r)}(\hat{y}^{(r)})^\top & 0 & \cdots & \hat{y}_{l_r}^{(r)} \end{bmatrix}.$$

We can observe that T has full row rank and also

$$rank(T) = rank(T_1) = rank(T_2) = l. \tag{123}$$

Note that $\bar{\gamma} = 0$ and in this situation, $Q^i(\bar{\gamma}) = \hat{y}^{(i)}\left(\hat{y}^{(i)}\right)^\top$, which is a rank one matrix, $i \in [k]$. Due to the special structure of $Q^i(\bar{\gamma})$, we can eliminate $Q(\bar{\gamma})_{(\Lambda_3^+,\cdot)}$, $Q(\bar{\gamma})_{(\Lambda_3^c,\cdot)}$, $Q(\bar{\gamma})_{(\Lambda_4,\cdot)}$, $Q(\bar{\gamma})_{(\Lambda_1,\cdot)}$ by applying a proper row transformation based on the thirteenth block where three is $\widehat{Y}$. Then conducting $\mathcal{A}(13,5)$, $\mathcal{A}(13,6)$, $\mathcal{B}(13,7)$, and $\mathcal{B}(13,8)$, the matrix becomes (recalling that $[n] = \Lambda_1 \cup \Lambda_2 \cup \Lambda_3^+ \cup \Lambda_3^c \cup \Lambda_4$):

$$\Gamma^1 \quad = \quad \begin{bmatrix} \mathbf{0}_{(\Lambda_3^+,L_1)} & \mathbf{0}_{(\Lambda_3^+,L_2)} & \mathbf{0}_{(\Lambda_3^+,L_3)} & \mathbf{0}_{(\Lambda_3^+,L_4)} & P_{(\Lambda_3^+,\cdot)} \\ \mathbf{0}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & \mathbf{0}_{(\Lambda_3^+,L_3)} & \mathbf{0}_{(\Lambda_3^c,L_4)} & P_{(\Lambda_3^c,\cdot)} \\ \mathbf{0}_{(\Lambda_4,L_1)} & \mathbf{0}_{(\Lambda_4,L_2)} & \mathbf{0}_{(\Lambda_4,L_3)} & \mathcal{I}_{(\Lambda_4,\cdot)} & P_{(\Lambda_4,\cdot)} \\ \mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathbf{0}_{(\Lambda_1,L_3)} & \mathbf{0}_{(\Lambda_1,L4)} & P_{(\Lambda_1,\cdot)} \\ \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_4)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_4)} & \mathbf{0}_{(\Lambda_1,L_5)} \\ \mathbf{e}_{(\Lambda_4,L_1)} & \mathbf{0}_{(\Lambda_4,L_2)} & -\mathcal{I}_{(\Lambda_4,\cdot)} & \mathbf{0}_{(\Lambda_4,L_4)} & \mathbf{0}_{(\Lambda_4,L_5)} \\ \mathbf{e}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & -\mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_4)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\ \mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathbf{0}_{(\Lambda_1,L_3)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_5)} \\ \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathbf{0}_{(\Lambda_2,L_3)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{0}_{(\Lambda_3^+,L_1)} & \mathbf{0}_{(\Lambda_3^+,L_2)} & \mathbf{0}_{(\Lambda_3^+,L_3)} & \mathcal{I}_{(\Lambda_3^+,\cdot)} & \mathbf{0}_{(\Lambda_3^+,L_5)} \\ \mathbf{0}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & \mathbf{0}_{(\Lambda_3^c,L_3)} & \mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\ \dot{y}^1 & \mathbf{0}_{([k],L_2)} & \hat{Y}^1 & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \\ \mathbf{0}_{(I_g,L_1)} & -\mathbf{e}_{(I_g,L_2)} & \mathbf{0}_{(I_g,L_3)} & \mathbf{0}_{(I_g,L_4)} & \mathbf{0}_{(I_g,L_5)} \end{bmatrix},$$

where $(\Lambda_3^+)^c$, $\hat{Y}^1$, and $\dot{y}^1$ are defined as in (88), (106), (95) respectively. Therefore,

$$rank(\Gamma) = rank(\Gamma^0) = rank(\Gamma^1).$$

For $\rho = (\rho_1, \cdots, \rho_{14})$ such that $0 = \rho^\top \Gamma^1 := [S_1\ S_2\ S_3\ S_4\ S_5]$, it holds that

$$\begin{aligned}
S_1 &= \rho_7^\top \mathbf{e}_{(\Lambda_4, L_1)} + \rho_8^\top \mathbf{e}_{(\Lambda_3^c, L_1)} + \rho_{13}^\top \dot{y}^1 = 0, && (124)\\
S_2 &= -\rho_{14} = 0, && (125)\\
S_3 &= \rho_5^\top \mathcal{I}_{(\Lambda_2,\cdot)} + \rho_6^\top \mathcal{I}_{(\Lambda_1,\cdot)} - \rho_7^\top \mathcal{I}_{(\Lambda_4,\cdot)} - \rho_8^\top \mathcal{I}_{(\Lambda_3^c,\cdot)} + \rho_{13}^\top \hat{Y}^1 = 0, && (126)\\
S_4 &= \rho_3^\top \mathcal{I}_{(\Lambda_4,\cdot)} + \rho_9^\top \mathcal{I}_{(\Lambda_1,\cdot)} + \rho_{10}^\top \mathcal{I}_{(\Lambda_2,\cdot)} + \rho_{11}^\top \mathcal{I}_{(\Lambda_3^+,\cdot)} + \rho_{12}^\top \mathcal{I}_{(\Lambda_3^c,\cdot)} = 0, &&\\
S_5 &= \rho_1^\top P_{(\Lambda_3^+,\cdot)} + \rho_2^\top P_{(\Lambda_3^c,\cdot)} + \rho_3^\top P_{(\Lambda_2,\cdot)} + \rho_4^\top P_{(\Lambda_1,\cdot)} = 0. && (127)
\end{aligned}$$

By (125), (126), and (127), we have $\rho_i = 0$ for $i = 3, 5, 6, 7, 8, 9, 10, 11, 12, 14$. Hence, (127) reduces to the following form:

$$\rho_1^\top P_{(\Lambda_3^+,\cdot)} + \rho_2^\top P_{(\Lambda_3^c,\cdot)} + \rho_4^\top P_{(\Lambda_1,\cdot)} = 0.$$

By (123), the row vectors in T_2 are linearly independent, given that $\rho_i = 0$ for $i = 1, 2, 4$.

Recall the definition of $\hat{Y}^1$ in (106), (124), and (126) reduce to

$$\rho_{13}^\top \dot{y}^1 = 0, \rho_{13}^\top \widehat{Y}_{([k],\Lambda_3^+)} = 0.$$

By the definition of K, without loss of geneարity, let $K^c = \{1, \cdots, l\}$ and $\Lambda_3^{+(i)} = j_i, i \in K^c$. Let ρ_{13} takes the partition as in (120). We know that by Assumption 5, $\widehat{Y}_{(k,\Lambda_3^+)}$ can be partitioned as follows:

$$Y_{(k,\Lambda_3^+)} = \begin{bmatrix} \hat{y}_{j_1}^{(1)} & 0 & \cdots & 0 \\ 0 & \hat{y}_{j_2}^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{y}_{j_l}^{(l)} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} := \begin{bmatrix} \bar{y}_{K^c} \\ \mathbf{0}_{(k-l,1)} \end{bmatrix}.$$

If Assumption 3 holds, $\mid K \mid = 0$. Therefore, $K^c = [k]$, and $Y_{([k],\Lambda_3^+)} = \bar{y}_{K^c}$. It gives that $\rho_{13}^\top Y_{([k],\Lambda_3^+)} = ((\rho_{13})_1 \hat{y}_{j_1}^{(1)}, \cdots, (\rho_{13})_k \hat{y}_{j_k}^{(k)}) = 0$. We obtain that ρ_{13}=0. MPEC-LICQ holds.

If Assumption 4 holds, it holds that $\mid K \mid = 1$. $\rho_{13}^\top Y_{([k,\Lambda_3^+])} = 0$ reduces to

$$\rho_{K^c}^\top \bar{y}_{K^c} = \left((\rho_{13})_1 \hat{y}_{j_1}^{(1)}, \cdots, (\rho_{13})_{k-1} \hat{y}_{j_{k-1}}^{(k-1)} \right) = 0,$$

which gives $\rho_{K^c} = 0$. Note that

$$\dot{y}^1 \;=\; \widehat{Y}_{([k],\Lambda_3^c\cup\Lambda_4)} \mathbf{e}_{|\Lambda_3^c\cup\Lambda_4|} = \widehat{Y}_{([k],\Lambda_4)} \mathbf{e}_{|\Lambda_4|} = \begin{bmatrix} \widehat{Y}_{(K,\Lambda_4)} \mathbf{e}_{|\Lambda_4|} \\ \widehat{Y}_{(K^c,\Lambda_4)} \mathbf{e}_{|\Lambda_4|} \end{bmatrix} := \begin{bmatrix} s_K \\ \dot{y}_{K^c}^1 \end{bmatrix}.$$

Hence, $\rho_{13}^\top \dot{y}^1 = 0$ reduces to $\rho_K^\top s_K = 0$. By $s_K \neq 0$, we obtain that $\rho_K = 0$. Therefore, the MPEC-LICQ holds at $\bar{v}$.

If both Assumptions 3 and 4 fail, by Lemma 38 (iii), MPEC-LICQ fails at $\bar{v}$.

(ii) If Assumption 5 fails, we have that the submatrix T has at least two row vectors which are linearly dependent. To show this, without loss of generality, assume that there exists one $i \in [k]$, say $i = 1$, with $1, 2 \in \Lambda_3^{(1)} \cup \Lambda_1^{(1)}$. Then we have that two rows in T take the following form:

$$\begin{bmatrix} \hat{y}_1^{(1)}(\hat{y}^{(1)})^\top & \mathbf{0}_{(1,n^{(2)})} & \cdots & \mathbf{0}_{(1,n^{(k)})} & \hat{y}_1^{(1)} & 0 & \mathbf{0}_{(1,k-2)} \\ \hat{y}_2^{(1)}(\hat{y}^{(1)})^\top & \mathbf{0}_{(1,n^{(2)})} & \cdots & \mathbf{0}_{(1,n^{(k)})} & \hat{y}_2^{(1)} & 0 & \mathbf{0}_{(1,k-2)} \end{bmatrix}.$$

The two row vectors are linearly dependent. Therefore, we can find two row vectors in Γ such that they are linearly dependent. Therefore, in this case, MPEC-LICQ fails. $\square$

Appendix E. Proofs for Section 6

Proof of Proposition 26

Note that we have

$$\begin{aligned} I_{0+} &= J_{G^1} \cup \{n + J_{G^2}\} = \Lambda_3 \cup \Lambda_4 \cup \{n + \Lambda_4\}, \\ I_{+0} &= J_{H^1} \cup \{n + J_{H^2}\} = \Lambda_2 \cup \{n + (\Lambda_1 \cup \Lambda_2 \cup \Lambda_3^+)\}, \\ I_{00} &= J_{GH^1} \cup \{n + J_{GH^2}\} = \Lambda_1 \cup \{n + \Lambda_3^c\}. \end{aligned}$$

By the definition of weak stationary point, it holds that

$$\begin{aligned} &\bar{\eta}_i = 0 \ \text{ for } \ i \in I_{+0} = \Lambda_2 \cup \{n + (\Lambda_1 \cup \Lambda_2 \cup \Lambda_3^+)\}, \\ &\bar{w}_i = 0 \ \text{ for } \ i \in I_{0+} = \Lambda_3 \cup \Lambda_4 \cup \{n + \Lambda_4\}. \end{aligned}$$

Hence, we get (59). $\square$

In the next result, we provide some precise representations of $\nabla^2_{vv}\mathbb{L}(\cdot)$ and $\mathcal{C}(\cdot)$ in the context of our bilevel hyperparameter optimization problem from the perspective of the KKT/MPEC reformulation in (MPEC-BHO).

Proposition 40. *If $\bar{v}$ be a weakly stationary point $\bar{v}$ of problem* (MPEC-BHO) *with associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$, the following statements hold true:*

(i) $\nabla^2_{vv}\mathbb{L}(v, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}) = \nabla^2_{vv} f(v) - \Delta(v)$ *with*

$$\nabla^2_{vv} f(v) = \begin{bmatrix} \nabla^2_{CC} f(v) & \nabla^2_{C\gamma} f(v) & \nabla^2_{C\alpha} f(v) & \mathbf{0}_{(L_1,L_4)} & \mathbf{0}_{(L_1,L_5)} \\ \nabla^2_{\gamma C} f(v) & \nabla^2_{\gamma\gamma} f(v) & \nabla^2_{\gamma\alpha} f(v) & \mathbf{0}_{(L_2,L_4)} & \mathbf{0}_{(L_2,L_5)} \\ \nabla^2_{C\alpha} f(v)^\top & \nabla^2_{\gamma\alpha} f(v)^\top & \nabla^2_{\alpha\alpha} f(v) & \mathbf{0}_{(L_3,L_4)} & \mathbf{0}_{(L_3,L_5)} \\ \mathbf{0}_{(L_4,L_1)} & \mathbf{0}_{(L_4,L_2)} & \mathbf{0}_{(L_4,L_3)} & \mathbf{0}_{(L_4,L_4)} & \mathbf{0}_{(L_4,L_5)} \\ \mathbf{0}_{(L_5,L_1)} & \mathbf{0}_{(L_5,L_2)} & \mathbf{0}_{(L_5,L_3)} & \mathbf{0}_{(L_5,L_4)} & \mathbf{0}_{(L_5,L_5)} \end{bmatrix}$$

and

$$\Delta(v) := \begin{bmatrix} \mathbf{0}_{(L_1,L_1)} & \mathbf{0}_{(L_1,L_2)} & \mathbf{0}_{(L_1,L_3)} & \mathbf{0}_{(L_1,L_4)} & \mathbf{0}_{(L_1,L_5)} \\ \mathbf{0}_{(L_2,L_1)} & \sum\limits_{i \in \Lambda_1 \cup \Lambda_3} \bar{\eta}_i^1 \nabla^2_{\gamma\gamma}\theta_i(v) & \sum\limits_{i \in \Lambda_1 \cup \Lambda_3} \bar{\eta}_i^1 \nabla^2_{\gamma\alpha}\theta_i(v) & \mathbf{0}_{(L_2,L_4)} & \mathbf{0}_{(L_2,L_5)} \\ \mathbf{0}_{(L_3,L_1)} & \sum\limits_{i \in \Lambda_1 \cup \Lambda_3} \bar{\eta}_i^1 \nabla^2_{\gamma\alpha}\theta_i(v)^\top & \mathbf{0}_{(L_3,L_3)} & \mathbf{0}_{(L_3,L_4)} & \mathbf{0}_{(L_3,L_5)} \\ \mathbf{0}_{(L_4,L_1)} & \mathbf{0}_{(L_4,L_2)} & \mathbf{0}_{(L_4,L_3)} & \mathbf{0}_{(L_4,L_4)} & \mathbf{0}_{(L_4,L_5)} \\ \mathbf{0}_{(L_5,L_1)} & \mathbf{0}_{(L_5,L_2)} & \mathbf{0}_{(L_5,L_3)} & \mathbf{0}_{(L_5,L_4)} & \mathbf{0}_{(L_5,L_5)} \end{bmatrix}.$$

(ii) $\mathcal{C}(\bar{v}) = \Big\{ d := \big(d^C,\ d^\gamma,\ (d^\alpha)^\top,\ (d^\sigma)^\top,\ (d^u)^\top\big)^\top \in \mathbb{R}^m : \quad Wd = 0 \Big\}$, *where*

$$W := \begin{bmatrix} \mathbf{0}_{(\Lambda_1^{\eta^1}, L_1)} & \nabla_\gamma \theta_{\Lambda_1^{\eta^1}}(\bar{v}) & Q(\bar{\gamma})_{(\Lambda_1^{\eta^1}, \cdot)} & \mathcal{I}_{(\Lambda_1^{\eta^1}, \cdot)} & P_{(\Lambda_1^{\eta^1}, \cdot)} \\ \mathbf{0}_{(\Lambda_3^{\eta^1}, L_1)} & \nabla_\gamma \theta_{\Lambda_3^{\eta^1}}(\bar{v}) & Q(\bar{\gamma})_{(\Lambda_3^{\eta^1}, \cdot)} & \mathcal{I}_{(\Lambda_3^{\eta^1}, \cdot)} & P_{(\Lambda_3^{\eta^1}, \cdot)} \\ \mathbf{0}_{(\Lambda_1^{\zeta^1}, L_1)} & \mathbf{0}_{(\Lambda_1^{\zeta^1}, L_2)} & \mathcal{I}_{(\Lambda_1^{\zeta^1}, \cdot)} & \mathbf{0}_{(\Lambda_1^{\zeta^1}, L_4)} & \mathbf{0}_{(\Lambda_1^{\zeta^1}, L_5)} \\ \mathbf{0}_{(\Lambda_2^{\zeta^1}, L_1)} & \mathbf{0}_{(\Lambda_2^{\zeta^1}, L_2)} & \mathcal{I}_{(\Lambda_2^{\zeta^1}, \cdot)} & \mathbf{0}_{(\Lambda_2^{\zeta^1}, L_4)} & \mathbf{0}_{(\Lambda_2^{\zeta^1}, L_5)} \\ \mathbf{e}_{((\Lambda_3^c)^{\eta^2}, L_1)} & \mathbf{0}_{((\Lambda_3^c)^{\eta^2}, L_2)} & -\mathcal{I}_{((\Lambda_3^c)^{\eta^2}, \cdot)} & \mathbf{0}_{((\Lambda_3^c)^{\eta^2}, L_4)} & \mathbf{0}_{((\Lambda_3^c)^{\eta^2}, L_5)} \\ \mathbf{e}_{(\Lambda_4^{\eta^2}, L_1)} & \mathbf{0}_{(\Lambda_4^{\eta^2}, L_2)} & -\mathcal{I}_{(\Lambda_4^{\eta^2}, \cdot)} & \mathbf{0}_{(\Lambda_4^{\eta^2}, L_4)} & \mathbf{0}_{(\Lambda_4^{\eta^2}, L_5)} \\ \mathbf{0}_{(\Lambda_1^{\zeta^2}, L_1)} & \mathbf{0}_{(\Lambda_1^{\zeta^2}, L_2)} & \mathbf{0}_{(\Lambda_1^{\zeta^2}, L_3)} & \mathcal{I}_{(\Lambda_1^{\zeta^2}, \cdot)} & \mathbf{0}_{(\Lambda_1^{\zeta^2}, L_5)} \\ \mathbf{0}_{(\Lambda_3^{\zeta^2}, L_1)} & \mathbf{0}_{(\Lambda_3^{\zeta^2}, L_2)} & \mathbf{0}_{(\Lambda_3^{\zeta^2}, L_3)} & \mathcal{I}_{(\Lambda_3^{\zeta^2}, \cdot)} & \mathbf{0}_{(\Lambda_3^{\zeta^2}, L_5)} \\ \mathbf{0}_{([k], L_1)} & \mathbf{0}_{([k], L_2)} & \widehat{Y} & \mathbf{0}_{([k], L_4)} & \mathbf{0}_{([k], L_5)} \\ \mathbf{0}_{(I_g, L_1)} & -\mathbf{e}_{(I_g, L_2)} & \mathbf{0}_{(I_g, L_3)} & \mathbf{0}_{(I_g, L_4)} & \mathbf{0}_{(I_g, L_5)} \end{bmatrix}.$$

Proof (i) Since $\bar{v}$ is a weakly stationary, we obtain (59). We discuss the following 2 cases.

(a) If $\bar{\gamma} > 0$, we have $\bar{\lambda} = 0$. If $\bar{\gamma} = 0$ and $\bar{\lambda} = 0$, we also have $\bar{\lambda} = 0$. Substituting (59) into the Lagrange function, we obtain that

$$\begin{aligned} \mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}) \quad = \quad & f(\bar{v}) + \sum_{i=1}^{k} \bar{\mu}^i h_i(\bar{v}) - \sum_{i \in \Lambda_1 \cup \Lambda_3 \cup \Lambda_4} \bar{\eta}_i^1 G_i^1(\bar{v}) \\ & - \sum_{i \in \Lambda_3^c \cup \Lambda_4} \bar{\eta}_i^2 G_i^2(\bar{v}) - \sum_{i \in \Lambda_1 \cup \Lambda_2} \bar{\zeta}_i^1 H_i^1(\bar{v}) \\ & - \sum_{i \in \Lambda_1 \cup \Lambda_2 \cup \Lambda_3} \bar{\zeta}_i^2 H_i^2(\bar{v}). \end{aligned} \tag{128}$$

By $\nabla_v \mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}) = 0$, we have

$$\begin{aligned} 0 \quad = \quad & \nabla f(\bar{v}) + \sum_{i=1}^{k} \bar{\mu}_i \nabla h_i(\bar{v}) - \sum_{i \in \Lambda_1 \cup \Lambda_3 \cup \Lambda_4} \bar{\eta}_i^1 \nabla G_i^1(\bar{v}) - \sum_{i \in \Lambda_3^c \cup \Lambda_4} \bar{\eta}_i^2 \nabla G_i^2(\bar{v}) \\ & - \sum_{i \in \Lambda_1 \cup \Lambda_2} \bar{\zeta}_i^1 \nabla H_i^1(\bar{v}) - \sum_{i \in \Lambda_1 \cup \Lambda_2 \cup \Lambda_3} \bar{\zeta}_i^2 \nabla H_i^2(\bar{v}). \end{aligned}$$

Denote by

$$\rho := \begin{bmatrix} 1 \\ -\bar{\eta}_{\Lambda_1}^1 \\ -\bar{\eta}_{\Lambda_3}^1 \\ -\bar{\eta}_{\Lambda_4}^1 \\ -\bar{\zeta}_{\Lambda_1}^1 \\ -\bar{\zeta}_{\Lambda_2}^1 \\ -\bar{\eta}_{\Lambda_3^c}^2 \\ -\bar{\eta}_{\Lambda_4}^2 \\ -\bar{\zeta}_{\Lambda_1}^2 \\ -\bar{\zeta}_{\Lambda_2}^2 \\ -\bar{\zeta}_{\Lambda_3}^2 \\ \bar{\mu} \end{bmatrix}.$$

The above conditions can be equivalently written as

$$
0 \quad = \begin{bmatrix} 1 \\ -\bar{\eta}^1_{\Lambda_1} \\ -\bar{\eta}^1_{\Lambda_3} \\ -\bar{\eta}^1_{\Lambda_4} \\ -\bar{\zeta}^1_{\Lambda_1} \\ -\bar{\zeta}^1_{\Lambda_2} \\ -\bar{\eta}^2_{\Lambda_3^c} \\ -\bar{\eta}^2_{\Lambda_4} \\ -\bar{\zeta}^2_{\Lambda_1} \\ -\bar{\zeta}^2_{\Lambda_2} \\ -\bar{\zeta}^2_{\Lambda_3} \\ \bar{\mu} \end{bmatrix}^\top \cdot \begin{bmatrix} \nabla_C f(\bar{v}) & \nabla_\gamma f(\bar{v}) & \nabla_\alpha f(\bar{v})^\top & \mathbf{0}_{(1,n)} & \mathbf{0}_{(1,k)} \\ \mathbf{0}_{(\Lambda_1,L_1)} & \nabla_\gamma \theta(\bar{v})_{\Lambda_1} & Q(\bar{\gamma})_{(\Lambda_1,\cdot\ )} & \mathcal{I}_{(\Lambda_1,\cdot)} & P_{(\Lambda_1,\cdot)} \\ \mathbf{0}_{(\Lambda_3,L_1)} & \nabla_\gamma \theta(\bar{v})_{\Lambda_3} & Q(\bar{\gamma})_{(\Lambda_3,\cdot\ )} & \mathcal{I}_{(\Lambda_3,\cdot)} & P_{(\Lambda_3,\cdot)} \\ \mathbf{0}_{(\Lambda_4,L_1)} & \nabla_\gamma \theta(\bar{v})_{\Lambda_4} & Q(\bar{\gamma})_{(\Lambda_4,\cdot\ )} & \mathcal{I}_{(\Lambda_4,\cdot)} & P_{(\Lambda_4,\cdot)} \\ \mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_4)} & \mathbf{0}_{(\Lambda_1,L_5)} \\ \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_4)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{e}_{(\Lambda_3^c,L_1)} & \mathbf{0}_{(\Lambda_3^c,L_2)} & -\mathcal{I}_{(\Lambda_3^c,\cdot)} & \mathbf{0}_{(\Lambda_3^c,L_4)} & \mathbf{0}_{(\Lambda_3^c,L_5)} \\ \mathbf{e}_{(\Lambda_4,L_1)} & \mathbf{0}_{(\Lambda_4,L_2)} & -\mathcal{I}_{(\Lambda_4,\cdot)} & \mathbf{0}_{(\Lambda_4,L_4)} & \mathbf{0}_{(\Lambda_4,L_5)} \\ \mathbf{0}_{(\Lambda_1,L_1)} & \mathbf{0}_{(\Lambda_1,L_2)} & \mathbf{0}_{(\Lambda_1,L_3)} & \mathcal{I}_{(\Lambda_1,\cdot)} & \mathbf{0}_{(\Lambda_1,L_5)} \\ \mathbf{0}_{(\Lambda_2,L_1)} & \mathbf{0}_{(\Lambda_2,L_2)} & \mathbf{0}_{(\Lambda_2,L_3)} & \mathcal{I}_{(\Lambda_2,\cdot)} & \mathbf{0}_{(\Lambda_2,L_5)} \\ \mathbf{0}_{(\Lambda_3,L_1)} & \mathbf{0}_{(\Lambda_3,L_2)} & \mathbf{0}_{(\Lambda_3,L_3)} & \mathcal{I}_{(\Lambda_3,\cdot)} & \mathbf{0}_{(\Lambda_3,L_5)} \\ \mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \widehat{Y} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \end{bmatrix} := \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{bmatrix},
$$

where

$$
\begin{aligned}
S_1 &:= \nabla_C f(\bar{v}) - \left(\bar{\eta}^2_{\Lambda_3^c}\right)^\top \mathbf{e}_{(\Lambda_3^c,L_1)} - \left(\bar{\eta}^2_{\Lambda_4}\right)^\top \mathbf{e}_{(\Lambda_4,L_1)} = 0, \\
S_2 &:= \nabla_\gamma f(\bar{v}) - \left(\bar{\eta}^1_{\Lambda_1}\right)^\top \nabla_\gamma \theta(\bar{v})_{\Lambda_1} - \left(\bar{\eta}^1_{\Lambda_3}\right)^\top \nabla_\gamma \theta(\bar{v})_{\Lambda_3} \\
&\quad - \left(\bar{\eta}^1_{\Lambda_4}\right)^\top \nabla_\gamma \theta(\bar{v})_{(\Lambda_4,\cdot)} = 0, \qquad (129) \\
S_3 &:= \nabla_\alpha f(\bar{v})^\top - \left(\bar{\eta}^1_{\Lambda_1}\right)^\top (Q(\gamma))_{(\Lambda_1,\cdot)} - \left(\bar{\eta}^1_{\Lambda_3}\right)^\top (Q(\gamma))_{(\Lambda_3,\cdot)} \\
&\quad - \left(\bar{\eta}^1_{\Lambda_4}\right)^\top (Q(\gamma))_{(\Lambda_4,\cdot)} - \left(\bar{\zeta}^1_{\Lambda_1}\right)^\top \mathcal{I}_{(\Lambda_1,\cdot)} - \left(\bar{\zeta}^1_{\Lambda_2}\right)^\top \mathcal{I}_{(\Lambda_2,\cdot)} \\
&\quad + \left(\bar{\eta}^2_{\Lambda_3^c}\right)^\top \mathcal{I}_{(\Lambda_3^c,\cdot)} + \left(\bar{\eta}^2_{\Lambda_4}\right)^\top \mathcal{I}_{(\Lambda_4,\cdot)} + \bar{\mu}^\top \widehat{Y} = 0, \qquad (130) \\
S_4 &:= -\left(\bar{\eta}^1_{\Lambda_1}\right)^\top \mathcal{I}_{(\Lambda_1,\cdot)} - \left(\bar{\eta}^1_{\Lambda_3}\right)^\top \mathcal{I}_{(\Lambda_3,\cdot)} - \left(\bar{\eta}^1_{\Lambda_4}\right)^\top \mathcal{I}_{(\Lambda_4,\cdot)} - \left(\bar{\zeta}^2_{\Lambda_1}\right)^\top \mathcal{I}_{(\Lambda_1,\cdot)} \\
&\quad - \left(\bar{\zeta}^2_{\Lambda_2}\right)^\top \mathcal{I}_{(\Lambda_2,\cdot)} - \left(\bar{\zeta}^2_{\Lambda_3}\right)^\top \mathcal{I}_{(\Lambda_3,\cdot)} = 0, \qquad (131) \\
S_5 &:= -\left(\bar{\eta}^1_{\Lambda_1}\right)^\top P_{(\Lambda_1,\cdot)} - \left(\bar{\eta}^1_{\Lambda_3}\right)^\top P_{(\Lambda_3,\cdot)} - \left(\bar{\eta}^1_{\Lambda_4}\right)^\top P_{(\Lambda_4,\cdot)} = 0. \qquad (132)
\end{aligned}
$$

By (131), $\bar{\eta}^1_{\Lambda_4} = 0$, $\bar{w}^2_{\Lambda_2} = 0$. Substituting them into the Lagrangian function in (128),

$$
\begin{aligned}
\mathbb{L}(v, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}) &= f(v) + \sum_{i=1}^{k} \bar{\mu}_i h_i(v) - \sum_{i \in \Lambda_1 \cup \Lambda_3} \bar{\eta}^1_i G^1_i(v) - \sum_{i \in \Lambda_3^c \cup \Lambda_4} \bar{\eta}^2_i G^2_i(v) \\
&\quad - \sum_{i \in \Lambda_1 \cup \Lambda_2} \bar{\zeta}^1_i H^1_i(v) - \sum_{i \in \Lambda_1 \cup \Lambda_3} \bar{\zeta}^2_i H^2_i(v). \qquad (133)
\end{aligned}
$$

It holds that

$$\begin{aligned}
\nabla^2_{vv}\mathbb{L}(\bar{v},\bar{\lambda},\bar{\mu},\bar{\eta},\bar{\zeta}) &= \nabla^2_{vv}f(\bar{v}) + \sum_{i=1}^{k}\bar{\mu}_i\nabla^2_{vv}h_i(\bar{v}) - \sum_{i\in\Lambda_1\cup\Lambda_3}\bar{\eta}^1_i\nabla^2_{vv}G^1_i(\bar{v}) \\
&\quad - \sum_{i\in\Lambda_3^c\cup\Lambda_u}\bar{\eta}^2_i\nabla^2_{vv}G^2_i(\bar{v}) - \sum_{i\in\Lambda_1\cup\Lambda_2}\bar{\zeta}^1_i\nabla^2_{vv}H^1_i(\bar{v}) \\
&\quad - \sum_{i\in\Lambda_1\cup\Lambda_3}\bar{\zeta}^2_i\nabla^2_{vv}H^2_i(\bar{v}) \\
&:= \nabla^2_{vv}f(\bar{v}) - \Delta(\bar{v}).
\end{aligned}$$

Since f is a function only related to C,γ,α, it holds that $\frac{\partial f}{\partial\sigma}=0$ and $\frac{\partial f}{\partial u}=0$. Therefore,

$$\nabla^2_{vv}f(\bar{v}) = \begin{bmatrix}
\nabla^2_{CC}f(\bar{v}) & \nabla^2_{C\gamma}f(\bar{v}) & \nabla^2_{C\alpha}f(\bar{v}) & \mathbf{0}_{(L_1,L_4)} & \mathbf{0}_{(L_1,L_5)} \\
\nabla^2_{\gamma C}f(\bar{v}) & \nabla^2_{\gamma\gamma}f(\bar{v}) & \nabla^2_{\gamma\alpha}f(\bar{v}) & \mathbf{0}_{(L_2,L_4)} & \mathbf{0}_{(L_2,L_5)} \\
\nabla^2_{C\alpha}f(\bar{v})^\top & \nabla^2_{\gamma\alpha}f(\bar{v})^\top & \nabla^2_{\alpha\alpha}f(\bar{v}) & \mathbf{0}_{(L_3,L_4)} & \mathbf{0}_{(L_3,L_5)} \\
\mathbf{0}_{(L_4,L_1)} & \mathbf{0}_{(L_4,L_2)} & \mathbf{0}_{(L_4,L_3)} & \mathbf{0}_{(L_4,L_4)} & \mathbf{0}_{(L_4,L_5)} \\
\mathbf{0}_{(L_5,L_1)} & \mathbf{0}_{(L_5,L_2)} & \mathbf{0}_{(L_5,L_3)} & \mathbf{0}_{(L_5,L_4)} & \mathbf{0}_{(L_5,L_5)}
\end{bmatrix}.$$

Note that $H(v),h(v)$ are both linear. Moreover, the $(n+1)$-th component to $2n$-th component in G are also linear. Therefore, we have

$$\Delta(\bar{v}) = \sum_{i\in\Lambda_1\cup\Lambda_3}\bar{\eta}^1_i\nabla^2_{vv}G^1_i(\bar{v}),$$

which reduces to the form in (i).

(b) If $\bar{\gamma}=0$ and $\bar{\lambda}>0$, it is easy to modify the above result, where (133) is replaced by

$$\begin{aligned}
\mathbb{L}(v,\bar{\lambda},\bar{\mu},\bar{\eta},\bar{\zeta}) &= f(v) + \sum_{i=1}^{k}\bar{\mu}_ih_i(v) - \sum_{i\in\Lambda_1\cup\Lambda_3}\bar{\eta}^1_iG^1_i(v) - \sum_{i\in\Lambda_3^c\cup\Lambda_4}\bar{\eta}^2_iG^2_i(v) \\
&\quad - \sum_{i\in\Lambda_1\cup\Lambda_2}\bar{\zeta}^1_iH^1_i(v) - \sum_{i\in\Lambda_1\cup\Lambda_3}\bar{\zeta}^2_iH^2_i(v) + \bar{\lambda}\nabla g(v).
\end{aligned}$$

Again, we obtain the result in (i).

(ii) If $\bar{\gamma}>0$ (or $\bar{\gamma}=0$ with $\bar{\lambda}=0$), by Proposition 27, it holds that

$$\mathrm{supp}(\bar{\eta}^1) = \Lambda_1^{\eta^1}\cup\Lambda_3^{\eta^1},\ \mathrm{supp}(\bar{\eta}^2) = (\Lambda_3^c)^{\eta_2}\cup\Lambda_4^{\eta_2},$$

$$\mathrm{supp}(\bar{\zeta}^1) = \Lambda_1^{\zeta^1}\cup\Lambda_2^{\zeta^1},\ \mathrm{supp}(\bar{\zeta}^2) = \Lambda_1^{\zeta^2}\cup\Lambda_3^{\zeta^2}.$$

By the definition of $\mathcal{C}(\bar{v})$, it holds that

$$\begin{aligned}
\mathcal{C}(\bar{v}) &= \{d\in\mathbb{R}^m : \nabla h_i(\bar{v})^\top d = 0, i\in[k]; \\
&\qquad G_i(\bar{v})^\top d = 0,\ \forall i\in\mathrm{supp}(\bar{\eta}); \nabla H_i(\bar{v})^\top d = 0,\ \forall\ i\in\mathrm{supp}(\bar{\zeta})\} \\
&= \{d\in\mathbb{R}^m\ :\ \nabla h_i(\bar{v})^\top d = 0, i\in[k]; \\
&\quad \nabla G^1_i(\bar{v})^\top d = 0,\ \forall\ i\in\Lambda_1^{\eta^1}\cup\Lambda_3^{\eta^1}; G^2_i(\bar{v})^\top d = 0, \forall\ i\in(\Lambda_3^c)^{\eta^2}\cup\Lambda_4^{\eta^2}; \\
&\quad \nabla H^1_i(\bar{v})^\top d = 0, \forall\ i\in\Lambda_1^{\zeta^1}\cup\Lambda_2^{\zeta^1}, \nabla H^2_i(\bar{v})^\top d = 0, \forall\ i\in\Lambda_1^{\zeta^2}\cup\Lambda_3^{\zeta^2}\}.
\end{aligned}$$

It gives that $d \in \mathcal{C}(\bar{v})$ if and only if d is the solution of linear system $Wd = 0$, where

$$W := \begin{bmatrix} \mathbf{0}_{(\Lambda_1^{\eta^1}, L_1)} & \nabla_\gamma \theta(\bar{v})_{\Lambda_1^{\eta^1}} & Q(\bar{\gamma})_{(\Lambda_1^{\eta^1}, \cdot)} & \mathcal{I}_{(\Lambda_1^{\eta^1}, \cdot)} & P_{(\Lambda_1^{\eta^1}, \cdot)} \\ \mathbf{0}_{(\Lambda_3^{\eta^1}, L_1)} & \nabla_\gamma \theta(\bar{v})_{\Lambda_3^{\eta^1}} & Q(\bar{\gamma})_{(\Lambda_3^{\eta^1}, \cdot)} & \mathcal{I}_{(\Lambda_3^{\eta^1}, \cdot)} & P_{(\Lambda_3^{\eta^1}, \cdot)} \\ \mathbf{0}_{(\Lambda_1^{\zeta^1}, L_1)} & \mathbf{0}_{(\Lambda_1^{\zeta^1}, L_2)} & \mathcal{I}_{(\Lambda_1^{\zeta^1}, \cdot)} & \mathbf{0}_{(\Lambda_1^{\zeta^1}, L_4)} & \mathbf{0}_{(\Lambda_1^{\zeta^1}, L_5)} \\ \mathbf{0}_{(\Lambda_2^{\zeta^1}, L_1)} & \mathbf{0}_{(\Lambda_2^{\zeta^1}, L_2)} & \mathcal{I}_{(\Lambda_2^{\zeta^1}, \cdot)} & \mathbf{0}_{(\Lambda_2^{\zeta^1}, L_4)} & \mathbf{0}_{(\Lambda_2^{\zeta^1}, L_5)} \\ \mathbf{e}_{((\Lambda_3^c)^{\eta^2}, L_1)} & \mathbf{0}_{((\Lambda_3^c)^{\eta^2}, L_2)} & -\mathcal{I}_{((\Lambda_3^c)^{\eta^2}, \cdot)} & \mathbf{0}_{((\Lambda_3^c)^{\eta^2}, L_4)} & \mathbf{0}_{((\Lambda_3^c)^{\eta^2}, L_5)} \\ \mathbf{e}_{(\Lambda_4^{\eta^2}, L_1)} & \mathbf{0}_{(\Lambda_4^{\eta^2}, L_2)} & -\mathcal{I}_{(\Lambda_4^{\eta^2}, \cdot)} & \mathbf{0}_{(\Lambda_4^{\eta^2}, L_4)} & \mathbf{0}_{(\Lambda_4^{\eta^2}, L_5)} \\ \mathbf{0}_{(\Lambda_1^{\zeta^2}, L_1)} & \mathbf{0}_{(\Lambda_1^{\zeta^2}, L_2)} & \mathbf{0}_{(\Lambda_1^{\zeta^2}, L_3)} & \mathcal{I}_{(\Lambda_1^{\zeta^2}, \cdot)} & \mathbf{0}_{(\Lambda_1^{\zeta^2}, L_5)} \\ \mathbf{0}_{(\Lambda_3^{\zeta^2}, L_1)} & \mathbf{0}_{(\Lambda_3^{\zeta^2}, L_2)} & \mathbf{0}_{(\Lambda_3^{\zeta^2}, L_3)} & \mathcal{I}_{(\Lambda_3^{\zeta^2}, \cdot)} & \mathbf{0}_{(\Lambda_3^{\zeta^2}, L_5)} \\ \mathbf{0}_{([k], L_1)} & \mathbf{0}_{([k], L_2)} & \widehat{Y} & \mathbf{0}_{([k], L_4)} & \mathbf{0}_{([k], L_5)} \end{bmatrix}.$$

If $\bar{\gamma} = 0$ with $\bar{\lambda} > 0$, then $\nabla g_{I_g}(\bar{v})$ should be added into the row of W, which give the last row block in W in (ii). Overall, we obtain (ii). ■

Proof of Theorem 28

(i) If $\Lambda_2 \cup \Lambda_4 \neq \emptyset$, by Proposition 40 (ii), we can see that $d^\sigma_{\Lambda_2 \cup \Lambda_4}$ does not appear in the equation $Wd = 0$. Therefore, by Proposition 40 (ii), we can choose d with $d^\sigma_{\Lambda_2 \cup \Lambda_4} \neq 0$ but the remaining components in d are zero. For such a vector d, we always have

$$d^\top \nabla^2_{vv} \mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}) d = 0$$

by the special structure of $\nabla^2_{vv} \mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$ based on Proposition 40 (i). Hence, the MPEC-SSOSC fails at $\bar{v}$.

(ii) If $\Lambda_2 \cup \Lambda_4 = \emptyset$, and $\Lambda^0 \neq \emptyset$, similar to the argument in (i), we can choose d with $d^\sigma_{\Lambda^0} \neq 0$ but the rest of d are zero. Now we found $d \neq 0$ such that

$$d^\top \nabla^2_{vv} \mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}) d = 0.$$

Therefore, MPEC-SSOSC does not hold at $\bar{v}$. □

Proof of Theorem 29

First start by observing that by Assumption 6,

$$\Lambda_1^{\eta^1} \cup \Lambda_1^{\zeta^2} = \Lambda_1 \ \text{ and } \ \Lambda_3^{\eta^1} \cup \Lambda_3^{\zeta^2} = \Lambda_3.$$

We do the following partition associated to the index sets $\Lambda_1^{\eta^1}$, $\Lambda_1^{\zeta^2}$, $(\Lambda_3^+)^{\eta^1}$, $(\Lambda_3^c)^{\eta^1}$, $\Lambda_1^{\zeta^1}$, $(\Lambda_3^c)^{\eta^2}$, $(\Lambda_3^+)^{\zeta^2}$, and $(\Lambda_3^c)^{\zeta^2}$. Let

$$\begin{cases} \Pi_1 := \Lambda_1^{\eta^1} \backslash \Lambda_1^{\zeta^2}, & \Pi_2 := \Lambda_1^{\eta^1} \cap \Lambda_1^{\zeta^2}, & \Pi_3 := \Lambda_1^{\zeta^2} \backslash \Lambda_1^{\eta^1}, \\ \Pi_{i1} := \Pi_i \cap \Lambda_1^{\zeta^1}, & \bar{\Pi}_{i1} := \Pi_i \backslash \Lambda_1^{\zeta^1}, & i = 1, 2, 3; \\ \phi_{\gamma 1} := (\Lambda_3^+)^{\eta^1} \backslash (\Lambda_3^+)^{\zeta^2}, & \phi_{\gamma 2} := (\Lambda_3^+)^{\eta^1} \cap (\Lambda_3^+)^{\zeta^2}, & \phi_{\gamma 3} := (\Lambda_3^+)^{\zeta^2} \backslash (\Lambda_3^+)^{\eta^1}, \\ \Theta_1 := (\Lambda_3^c)^{\eta^1} \backslash (\Lambda_3^c)^{\zeta^2}, & \Theta_2 := (\Lambda_3^c)^{\eta^1} \cap (\Lambda_3^c)^{\zeta^2}, & \Theta_3 := (\Lambda_3^c)^{\zeta^2} \backslash (\Lambda_3^c)^{\eta^1}, \\ \Theta_{i2} := \Theta_i \cap (\Lambda_3^c)^{\eta^2}, & \bar{\Theta}_{i2} := \Theta_i \backslash (\Lambda_3^c)^{\eta^2}, & i = 1, 2, 3. \end{cases}$$

See the following figure for the partitions of the above sets:

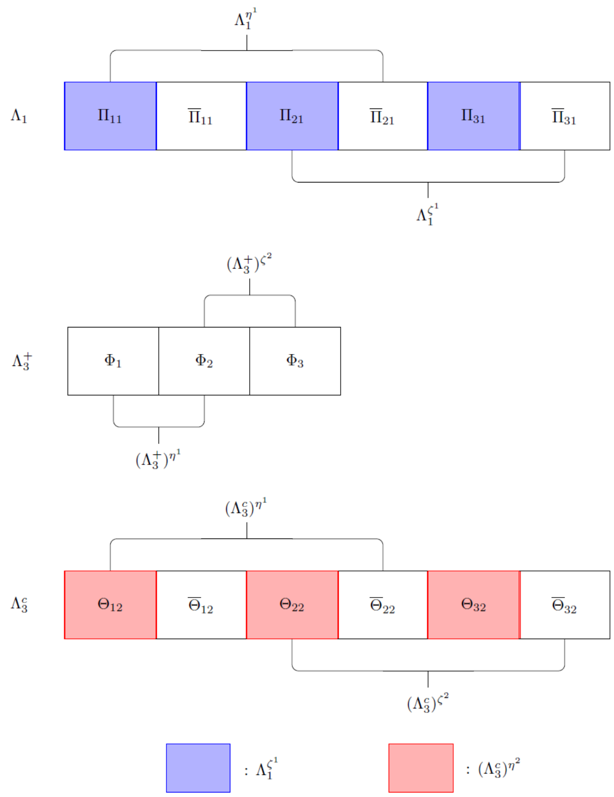


Figure 5: Partitions of Λ_1, Λ_3^+, and Λ_3^c, respectively.

We have the following relationships:

$$\Lambda_1 = \cup_{i=1,2,3}\Pi_i, \quad \Lambda_1^{\eta^1} = \Pi_1 \cup \Pi_2, \quad \Lambda_1^{\zeta^2} = \Pi_2 \cup \Pi_3, \quad \Lambda_1^{\zeta^1} = \cup_{i=1,2,3}\Pi_{i1},$$

$$\Lambda_3^+ = \cup_{i=1,2,3}\phi_{\gamma i}, \quad \Lambda_3^c = \cup_{i=1,2,3}\Theta_i, \quad \Lambda_3^{\eta^1} = \phi_{\gamma 1} \cup \phi_{\gamma 2} \cup \Theta_1 \cup \Theta_2,$$

$$\Lambda_3^{\zeta^2} = \phi_{\gamma 2} \cup \phi_{\gamma 3} \cup \Theta_2 \cup \Theta_3, \quad (\Lambda_3^c)^{\eta^2} = \cup_{i=1,2,3}\Theta_{i2}.$$

Proposition 41. *Let $\bar{v}$ be a weakly stationary for* (MPEC-BHO) *with multiplier vector* $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{w})$*. With Assumption* 6*, W in Proposition 40 (ii) reduces to the form*

$$W = \begin{bmatrix}
\mathbf{0}_{(\Pi_{11},L_1)} & \nabla_\gamma\theta(\bar{v})_{\Pi_{11}} & Q(\bar{\gamma})_{(\Pi_{11},\cdot)} & \mathcal{I}_{(\Pi_{11},\cdot)} & P_{(\Pi_{11},\cdot)} \\
\mathbf{0}_{(\bar{\Pi}_{11},L_1)} & \nabla_\gamma\theta(\bar{v})_{\bar{\Pi}_{11}} & Q(\bar{\gamma})_{(\bar{\Pi}_{11},\cdot)} & \mathcal{I}_{(\bar{\Pi}_{11},\cdot)} & P_{(\bar{\Pi}_{11},\cdot)} \\
\mathbf{0}_{(\Pi_{21},L_1)} & \nabla_\gamma\theta(\bar{v})_{\Pi_{21}} & Q(\bar{\gamma})_{(\Pi_{21},\cdot)} & \mathcal{I}_{(\Pi_{21},\cdot)} & P_{(\Pi_{21},\cdot)} \\
\mathbf{0}_{(\bar{\Pi}_{21},L_1)} & \nabla_\gamma\theta(\bar{v})_{\bar{\Pi}_{21}} & Q(\bar{\gamma})_{(\bar{\Pi}_{21},\cdot)} & \mathcal{I}_{(\bar{\Pi}_{21},\cdot)} & P_{(\bar{\Pi}_{21},\cdot)} \\
\mathbf{0}_{(\phi_{\gamma 1},L_1)} & \nabla_\gamma\theta(\bar{v})_{\phi_{\gamma 1}} & Q(\bar{\gamma})_{(\phi_{\gamma 1},\cdot)} & \mathcal{I}_{(\phi_{\gamma 1},\cdot)} & P_{(\phi_{\gamma 1},\cdot)} \\
\mathbf{0}_{(\phi_{\gamma 2},L_1)} & \nabla_\gamma\theta(\bar{v})_{\phi_{\gamma 2}} & Q(\bar{\gamma})_{(\phi_{\gamma 2},\cdot)} & \mathcal{I}_{(\phi_{\gamma 2},\cdot)} & P_{(\phi_{\gamma 2},\cdot)} \\
\mathbf{0}_{(\Theta_{12},L_1)} & \nabla_\gamma\theta(\bar{v})_{\Theta_{12}} & Q(\bar{\gamma})_{(\Theta_{12},\cdot)} & \mathcal{I}_{(\Theta_{12},\cdot)} & P_{(\Theta_{12},\cdot)} \\
\mathbf{0}_{(\bar{\Theta}_{12},L_1)} & \nabla_\gamma\theta(\bar{v})_{\bar{\Theta}_{12}} & Q(\bar{\gamma})_{(\bar{\Theta}_{12},\cdot)} & \mathcal{I}_{(\bar{\Theta}_{12},\cdot)} & P_{(\bar{\Theta}_{12},\cdot)} \\
\mathbf{0}_{(\Theta_{22},L_1)} & \nabla_\gamma\theta(\bar{v})_{\Theta_{22}} & Q(\bar{\gamma})_{(\Theta_{22},\cdot)} & \mathcal{I}_{(\Theta_{22},\cdot)} & P_{(\Theta_{22},\cdot)} \\
\mathbf{0}_{(\bar{\Theta}_{22},L_1)} & \nabla_\gamma\theta(\bar{v})_{\bar{\Theta}_{22}} & Q(\bar{\gamma})_{(\bar{\Theta}_{22},\cdot)} & \mathcal{I}_{(\bar{\Theta}_{22},\cdot)} & P_{(\bar{\Theta}_{22},\cdot)} \\
\mathbf{0}_{(\Pi_{11},L_1)} & \mathbf{0}_{(\Pi_{11},L_2)} & \mathcal{I}_{(\Pi_{11},\cdot)} & \mathbf{0}_{(\Pi_{11},L_4)} & \mathbf{0}_{(\Pi_{11},L_5)} \\
\mathbf{0}_{(\Pi_{21},L_1)} & \mathbf{0}_{(\Pi_{21},L_2)} & \mathcal{I}_{(\Pi_{21},\cdot)} & \mathbf{0}_{(\Pi_{21},L_4)} & \mathbf{0}_{(\Pi_{21},L_5)} \\
\mathbf{0}_{(\Pi_{31},L_1)} & \mathbf{0}_{(\Pi_{31},L_2)} & \mathcal{I}_{(\Pi_{31},\cdot)} & \mathbf{0}_{(\Pi_{31},L_4)} & \mathbf{0}_{(\Pi_{31},L_5)} \\
\mathbf{e}_{(\Theta_{12},L_1)} & \mathbf{0}_{((\Theta_{12},L_2)} & -\mathcal{I}_{(\Theta_{12},\cdot)} & \mathbf{0}_{(\Theta_{12},L_4)} & \mathbf{0}_{(\Theta_{12},L_5)} \\
\mathbf{e}_{(\Theta_{22},L_1)} & \mathbf{0}_{((\Theta_{22},L_2)} & -\mathcal{I}_{(\Theta_{22},\cdot)} & \mathbf{0}_{(\Theta_{22},L_4)} & \mathbf{0}_{(\Theta_{22},L_5)} \\
\mathbf{e}_{(\Theta_{32},L_1)} & \mathbf{0}_{((\Theta_{32},L_2)} & -\mathcal{I}_{(\Theta_{32},\cdot)} & \mathbf{0}_{(\Theta_{32},L_4)} & \mathbf{0}_{(\Theta_{32},L_5)} \\
\mathbf{0}_{(\Pi_{21},L_1)} & \mathbf{0}_{(\Pi_{21},L_2)} & \mathbf{0}_{(\Pi_{21},L_3)} & \mathcal{I}_{(\Pi_{21},\cdot)} & \mathbf{0}_{(\Pi_{11},L_5)} \\
\mathbf{0}_{(\bar{\Pi}_{21},L_1)} & \mathbf{0}_{(\bar{\Pi}_{21},L_2)} & \mathbf{0}_{(\bar{\Pi}_{21},L_3)} & \mathcal{I}_{(\bar{\Pi}_{21},\cdot)} & \mathbf{0}_{(\bar{\Pi}_{21},L_5)} \\
\mathbf{0}_{(\Pi_{31},L_1)} & \mathbf{0}_{(\Pi_{31},L_2)} & \mathbf{0}_{(\Pi_{31},L_3)} & \mathcal{I}_{(\Pi_{31},\cdot)} & \mathbf{0}_{(\Pi_{31},L_5)} \\
\mathbf{0}_{(\bar{\Pi}_{31},L_1)} & \mathbf{0}_{(\bar{\Pi}_{31},L_2)} & \mathbf{0}_{(\bar{\Pi}_{31},L_3)} & \mathcal{I}_{(\bar{\Pi}_{31},\cdot)} & \mathbf{0}_{(\bar{\Pi}_{31},L_5)} \\
\mathbf{0}_{(\phi_{\gamma 2},L_1)} & \mathbf{0}_{(\phi_{\gamma 2},L_2)} & \mathbf{0}_{(\phi_{\gamma 2},L_3)} & \mathcal{I}_{(\phi_{\gamma 2},\cdot)} & \mathbf{0}_{(\phi_{\gamma 2},L_5)} \\
\mathbf{0}_{(\phi_{\gamma 3},L_1)} & \mathbf{0}_{(\phi_{\gamma 3},L_2)} & \mathbf{0}_{(\phi_{\gamma 3},L_3)} & \mathcal{I}_{(\phi_{\gamma 3},\cdot)} & \mathbf{0}_{(\phi_{\gamma 3},L_5)} \\
\mathbf{0}_{(\Theta_{22},L_1)} & \mathbf{0}_{(\Theta_{22},L_2)} & \mathbf{0}_{(\Theta_{22},L_3)} & \mathcal{I}_{(\Theta_{22},\cdot)} & \mathbf{0}_{(\Theta_{22},L_5)} \\
\mathbf{0}_{(\Theta_{32},L_1)} & \mathbf{0}_{(\Theta_{32},L_2)} & \mathbf{0}_{(\Theta_{32},L_3)} & \mathcal{I}_{(\Theta_{32},\cdot)} & \mathbf{0}_{(\Theta_{32},L_5)} \\
\mathbf{0}_{(\bar{\Theta}_{22},L_1)} & \mathbf{0}_{(\bar{\Theta}_{22},L_2)} & \mathbf{0}_{(\bar{\Theta}_{22},L_3)} & \mathcal{I}_{(\bar{\Theta}_{22},\cdot)} & \mathbf{0}_{(\bar{\Theta}_{22},L_5)} \\
\mathbf{0}_{(\bar{\Theta}_{32},L_1)} & \mathbf{0}_{(\bar{\Theta}_{32},L_2)} & \mathbf{0}_{(\bar{\Theta}_{32},L_3)} & \mathcal{I}_{(\bar{\Theta}_{32},\cdot)} & \mathbf{0}_{(\bar{\Theta}_{32},L_5)} \\
\mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \widehat{Y} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \\
\mathbf{0}_{(I_g,L_1)} & -\mathbf{e}_{(I_g,L_2)} & \mathbf{0}_{(I_g,L_3)} & \mathbf{0}_{(I_g,L_4)} & \mathbf{0}_{(I_g,L_5)}
\end{bmatrix}. \tag{134}$$

Proof Note that in this case, W reduces to the following form

$$W = \begin{bmatrix}
\mathbf{0}_{(\Lambda_1^{\eta^1},L_1)} & \nabla_\gamma\theta(\bar{v})_{\Lambda_1^{\eta^1}} & Q(\bar{\gamma})_{(\Lambda_1^{\eta^1},\cdot)} & \mathcal{I}_{(\Lambda_1^{\eta^1},\cdot)} & P_{(\Lambda_1^{\eta^1},\cdot)} \\
\mathbf{0}_{(\Lambda_3^{\eta^1},L_1)} & \nabla_\gamma\theta(\bar{v})_{\Lambda_3^{\eta^1}} & Q(\bar{\gamma})_{(\Lambda_3^{\eta^1},\cdot)} & \mathcal{I}_{(\Lambda_3^{\eta^1},\cdot)} & P_{(\Lambda_3^{\eta^1},\cdot)} \\
\mathbf{0}_{(\Lambda_1^{\zeta^1},L_1)} & \mathbf{0}_{(\Lambda_1^{\zeta^1},L_2)} & \mathcal{I}_{(\Lambda_1^{\zeta^1},\cdot)} & \mathbf{0}_{(\Lambda_1^{\zeta^1},L_4)} & \mathbf{0}_{(\Lambda_1^{\zeta^1},L_5)} \\
\mathbf{e}_{((\Lambda_3^c)^{\eta^2},L_1)} & \mathbf{0}_{((\Lambda_3^c)^{\eta^2},L_2)} & -\mathcal{I}_{((\Lambda_3^c)^{\eta^2},\cdot)} & \mathbf{0}_{((\Lambda_3^c)^{\eta^2},L_4)} & \mathbf{0}_{((\Lambda_3^c)^{\eta^2},L_5)} \\
\mathbf{0}_{(\Lambda_1^{\zeta^2},L_1)} & \mathbf{0}_{(\Lambda_1^{\zeta^2},L_2)} & \mathbf{0}_{(\Lambda_1^{w^2},L_3)} & \mathcal{I}_{(\Lambda_1^{\zeta^2},\cdot)} & \mathbf{0}_{(\Lambda_1^{\zeta^2},L_5)} \\
\mathbf{0}_{(\Lambda_3^{\zeta^2},L_1)} & \mathbf{0}_{(\Lambda_3^{\zeta^2},L_2)} & \mathbf{0}_{(\Lambda_3^{\zeta^2},L_3)} & \mathcal{I}_{(\Lambda_3^{\zeta^2},\cdot)} & \mathbf{0}_{(\Lambda_3^{\zeta^2},L_5)} \\
\mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \widehat{Y} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \\
\mathbf{0}_{(I_g,L_1)} & -\mathbf{e}_{(I_g,L_2)} & \mathbf{0}_{(I_g,L_3)} & \mathbf{0}_{(I_g,L_4)} & \mathbf{0}_{(I_g,L_5)}
\end{bmatrix}$$

The result is obvious by directly applying the assumption. ∎

Recall the definition of $\widetilde{A}^1$ and $\widetilde{A}^2$ in (64) and consider

$$
\begin{cases}
\bar{a}^1 := -(Q(\bar{\gamma})_{(I_1,I_1)})^{-1}Q(\bar{\gamma})_{(I_1,I_3)}\mathbf{e}_{|I_3|}, & \bar{a}^2 := Q(\bar{\gamma})_{(I',I_3)}\mathbf{e}_{|I_3|} - \widetilde{A}_1 Q(\bar{\gamma})_{(I_1,I_3)}\mathbf{e}_{|I_3|}, \\
\bar{b}^1 := -(Q(\bar{\gamma})_{(I_1,I_1)})^{-1}\nabla_\gamma \theta(\bar{v})_{I_1}, & \bar{b}^2 := \nabla_\gamma \theta(\bar{v})_{I'} - \widetilde{A}_1 \nabla_\gamma \theta(\bar{v})_{I_1}, \\
\bar{Q}^1 := -(Q(\bar{\gamma})_{(I_1,I_1)})^{-1}Q(\bar{\gamma})_{(I_1,I_2)}, & \bar{Q}^2 := Q(\bar{\gamma})_{(I',I_2)} - \widetilde{A}_1 Q(\bar{\gamma})_{(I_1,I_2)}, \\
\bar{P}^1 := -(Q(\bar{\gamma})_{(I_1,I_1)})^{-1}P_{(I_1,\cdot)}, & \bar{P}^2 := P_{(I',\cdot)} - \widetilde{A}_1 P_{(I_1,\cdot)}, \\
\tilde{y}^1 := \widehat{Y}_{([k],I_3)}\mathbf{e}_{|I_3|} - \widetilde{A}_2 Q(\bar{\gamma})_{(I_1,I_3)}\mathbf{e}_{|I_3|}, & \tilde{y}^2 := -\widetilde{A}_2 \nabla_\gamma \theta(\bar{v})_{I_1}, \\
\tilde{y}^3 := -\widetilde{A}_2 P_{(I_1,\cdot)}, & \mathbb{Y}^1 := \widehat{Y}_{([k],I_2)} - \widetilde{A}_2 Q(\bar{\gamma})_{(I_1,I_2)}.
\end{cases}
\tag{135}
$$

Furthermore, let

$$
\begin{array}{ll}
I_1 := \bar{\Pi}_{21} \cup \phi_{\gamma 2} \cup \bar{\Theta}_{22}, & I_2 := \bar{\Pi}_{11} \cup \phi_{\gamma 1} \cup \bar{\Theta}_{12} \cup \bar{\Pi}_{31} \cup \phi_{\gamma 3} \cup \bar{\Theta}_{32}, \\
I_3 := \bigcup\limits_{i=1,2,3} \Theta_{i2}, & I' := \Pi_{21} \cup \Theta_{22}.
\end{array}
$$

$\bar{I} := [n] \backslash (I_1 \cup I_2 \cup I_3)$, $I^0 := \Pi_{11} \cup \bar{\Pi}_{11} \cup \Theta_{12} \cup \bar{\Theta}_{12} \cup \phi_{\gamma 1}$, $\bar{I}^0 := [n] \backslash I^0$.

Proposition 42. *Let $\bar{\Lambda}_1^{\zeta^1} := I \backslash \Lambda_1^{\zeta^1}$. It holds that $\bar{\Lambda}_1^{\zeta^1} = I_1 \cup I_2 \cup I_3$.*

Proof By the definition of I_1, I_2 and I_3, it is easy to see that

$$
\begin{aligned}
I_1 \cup I_2 \cup I_3 &= \bar{\Pi}_{21} \cup \phi_{\gamma 2} \cup \bar{\Theta}_{22} \cup \bar{\Pi}_{11} \cup \phi_{\gamma 1} \cup \bar{\Theta}_{12} \cup \bar{\Pi}_{31} \cup \phi_{\gamma 3} \cup \bar{\Theta}_{32} \cup (\bigcup_{i=1,2,3} \Theta_{i2}) \\
&= (\bigcup_{i=1,2,3} \Theta_i) \cup (\bigcup_{i=1,2,3} \phi_{\gamma i}) \cup (\bigcup_{i=1,2,3} \bar{\Pi}_{i1}) \\
&= \Lambda_3^+ \cup \Lambda_3^c \cup (\Lambda_1 \backslash \Lambda_1^{\zeta^1}) \\
&= I \backslash \Lambda_1^{\zeta^1} \\
&= \bar{\Lambda}_1^{\zeta^1}.
\end{aligned}
$$

This concludes the proof. ∎

Lemma 43. *Let $\bar{v}$ be a weakly stationary point of problem* (MPEC-BHO) *with $(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$ as a corresponding Lagrange multiplier vector. Let Assumption* 6 *hold.*

(i) *If $\bar{\gamma} > 0$, for any $d \in \mathcal{C}(\bar{v})$, it holds that*

$$
d^\top \nabla^2_{vv} \mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}) d = \bar{d}^\top V^\top D V \bar{d}
$$

with

$$
\bar{d} := \begin{bmatrix} d^C \\ d^\gamma \\ d^\alpha_{I_2} \\ d^u \end{bmatrix} \quad \text{and} \quad V := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \bar{a}^1 & \bar{b}^1 & \bar{Q}^1 & \bar{P}^1 \\ 0 & 0 & I_{(I_2,I_2)} & 0 \\ \mathbf{e}_{|I_3|} & 0 & 0 & 0 \end{bmatrix}. \tag{136}
$$

Here, $d^{\alpha}_{I_2}$, d^C, d^{γ}, *and* d^u *satisfy*

$$M\bar{d} = 0 \quad with \quad M \quad := \quad \begin{bmatrix} \bar{a}^2 & \bar{b}^2 & \bar{Q}^2 & \bar{P}^2 \\ \tilde{y}^1 & \tilde{y}^2 & \mathbb{Y}^1 & \tilde{y}^3 \end{bmatrix} \tag{137}$$

and $D := D^1 - D^2$, *where*

$$D^1 := \begin{bmatrix} \nabla^2_{CC} f(\bar{v}) & \nabla_{C\gamma} f(\bar{v}) & \nabla^2_{C\alpha_{\bar{\Lambda}_1^{\zeta^1}}} f(\bar{v}) \\ \nabla^2_{\gamma C} f(\bar{v}) & \nabla^2_{\gamma\gamma} f(\bar{v}) & \nabla^2_{\gamma\alpha_{\bar{\Lambda}_1^{\zeta^1}}} f(\bar{v}) \\ \nabla^2_{C\alpha_{\bar{\Lambda}_1^{\zeta^1}}} f(\bar{v})^\top & \nabla^2_{\gamma\alpha_{\bar{\Lambda}_1^{\zeta^1}}} f(\bar{v})^\top & \nabla^2_{\alpha_{\bar{\Lambda}_1^{\zeta^1}}\alpha_{\bar{\Lambda}_1^{\zeta^1}}} f(\bar{v}) \end{bmatrix}, \tag{138}$$

$$D^2 := \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sum\limits_{i\in\Lambda_3} \bar{\eta}^1_i \nabla^2_{\gamma\gamma}\theta_i(\bar{v}) & \sum\limits_{i\in\Lambda_3} \bar{\eta}^1_i \nabla^2_{\gamma\alpha_{\bar{\Lambda}_1^{\zeta^1}}}\theta_i(\bar{v}) \\ 0 & \sum\limits_{i\in\Lambda_3} \bar{\eta}^1_i \nabla^2_{\gamma\alpha_{\bar{\Lambda}_1^{\zeta^1}}}\theta_i(\bar{v})^\top & \mathbf{0}_{(\bar{\Lambda}_1^{\zeta^1}, \bar{\Lambda}_1^{\zeta^1})} \end{bmatrix}. \tag{139}$$

(ii) *If* $\bar{\gamma} = 0$ *and* $\bar{\lambda} > 0$, *for any* $d \in \mathcal{C}(\bar{v})$, *we have*

$$d^\top \nabla^2_{vv}\mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})d = \bar{d}^\top V^\top D V \bar{d} > 0, \quad \forall \bar{d} \tag{140}$$

with

$$\bar{d} := \begin{bmatrix} d^C \\ d^{\alpha}_{I_1} \\ d^{\alpha}_{I_2} \end{bmatrix} \quad and \quad V := \begin{bmatrix} 1 & 0 & 0 \\ 0 & \mathbf{I}_{(|I_1|,|I_1|)} & 0 \\ 0 & 0 & I_{(|I_2|,|I_2|)} \\ \mathbf{e}_{|I_3|} & 0 & 0 \end{bmatrix}. \tag{141}$$

Here $d^{\alpha}_{I_1}, d^{\alpha}_{I_2}$, d^C, d^u *satisfy*

$$\begin{cases} \begin{bmatrix} Q(\bar{\gamma})_{(I_1,I_3)}\mathbf{e}_{|I_3|} & Q(\bar{\gamma})_{(I_1,I_1)} & Q(\bar{\gamma})_{(I_1,I_2)} & P_{(I_1,\cdot)} \\ Q(\bar{\gamma})_{(I',I_3)}\mathbf{e}_{|I_3|} & Q(\bar{\gamma})_{(I',I_1)} & Q(\bar{\gamma})_{(I',I_2)} & P_{(I',\cdot)} \\ \widehat{Y}_{([k],I_3)}\mathbf{e}_{|I_3|} & \widehat{Y}_{([k],I_1)} & \widehat{Y}_{([k],I_2)} & \mathbf{0}_{([k],L_5)} \end{bmatrix} \cdot \begin{bmatrix} d^C \\ d^{\alpha}_{I_1} \\ d^{\alpha}_{I_2} \\ d^u \end{bmatrix} = 0, \\ D := \begin{bmatrix} \nabla^2_{CC} f(\bar{v}) & \nabla^2_{C\alpha_{\bar{\Lambda}_1^{\zeta^1}}} f(\bar{v}) \\ \nabla^2_{C\alpha_{\bar{\Lambda}_1^{\zeta^1}}} f(\bar{v})^\top & \nabla^2_{\alpha_{\bar{\Lambda}_1^{\zeta^1}}\alpha_{\bar{\Lambda}_1^{\zeta^1}}} f(\bar{v}) \end{bmatrix}. \end{cases} \tag{142}$$

Proof (i) Firstly, by the basic row transformations, W in Proposition 41 can be equivalently written as the following form

$$
W = \begin{bmatrix}
\mathbf{0}_{(\Pi_{11},L_1)} & \nabla_\gamma\theta(\bar v)_{\Pi_{11}} & Q(\bar\gamma)_{(\Pi_{11},\cdot)} & \mathcal{I}_{(\Pi_{11},\cdot)} & P_{(\Pi_{11},\cdot)} \\
\mathbf{0}_{(\bar\Pi_{11},L_1)} & \nabla_\gamma\theta(\bar v)_{\bar\Pi_{11}} & Q(\bar\gamma)_{(\bar\Pi_{11},\cdot)} & \mathcal{I}_{(\bar\Pi_{11},\cdot)} & P_{(\bar\Pi_{11},\cdot)} \\
\mathbf{0}_{(\Pi_{21},L_1)} & \nabla_\gamma\theta(\bar v)_{\Pi_{21}} & Q(\bar\gamma)_{(\Pi_{21},\cdot)} & \mathbf{0}_{(\Pi_{21},L4)} & P_{(\Pi_{21},\cdot)} \\
\mathbf{0}_{(\bar\Pi_{21},L_1)} & \nabla_\gamma\theta\bar v_{\bar\Pi_{21}} & Q(\bar\gamma)_{(\bar\Pi_{21},\cdot)} & \mathbf{0}_{(\bar\Pi_{21},L_4)} & P_{(\bar\Pi_{21},\cdot)} \\
\mathbf{0}_{(\phi_{\gamma1},L_1)} & \nabla_\gamma\theta(\bar v)_{\phi_{\gamma1}} & Q(\bar\gamma)_{(\phi_{\gamma1},\cdot)} & \mathcal{I}_{(\phi_{\gamma1},L_4)} & P_{(\phi_{\gamma1},\cdot)} \\
\mathbf{0}_{(\phi_{\gamma2},L_1)} & \nabla_\gamma\theta\bar v_{\phi_{\gamma2}} & Q(\bar\gamma)_{(\phi_{\gamma2},\cdot)} & \mathbf{0}_{(\phi_{\gamma2},L_4)} & P_{(\phi_{\gamma2},\cdot)} \\
\mathbf{0}_{(\Theta_{12},L_1)} & \nabla_\gamma\theta(\bar v)_{\Theta_{12}} & Q(\bar\gamma)_{(\Theta_{12},\cdot)} & \mathcal{I}_{(\Theta_{12},L_4)} & P_{(\Theta_{12},\cdot)} \\
\mathbf{0}_{(\bar\Theta_{12},L_1)} & \nabla_\gamma\theta(\bar v)_{\bar\Theta_{12}} & Q(\bar\gamma)_{(\bar\Theta_{12},\cdot)} & \mathcal{I}_{(\bar\Theta_{12},L_4)} & P_{(\bar\Theta_{12},\cdot)} \\
\mathbf{0}_{(\Theta_{22},L_1)} & \nabla_\gamma\theta(\bar v)_{\Theta_{22}} & Q(\bar\gamma)_{(\Theta_{22},\cdot)} & \mathbf{0}_{(\Theta_{22},L_4)} & P_{(\Theta_{22},\cdot)} \\
\mathbf{0}_{(\bar\Theta_{22},L_1)} & \nabla_\gamma\theta\bar v_{\bar\Theta_{22}} & Q(\bar\gamma)_{(\bar\Theta_{22},\cdot)} & \mathbf{0}_{(\bar\Theta_{22},L_4)} & P_{(\bar\Theta_{22},\cdot)} \\
\mathbf{0}_{(\Pi_{11},L_1)} & \mathbf{0}_{(\Pi_{11},L_2)} & \mathcal{I}_{(\Pi_{11},\cdot)} & \mathbf{0}_{(\Pi_{11},L_4)} & \mathbf{0}_{(\Pi_{11},L_5)} \\
\mathbf{0}_{(\Pi_{21},L_1)} & \mathbf{0}_{(\Pi_{21},L_2)} & \mathcal{I}_{(\Pi_{21},\cdot)} & \mathbf{0}_{(\Pi_{21},L_4)} & \mathbf{0}_{(\Pi_{21},L_5)} \\
\mathbf{0}_{(\Pi_{31},L_1)} & \mathbf{0}_{(\Pi_{31},L_2)} & \mathcal{I}_{(\Pi_{31},\cdot)} & \mathbf{0}_{(\Pi_{31},L_4)} & \mathbf{0}_{(\Pi_{31},L_5)} \\
\mathbf{e}_{(\Theta_{12},L_1)} & \mathbf{0}_{((\Theta_{12},L_2)} & -\mathcal{I}_{(\Theta_{12},\cdot)} & \mathbf{0}_{(\Theta_{12},L_4)} & \mathbf{0}_{(\Theta_{12},L_5)} \\
\mathbf{e}_{(\Theta_{22},L_1)} & \mathbf{0}_{((\Theta_{22},L_2)} & -\mathcal{I}_{(\Theta_{22},\cdot)} & \mathbf{0}_{(\Theta_{22},L_4)} & \mathbf{0}_{(\Theta_{22},L_5)} \\
\mathbf{e}_{(\Theta_{32},L_1)} & \mathbf{0}_{((\Theta_{32},L_2)} & -\mathcal{I}_{(\Theta_{32},\cdot)} & \mathbf{0}_{(\Theta_{32},L_4)} & \mathbf{0}_{(\Theta_{32},L_5)} \\
\mathbf{0}_{(\Pi_{21},L_1)} & \mathbf{0}_{(\Pi_{21},L_2)} & \mathbf{0}_{(\Pi_{21},L_3)} & \mathcal{I}_{(\Pi_{21},\cdot)} & \mathbf{0}_{(\Pi_{11},L_5)} \\
\mathbf{0}_{(\bar\Pi_{21},L_1)} & \mathbf{0}_{(\bar\Pi_{21},L_2)} & \mathbf{0}_{(\bar\Pi_{21},L_3)} & \mathcal{I}_{(\bar\Pi_{21},\cdot)} & \mathbf{0}_{(\bar\Pi_{21},L_5)} \\
\mathbf{0}_{(\Pi_{31},L_1)} & \mathbf{0}_{(\Pi_{31},L_2)} & \mathbf{0}_{(\Pi_{31},L_3)} & \mathcal{I}_{(\Pi_{31},\cdot)} & \mathbf{0}_{(\Pi_{31},L_5)} \\
\mathbf{0}_{(\bar\Pi_{31},L_1)} & \mathbf{0}_{(\bar\Pi_{31},L_2)} & \mathbf{0}_{(\bar\Pi_{31},L_3)} & \mathcal{I}_{(\bar\Pi_{31},\cdot)} & \mathbf{0}_{(\bar\Pi_{31},L_5)} \\
\mathbf{0}_{(\phi_{\gamma2},L_1)} & \mathbf{0}_{(\phi_{\gamma2},L_2)} & \mathbf{0}_{(\phi_{\gamma2},L_3)} & \mathcal{I}_{(\phi_{\gamma2},\cdot)} & \mathbf{0}_{(\phi_{\gamma2},L_5)} \\
\mathbf{0}_{(\phi_{\gamma3},L_1)} & \mathbf{0}_{(\phi_{\gamma3},L_2)} & \mathbf{0}_{(\phi_{\gamma3},L_3)} & \mathcal{I}_{(\phi_{\gamma3},\cdot)} & \mathbf{0}_{(\phi_{\gamma3},L_5)} \\
\mathbf{0}_{(\Theta_{22},L_1)} & \mathbf{0}_{(\Theta_{22},L_2)} & \mathbf{0}_{(\Theta_{22},L_3)} & \mathcal{I}_{(\Theta_{22},\cdot)} & \mathbf{0}_{(\Theta_{22},L_5)} \\
\mathbf{0}_{(\Theta_{32},L_1)} & \mathbf{0}_{(\Theta_{32},L_2)} & \mathbf{0}_{(\Theta_{32},L_3)} & \mathcal{I}_{(\Theta_{32},\cdot)} & \mathbf{0}_{(\Theta_{32},L_5)} \\
\mathbf{0}_{(\bar\Theta_{22},L_1)} & \mathbf{0}_{(\bar\Theta_{22},L_2)} & \mathbf{0}_{(\bar\Theta_{22},L_3)} & \mathcal{I}_{(\bar\Theta_{22},\cdot)} & \mathbf{0}_{(\bar\Theta_{22},L_5)} \\
\mathbf{0}_{(\bar\Theta_{32},L_1)} & \mathbf{0}_{(\bar\Theta_{32},L_2)} & \mathbf{0}_{(\bar\Theta_{32},L_3)} & \mathcal{I}_{(\bar\Theta_{32},\cdot)} & \mathbf{0}_{(\bar\Theta_{32},L_5)} \\
\mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \widehat{Y} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
Q(\bar\gamma)_{(I^0,I_3)}\mathbf{e}_{|I_3|} & \nabla_\gamma\theta(\bar v)_{I^0} & Q(\bar\gamma)_{(I^0,I_1\cup I_2)} & \mathbf{0}_{(I^0,I_3\cup\bar I)} & \mathcal{I}_{(I^0,L4)} & P_{(I^0,\cdot)} \\
Q(\bar\gamma)_{(I_1,I_3)}\mathbf{e}_{|I_3|} & \nabla_\gamma\theta(\bar v)_{I_1} & Q(\bar\gamma)_{(I_1,I_1\cup I_2)} & \mathbf{0}_{(I_1,I_3\cup\bar I)} & \mathbf{0}_{(I_1,L_4)} & P_{(I_1,\cdot)} \\
Q(\bar\gamma)_{(I',I_3)}\mathbf{e}_{|I_3|} & \nabla_\gamma\theta(\bar v)_{I'} & Q(\bar\gamma)_{(I',I_1\cup I_2)} & \mathbf{0}_{(I',I_3\cup\bar I)} & \mathbf{0}_{(I',L_4)} & P_{(I',\cdot)} \\
\mathbf{0}_{(\bar I,L_1)} & \mathbf{0}_{(\bar I,L_2)} & \mathbf{0}_{(\bar I,I_1\cup I_2)} & -\mathcal{I}_{(\bar I,I_3\cup\bar I)} & \mathbf{0}_{(\bar I,L_4)} & \mathbf{0}_{(\bar I,L_5)} \\
\mathbf{e}_{(I_3,L_1)} & \mathbf{0}_{(I_3,L_2)} & \mathbf{0}_{(I_3,I_1\cup I_2)} & -\mathcal{I}_{(I_3,I_3\cup\bar I)} & \mathbf{0}_{(I_3,L_4)} & \mathbf{0}_{(I_3,L_5)} \\
\mathbf{0}_{(\bar I^0,L_1)} & \mathbf{0}_{(\bar I^0,L_2)} & \mathbf{0}_{(\bar I^0,I_1\cup I_2)} & \mathbf{0}_{(\bar I^0,I_3\cup\bar I)} & \mathcal{I}_{(\bar I^0,\cdot)} & \mathbf{0}_{(\bar I^0,L_5)} \\
\widehat{Y}_{([k],I_3)}\mathbf{e}_{|I_3|} & \mathbf{0}_{([k],L_2)} & \widehat{Y}_{([k],I_1\cup I_2)} & \mathbf{0}_{([k],I_3\cup\bar I)} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)}
\end{bmatrix}.
$$

By solving $\overline{W}d = 0$, we obtain that $d^\alpha_{\bar I} = 0$, $d^\alpha_{I_3} = -\mathbf{e}_{|I_3|}d^C$, $d^\sigma_{\bar I^0} = 0$, which gives

$$d^\alpha_{\Pi_{i1}} = 0 \text{ and } d^\alpha_{\Theta_{i2}} = -\mathbf{e}_{|\Theta_{i2}|}d^C \text{ for } i = 1,\,2,\,3, \tag{143}$$

$$d^\sigma_{\Pi_{i1}} = 0,\; d^\sigma_{\bar\Pi_{i1}} = 0,\; d^\sigma_{\phi_{\gamma i}} = 0,\; d^\sigma_{\Theta_{i2}} = 0,\; \text{ and } d^\sigma_{\bar\Theta_{i2}} = 0 \text{ for } i = 2,\,3, \tag{144}$$

$$d^\sigma_{I^0} = -Q(\bar\gamma)_{(I_0,I_3)}\mathbf{e}_{|I_3|}d^C - \nabla_\gamma\theta(\bar v)_{I^0}d^\gamma - Q(\bar\gamma)_{(I^0,I_1\cup I_2)}d^\alpha_{I_1\cup I_2} - P_{(I^0,\cdot)}d^u, \tag{145}$$

and

$$0 \;=\; \begin{bmatrix} Q(\bar\gamma)_{(I_1,I_3)}\mathbf{e}_{|I_3|} & \nabla_\gamma\theta(\bar v)_{I_1} & Q(\bar\gamma)_{(I_1,I_1)} & Q(\bar\gamma)_{(I_1,I_2)} & P_{(I_1,\cdot)} \\ Q(\bar\gamma)_{(I',I_3)}\mathbf{e}_{|I_3|} & \nabla_\gamma\theta(\bar v)_{I'} & Q(\bar\gamma)_{(I',I_1)} & Q(\bar\gamma)_{(I',I_2)} & P_{(I',\cdot)} \\ \widehat{Y}_{([k],I_3)}\mathbf{e}_{|I_3|} & \mathbf{0}_{([k],L_2)} & \widehat{Y}_{([k],I_1)} & \widehat{Y}_{([k],I_2)} & \mathbf{0}_{([k],L_5)} \end{bmatrix} \cdot \begin{bmatrix} d^C \\ d^\gamma \\ d^\alpha_{I_1} \\ d^\alpha_{I_2} \\ d^u \end{bmatrix}$$

Note that $Q(\bar\gamma)_{(I_1,I_1)}$ is positive definite. Let $\widetilde{A}_1$ and $\widetilde{A}_2$ be matrices of suitable sizes such that (64) holds. After proper basic row transformations, it holds that

$$\begin{bmatrix} Q(\bar\gamma)_{(I_1,I_3)}\mathbf{e}_{|I_3|} & \nabla_\gamma\theta(\bar v)_{I_1} & Q(\bar\gamma)_{(I_1,I_1)} & Q(\bar\gamma)_{(I_1,I_2)} & P_{(I_1,\cdot)} \\ \bar a^2 & \bar b^2 & \mathbf{0}_{(I',I_1)} & \bar Q^2 & \bar P^2 \\ \tilde y^1 & \tilde y^2 & \mathbf{0}_{([k],I_1)} & \mathbb{Y}^1 & \tilde y^3 \end{bmatrix} \begin{bmatrix} d^C \\ d^\gamma \\ d^\alpha_{I_1} \\ d^\alpha_{I_2} \\ d^u \end{bmatrix} = 0. \tag{146}$$

Here $\bar a^2$, $\bar b^2$, $\bar Q^2$, $\bar P^2$, $\tilde y^1$, $\tilde y^2$, $\mathbb{Y}^1$, and $\tilde y^3$ are defined in (66). Then we obtain that

$$d^\alpha_{I_1} = \bar a_1 d^C + \bar b_1 d^\gamma + \bar Q^1 d^\alpha_{I_2} + \bar P^1 d^u, \tag{147}$$

where $d^\alpha_{I_2}$, d^C, d^γ, and d^u satisfy the following linear system of equations:

$$\begin{aligned} \bar a^2 d^C + \bar b^2 d^\gamma + \bar Q^2 d^\alpha_{I_2} + \bar P^2 d^u &= 0, \\ \tilde y^1 d^C + \tilde y^2 d^\gamma + \mathbb{Y}^1 d^\alpha_{I_2} + \tilde y^3 d^u &= 0. \end{aligned}$$

(145) reduces to the following

$$d^\sigma_{I^0} = z^1 d^C + z^2 d^\gamma + z^3 d^\alpha_{I_2} + z^4 d^u, \tag{148}$$

where

$$z^1 := Q(\bar\gamma)_{(I^0,I_3)}\mathbf{e}_{|I_3|} - Q(\bar\gamma)_{(I^0,I_1)}\bar a^1, \;\; z^2 := -\nabla_\gamma\theta(\bar v)_{I^0} - Q(\bar\gamma)_{(I^0,I_1)}\bar b^1,$$

$$z^3 := -Q(\bar\gamma)_{(I^0,I_2)} - Q(\bar\gamma)_{(I^0,I_1)}\bar Q^1, \;\; z^4 := -\Omega_{(I^0,L_5)} - Q(\bar\gamma)_{(I^0,I_1)}\bar P^1.$$

Then $\mathcal{C}(\bar v)$ can be characterized as follows

$$\begin{aligned} \mathcal{C}(\bar v) \;=\; & \{d =: (d^C,\ d^\gamma,\ d^\alpha,\ d^\sigma,\ d^u) : \\ & d \text{ satisfies (137), (143), (148), other components in } d \text{ are } 0\}. \end{aligned}$$

Therefore, we have

$$d \;=\; \begin{bmatrix} d^C \\ d^\gamma \\ d^\alpha_{I_1} \\ d^\alpha_{I_2} \\ d^\alpha_{I_3} \\ d^\alpha_{\bar I} \\ d^\sigma_{I^0} \\ d^\sigma_{\bar I^0} \\ d^u \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \bar a^1 & \bar b^1 & \bar Q^1 & \bar P^1 \\ 0 & 0 & I_{(I_2,I_2)} & 0 \\ \mathbf{e}_{|I_3|} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{k\times k} \end{bmatrix} \cdot \begin{bmatrix} d^C \\ d^\gamma \\ d^\alpha_{I_2} \\ d^u \end{bmatrix}.$$

By the definition in (136), it is easy to verify that

$$V\bar{d} = \begin{bmatrix} d^C \\ d^\gamma \\ d^\alpha_{I_1} \\ d^\alpha_{I_2} \\ d^\alpha_{I_3} \end{bmatrix}.$$

Recall that $\nabla^2_{vv}\mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\eta})$ is 0 with respect to the block $\underline{\lambda}$, and u. Therefore, we obtain that $d^\top \nabla^2_{vv}\mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\eta})d = \bar{d}^\top V^\top D V \bar{d}$, where $D = D^1 - D^2$, with D^1 and D^2 given by (138) and (139). By Prpopsotion 42, $I_1 \cup I_2 \cup I_3 = \bar{\Lambda}^{\zeta^1}_1$. We obtain the result in (i).

(ii) The proof is similar to that in (i) but with the fact that $d^\gamma = 0$. Firstly, by basic row transformations, W can be equivalently written as the following form

$$W^1 \quad := \quad \begin{bmatrix}
\mathbf{0}_{(I^0,L_1)} & \mathbf{0}_{(I^0,L_2)} & Q(\bar{\gamma})_{(I^0,I_1\cup I_2)} & Q(\bar{\gamma})_{(I^0,I_3\cup \bar{I})} & \mathcal{I}_{(I^0,L_4)} & P_{(I^0,\cdot)} \\
\mathbf{0}_{(I_1,L_1)} & \mathbf{0}_{(I_1,L_2)} & Q(\bar{\gamma})_{(I_1,I_1\cup I_2)} & Q(\bar{\gamma})_{(I_1,I_3\cup \bar{I})} & \mathbf{0}_{(I_1,L_4)} & P_{(I_1,\cdot)} \\
\mathbf{0}_{(I',L_1)} & \mathbf{0}_{(I',L_2)} & Q(\bar{\gamma})_{(I',I_1\cup I_2)} & Q(\bar{\gamma})_{(I',I_3\cup \bar{I})} & \mathbf{0}_{(I',L_4)} & P_{(I',\cdot)} \\
\mathbf{0}_{(\bar{I},L_1)} & \mathbf{0}_{(\bar{I},L_2)} & \mathbf{0}_{(\bar{I},I_1\cup I_2)} & -\mathcal{I}_{(\bar{I},I_3\cup \bar{I})} & \mathbf{0}_{(\bar{I},L_4)} & \mathbf{0}_{(\bar{I},L_5)} \\
\mathbf{e}_{(I_3,L_1)} & \mathbf{0}_{(I_3,L_2)} & \mathbf{0}_{(I_3,I_1\cup I_2)} & -\mathcal{I}_{(I_3,I_3\cup \bar{I})} & \mathbf{0}_{(I_3,L_4)} & \mathbf{0}_{(I_3,L_5)} \\
\mathbf{0}_{(\bar{I}^0,L_1)} & \mathbf{0}_{(\bar{I}^0,L_2)} & \mathbf{0}_{(\bar{I}^0,I_1\cup I_2)} & \mathbf{0}_{(\bar{I}^0,I_3\cup \bar{I})} & \mathcal{I}_{(\bar{I}^0,\cdot)} & \mathbf{0}_{(\bar{I}^0,L_5)} \\
\mathbf{0}_{([k],L_1)} & \mathbf{0}_{([k],L_2)} & \widehat{Y}_{([k],I_1\cup I_2)} & \widehat{Y}_{([k],I_3\cup \bar{I})} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \\
\mathbf{0}_{(I_g,L_1)} & -\mathbf{e}_{(I_g,L_2)} & \mathbf{0}_{(I_g,I_1\cup I_2)} & \mathbf{0}_{(I_g,I_3\cup \bar{I})} & \mathbf{0}_{(I_g,L_4)} & \mathbf{0}_{(I_g,L_5)}
\end{bmatrix}.$$

By conducting basic row transformations, we obtain the following matrix

$$\overline{W} \quad := \quad \begin{bmatrix}
Q(\bar{\gamma})_{(I^0,I_3)} & \mathbf{0}_{(I^0,L_2)} & Q(\bar{\gamma})_{(I^0,I_1\cup I_2)} & \mathbf{0}_{(I^0,I_3\cup \bar{I})} & \mathcal{I}_{(I^0,L_4)} & P_{(I^0,\cdot)} \\
Q(\bar{\gamma})_{(I_1,I_3)}\mathbf{e}_{|I_3|} & \mathbf{0}_{(I_1,L_2)} & Q(\bar{\gamma})_{(I_1,I_1\cup I_2)} & \mathbf{0}_{(I_1,I_3\cup \bar{I})} & \mathbf{0}_{(I_1,L4)} & P_{(I_1,\cdot)} \\
Q(\bar{\gamma})_{(I',I_3)}\mathbf{e}_{|I_3|} & \mathbf{0}_{(I',L_2)} & Q(\bar{\gamma})_{(I',I_1\cup I_2)} & \mathbf{0}_{(I',I_3\cup \bar{I})} & \mathbf{0}_{(I',L4)} & P_{(I',\cdot)} \\
\mathbf{0}_{(\bar{I},L_1)} & \mathbf{0}_{(\bar{I},L_2)} & \mathbf{0}_{(\bar{I},I_1\cup I_2)} & \mathcal{I}_{(\bar{I},I_3\cup \bar{I})} & \mathbf{0}_{(\bar{I},L_4)} & \mathbf{0}_{(\bar{I},L_5)} \\
\mathbf{e}_{(I_3,L_1)} & \mathbf{0}_{(I_3,L_2)} & \mathbf{0}_{(I_3,I_1\cup I_2)} & \mathcal{I}_{(I_3,I_3\cup \bar{I})} & \mathbf{0}_{(I_3,L_4)} & \mathbf{0}_{(I_3,L_5)} \\
\mathbf{0}_{(\bar{I}^0,L_1)} & \mathbf{0}_{(\bar{I}^0,L_2)} & \mathbf{0}_{(\bar{I}^0,I_1\cup I_2)} & \mathbf{0}_{(\bar{I}^0,I_3\cup \bar{I})} & \mathcal{I}_{(\bar{I}^0,\cdot)} & \mathbf{0}_{(\bar{I}^0,L_5)} \\
\widehat{Y}_{([k],I_3)}\mathbf{e}_{|I_3|} & \mathbf{0}_{([k],L_2)} & \widehat{Y}_{([k],I_1\cup I_2)} & \mathbf{0}_{([k],I_3\cup \bar{I})} & \mathbf{0}_{([k],L_4)} & \mathbf{0}_{([k],L_5)} \\
\mathbf{0}_{(I_g,L_1)} & -\mathbf{e}_{(I_g,L_2)} & \mathbf{0}_{(I_g,I_1\cup I_2)} & \mathbf{0}_{(I_g,I_3\cup \bar{I})} & \mathbf{0}_{(I_g,L_4)} & \mathbf{0}_{(I_g,L_5)}
\end{bmatrix}$$

By solving $\overline{W}d = 0$, we obtain that (143), (144), $d^\gamma = 0$,

$$\begin{aligned} d^\sigma_{I^0} &= -Q(\bar{\gamma})_{(I_0,I_3)}\mathbf{e}_{|I_3|}d^C - Q(\bar{\gamma})_{(I^0,I_1\cup I_2)}d^\alpha_{I_1\cup I_2} - P_{(I^0,\cdot)}d^u \\ &:= \tilde{z}^1 d^C + \tilde{z}^2 d^\alpha_{I_1} + \tilde{z}^3 d^\alpha_{I_2} + \tilde{z}^4 d^u, \end{aligned} \tag{149}$$

and

$$0 \quad = \quad \begin{bmatrix} Q(\bar{\gamma})_{(I_1,I_3)}\mathbf{e}_{|I_3|} & Q(\bar{\gamma})_{(I_1,I_1)} & Q(\bar{\gamma})_{(I_1,I_2)} & P_{(I_1,\cdot)} \\ Q(\bar{\gamma})_{(I',I_3)}\mathbf{e}_{|I_3|} & Q(\bar{\gamma})_{(I',I_1)} & Q(\bar{\gamma})_{(I',I_2)} & P_{(I',\cdot)} \\ \widehat{Y}_{([k],I_3)}\mathbf{e}_{|I_3|} & \widehat{Y}_{([k],I_1)} & \widehat{Y}_{([k],I_2)} & \mathbf{0}_{([k],L_5)} \end{bmatrix} \cdot \begin{bmatrix} d^C \\ d^\alpha_{I_1} \\ d^\alpha_{I_2} \\ d^u \end{bmatrix}. \tag{150}$$

Then $\mathcal{C}(\bar{v})$ can be characterized as follows

$$\begin{aligned}\mathcal{C}(\bar{v}) &= \{d := (d^C\ d^\gamma\ d^\alpha\ d^\sigma\ d^u) : \\ &\qquad d \text{ satisfies (150), (149), other components in } d \text{ are } 0\}.\end{aligned}$$

Therefore, we have

$$d = \begin{bmatrix} d^C \\ d^\gamma \\ d^\alpha_{I_1} \\ d^\alpha_{I_2} \\ d^\alpha_{I_3} \\ d^\alpha_{\bar{I}} \\ d^\sigma_{I^0} \\ d^\sigma_{\bar{I}^0} \\ d^u \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \mathbf{I}_{(|I_1|,|I_1|)} & 0 & 0 \\ 0 & 0 & I_{(|I_2|,|I_2|)} & 0 \\ \mathbf{e}_{|I_3|} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \tilde{z}^1 & \tilde{z}^2 & \tilde{z}^3 & \tilde{z}^4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{k\times k} \end{bmatrix} \cdot \begin{bmatrix} d^C \\ d^\alpha_{I_1} \\ d^\alpha_{I_2} \\ d^u \end{bmatrix}.$$

By the definition in (141), it is easy to verify that

$$V\bar{d} = \begin{bmatrix} d^C \\ d^\alpha_{I_1} \\ d^\alpha_{I_2} \\ d^\alpha_{I_3} \end{bmatrix}.$$

Recall that $\nabla^2_{vv}\mathbb{L}(\bar{v},\bar{\lambda},\bar{\mu},\bar{\eta},\bar{\zeta})$ is 0 w.r.t. the blocks $\underline{\lambda}$ and u. Therefore, we obtain that $d^\top\nabla^2_{vv}\mathbb{L}(\bar{v},\bar{\lambda},\bar{\mu},\bar{\eta},\bar{\zeta})d = \bar{d}^\top V^\top D^1 V\bar{d}$, where with D^1 given by (165). Hence, the proof. ■

Proposition 44. *For a weakly stationary point $\bar{v}$ of the* (MPEC-BHO) *with multiplier vector $(\bar{\lambda},\bar{\mu},\bar{\eta},\bar{\zeta})$ satisfying $\Lambda_2 \cup \Lambda_4 = \emptyset$. Let Assumption* 6 *and Assumption* 7 *hold.*

(i) If $\bar{\gamma} > 0$ satisfying $\mid J_{GH^1} \mid + \mid J_{GH^2} \mid = 0$ for any $d \in \mathcal{C}(\bar{v})$, the following result holds

$$d^\top\nabla^2_{vv}\mathbb{L}(\bar{v},\bar{\lambda},\bar{\mu},\bar{\eta},\bar{\zeta})d = \bar{d}^\top V^\top DV\bar{d}, \tag{151}$$

where $\bar{d}$ and V are defined by

$$\bar{d} := \begin{bmatrix} d^C \\ d^\gamma \\ d^u \end{bmatrix} \quad \text{and} \quad V := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \bar{a}^1 & \bar{b}^1 & \bar{P}^1 \\ \mathbf{e}_{|I_3|} & 0 & 0 \end{bmatrix}. \tag{152}$$

Here d^C, d^γ, d^u satisfy

$$M\bar{d} = 0 \quad \text{with} \quad M := \begin{bmatrix} \bar{a} & \bar{b} & \bar{P} \\ \check{y}_1 & \check{y}_2 & \tilde{y}_3 \end{bmatrix} \tag{153}$$

and $D := D^1 - D^2$, where D^1 and D^2 are respectively given by

$$D^1 := \begin{bmatrix} \nabla^2_{CC} f(\bar{v}) & \nabla^2_{C\gamma} f(\bar{v}) & \nabla^2_{C\alpha_{\Lambda_3}} f(\bar{v}) \\ \nabla^2_{\gamma C} f(\bar{v}) & \nabla^2_{\gamma\gamma} f(\bar{v}) & \nabla^2_{\gamma\alpha_{\Lambda_3}} f(\bar{v}) \\ \nabla^2_{C\alpha_{\Lambda_3}} f(\bar{v})^\top & \nabla^2_{\gamma\alpha_{\Lambda_3}} f(\bar{v})^\top & \nabla^2_{\alpha_{\Lambda_3}\alpha_{\Lambda_3}} f(\bar{v}) \end{bmatrix}, \tag{154}$$

$$D^2 := \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sum\limits_{i\in\Lambda_3} \bar{\eta}^1_i \nabla^2_{\gamma\gamma}\theta_i(\bar{v}) & \sum\limits_{i\in\Lambda_3} \bar{\eta}^1_i \nabla^2_{\gamma\alpha_{\Lambda_3}}\theta_i(\bar{v}) \\ 0 & \sum\limits_{i\in\Lambda_3} \bar{\eta}^1_i \nabla^2_{\gamma\alpha_{\Lambda_3}}\theta_i(\bar{v})^\top & \mathbf{0}_{(\Lambda_3,\Lambda_3)} \end{bmatrix}. \tag{155}$$

(ii) If $\bar{\gamma} = 0$ with $\bar{\lambda} > 0$, for any $d \in \mathcal{C}(\bar{v})$, the following result holds

$$d^\top \nabla^2_{vv} \mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}) d = \bar{d}^\top V^\top D V \bar{d},$$

where

$$\bar{d} = \begin{bmatrix} d^C \\ d^\alpha_{\Lambda_3^+} \end{bmatrix}, \quad V = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{I}_{(|\Lambda_3^+|,|\Lambda_3^+|)} \\ \mathbf{e}_{|\Lambda_3^c|} & 0 \end{bmatrix}, \tag{156}$$

D is given by

$$\begin{bmatrix} \nabla^2_{CC} f(\bar{v}) & \nabla^2_{C\alpha_{\Lambda_3}} f(\bar{v}) \\ \nabla^2_{C\alpha_{\Lambda_3}} f(\bar{v})^\top & \nabla^2_{\alpha_{\Lambda_3}\alpha_{\Lambda_3}} f(\bar{v}) \end{bmatrix}, \tag{157}$$

and d^C, $d^\alpha_{\Lambda_3^+}$, d^u satisfy condition (70).

Proof If $\Lambda_2 \cup \Lambda_4 = \emptyset$, by Assumption 7, $\Pi_i = 0,\ \phi_{\gamma i} = 0,\ \Theta_i = 0,\ i = 1,3;\ \bar{\Pi}_{21} = 0,\ \bar{\Theta}_{22} = 0,\ \Pi_{21} = \Lambda_1, \phi_{\gamma 2} = \Lambda_3^+, \Theta_{22} = \Lambda_3^c$. By the definition of $I_i, i = 1,2,3,\ \bar{I}$ and I', we have

$$I_1 = \Lambda_3^+, I_2 = \emptyset, I_3 = \Lambda_3^c,\ \bar{I} = \Lambda_1,\ I' = \Lambda_1 \cup \Lambda_3^c.$$

Hence $\bar{a}^1$, $\bar{b}^1$ and $\bar{P}^1$ reduce to the following form

$$\begin{cases} \bar{a}^1 &= -\left(Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^+)}\right)^{-1} Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^c)|\Lambda_3^c|}, \\ \bar{b}^1 &= -Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^+)})^{-1} \nabla_\gamma \theta(\bar{v})_{\Lambda_3^+}, \\ \bar{P}^1 &= -\left(Q(\bar{\gamma})_{(\Lambda_3^+,\Lambda_3^+)}\right)^{-1} P_{(\Lambda_3^+,\cdot)}. \end{cases} \tag{158}$$

$\bar{a}^2$, $\bar{b}^2$, $\bar{P}^2$ $\tilde{y}^1$, $\tilde{y}^2$ in (135) reduce to $\bar{a}$, $\bar{b}$, $\bar{P}$, $\check{y}^1$, $\check{y}^2$ in (66), respectively If $\bar{\gamma} > 0$, by Lemma 43 (i), we have the results in (i).

If $\bar{\gamma} = 0$, by Lemma 43 (ii), note that $\Lambda_3 \cup \Lambda_1 = [n]$, the coefficient matrix in the first equation system in (142) reduces to the following matrix

$$\begin{bmatrix} Q(\bar{\gamma})_{(I_1,I_3)}\mathbf{e}_{|I_3|} & Q(\bar{\gamma})_{(I_1,I_1)} & P_{(I_1,\cdot)} \\ Q(\bar{\gamma})_{(I',I_3)}\mathbf{e}_{|I_3|} & Q(\bar{\gamma})_{(I',I_1)} & P_{(I',\cdot)} \\ \widehat{Y}_{([k],I_3)}\mathbf{e}_{|I_3|} & \widehat{Y}_{([k],I_1)} & \mathbf{0}_{([k],L_5)} \end{bmatrix} = \begin{bmatrix} Q(\bar{\gamma})_{([n],\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} & Q(\bar{\gamma})_{([n],\Lambda_3^+)} & P_{([n],\cdot)} \\ \widehat{Y}_{([k],\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} & \widehat{Y}_{([k],\Lambda_3^+)} & \mathbf{0}_{([k],L_5)} \end{bmatrix}.$$

We obtain the results in (ii). This concludes the proof. ■

Proof of Theorem 29

(i) By Proposition 18 and Assumption 6, it holds that $\Lambda_2 = \emptyset$, which gives

$$\mid \Lambda_1^{(i)} \cup \Lambda_2^{(i)} \mid = \mid \Lambda_1^{(i)} \mid \geq 2, \quad i \in [k].$$

Therefore, $\mid \Lambda_1 \cup \Lambda_3^c \mid \geq \mid \Lambda_1 \mid = \sum_{i \in [k]} \mid \Lambda_1^{(i)} \mid \geq 2k$. On the other hand, by Proposition 44 (i), for M defined in (153), $M \in \mathbb{R}^{(k+|\Lambda_1 \cup \Lambda_3^c|) \times (k+2)}$. Therefore, $k + \mid \Lambda_1 \cup \Lambda_3^c \mid \geq 3k > k+2$, which gives $rank(M) \leq k+2$.

Let J^0 and $\bar{J}^0$ be defined as in (63). We can rewrite matrix M in (153) as follows

$$M = \begin{bmatrix} \bar{a}_{J^0} & \bar{b}_{J^0} & \bar{P}_{(J^0,\cdot)} \\ \bar{a}_{\bar{J}^0} & \bar{b}_{\bar{J}^0} & \bar{P}_{(\bar{J}^0,\cdot)} \\ \check{y}^1 & \check{y}^2 & \tilde{y}^3 \end{bmatrix}.$$

Note that

$$\bar{P}_{(J^0,\cdot)} = \begin{bmatrix} \hat{y}_{j_1}^{(1)} & 0 & \cdots & 0 \\ 0 & \hat{y}_{j_2}^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{y}_{j_k}^{(k)} \end{bmatrix},$$

implying that $rank(\bar{P}_{(J^0,\cdot)}) = k$. Let $\bar{A}^1$, $\bar{A}_2$ be defined as in (65). By conducting basic row transformations, we obtain that (153) is equivalent to

$$\begin{bmatrix} \bar{a}_{J^0} & \bar{b}_{J^0} & \bar{P}_{(J^0,\cdot)} \\ \bar{a}_{\bar{J}^0} - \bar{A}^1 \bar{a}_{J^0} & \bar{b}_{\bar{J}^0} - \bar{A}^1 \bar{b}_{J^0} & 0 \\ \check{y}^1 - \bar{A}^2 \bar{a}_{J^0} & \check{y}^2 - \bar{A}^2 \bar{b}_{J^0} & 0 \end{bmatrix} \cdot \begin{bmatrix} d^C \\ d^{\Lambda_3^+} \\ d^u \end{bmatrix} = 0.$$

Then we obtain that (recall $\widetilde{M}$ defined in Assumption 9)

$$d^u = -(\bar{P}_{(J^0,\cdot)})^{-1} \left(\bar{a}_{J^0} d^C + \bar{b}_{J^0} d^\gamma \right), \tag{159}$$

$$\widetilde{M} \begin{bmatrix} d^C \\ d^\gamma \end{bmatrix} = 0. \tag{160}$$

If Assumption 9 holds, the solution to (160) is $d^C = 0$, $d^\gamma = 0$; hence, implying that $d^u = 0$, $d^\alpha_{\Lambda_3^c} = 0$. Therefore, $\mathcal{C}(\bar{v}) = \{0\}$, and the MPEC-SSOSC holds automatically.

(ii) By Proposition 44 (ii), note that $\mid \Lambda_3 \cup \Lambda_1 \mid = n$, for the coefficient matrix defined in (70), it is in $\mathbb{R}^{(k+n) \times (1+k+|\Lambda_3^+|)}$. Similar to the argument in (i-i), define J^0 and $\bar{J}^0$ by (63). We can rewrite the coefficient matrix in (70) as follows

$$\begin{bmatrix} Q(\bar{\gamma})_{(J^0,\Lambda_3^c)} \mathbf{e}_{|\Lambda_3^c|} & Q(\bar{\gamma})_{(J^0,\Lambda_3^+)} & P_{(J^0,L_5)} \\ Q(\bar{\gamma})_{(\bar{J}^0,\Lambda_3^c)} \mathbf{e}_{|\Lambda_3^c|} & Q(\bar{\gamma})_{(\bar{J}^0,\Lambda_3^+)} & P_{(\bar{J}^0,L_5)} \\ \widehat{Y}_{([k],\Lambda_3^c)} \mathbf{e}_{|\Lambda_3^c|} & \widehat{Y}_{([k],\Lambda_3^+)} & \mathbf{0}_{([k],L_5)} \end{bmatrix}.$$

Observe that

$$P_{(J^0,L_5)} = \begin{bmatrix} \hat{y}^{(1)}_{j_1} & 0 & \cdots & 0 \\ 0 & \hat{y}^{(2)}_{j_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{y}^{(k)}_{j_k} \end{bmatrix},$$

implying that $rank(P_{(J^0,L_5)}) = k$. Assume that there exists $\bar{A}^3$, such that (67) holds. By conducting basic row transformations, (70) is equivalent to

$$\begin{bmatrix} Q(\bar{\gamma})_{(J^0,\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} & Q(\bar{\gamma})_{(J^0,\Lambda_3^+)} & P_{(J^0,L_5)} \\ Q(\bar{\gamma})_{(\bar{J}^0,\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} - \bar{A}^3 Q(\bar{\gamma})_{(J^0,\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} & Q(\bar{\gamma})_{(\bar{J}^0,\Lambda_3^+)} - \bar{A}^3 Q(\bar{\gamma})_{(J^0,\Lambda_3^+)} & 0 \\ \widehat{Y}_{([k],\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|} & \widehat{Y}_{([k],\Lambda_3^+)} & \mathbf{0}_{([k],L_5)} \end{bmatrix} \begin{bmatrix} d^C \\ d^\alpha_{\Lambda_3^+} \\ d^u \end{bmatrix} = 0.$$

Hence, we obtain that

$$d^u = -(P_{(J^0,L_5)})^{-1}\left(Q(\bar{\gamma})_{(J^0,\Lambda_3^c)}\mathbf{e}_{|\Lambda_3^c|}d^C + Q(\bar{\gamma})_{(J^0,\Lambda_3^+)}d^\alpha_{\Lambda_3^+}\right), \tag{161}$$

where d^C and $d^\alpha_{\Lambda_3^+}$ satisfy

$$\widehat{M}\begin{bmatrix} d^C \\ d^\alpha_{\Lambda_3^+} \end{bmatrix} = 0. \tag{162}$$

Here $\widehat{M}$ is defined in Assumption 10. Therefore, if Assumption 10 holds, the solution to (162) is $d^C = 0$, $d^\alpha_{\Lambda_3^+} = 0$, implying that $d^\alpha_{\Lambda_3^c} = 0$. Therefore, $\mathcal{C}(\bar{v}) = \{0\}$. This means that MPEC-SSOSC holds automatically. □

To prove Theorem 30, we need the following proposition.

Proposition 45. *Let $f(v)$ satisfy Assumption 11. Consider a weakly stationary point $\bar{v}$ of (MPEC-BHO) with multiplier vector $(\bar{\lambda},\bar{\mu},\bar{\eta},\bar{\zeta})$ satisfying Assumptions 6 and 7.*

(i) *For $\bar{\gamma} > 0$ satisfying $\mid J_{GH^1} \mid + \mid J_{GH^2} \mid = 0$, if $rank(\widetilde{M}) < 2$, for any $d \in \mathcal{C}(\bar{v})$, it holds that*

$$d^\top \nabla^2_{vv}\mathbb{L}(\bar{v},\bar{\lambda},\bar{\mu},\bar{\eta},\bar{\zeta})d = d^\gamma(\dot{V}^f)^\top \dot{D}^f \dot{V}^f d^\gamma,$$

where $\dot{V}^f$, $\dot{D}^f$ are defined by (68).

(ii) *For $\bar{\gamma} = 0$ and $\bar{\lambda} > 0$, if $rank(\widehat{M}) <\mid \Lambda_3^+ \mid +1$, for any $d \in \mathcal{C}(\bar{v})$, we have*

$$d^\top \nabla^2_{vv}\mathbb{L}(\bar{v},\bar{\lambda},\bar{\mu},\bar{\eta},\bar{\zeta})d = \hat{d}^\top (V^f)^\top D^f V^f \hat{d}$$

with $\hat{d}$, V^f and D^f given in Assumption 13, $d^C, d^\alpha_{\Lambda_3^+}, d^u$ satisfying (70).

Proof (i) For $\bar{\gamma} > 0$, we conduct the same process as in the proof of Theorem 30 (i), one can also obtain (159) and (160). If $rank(\widetilde{M}) < 2$, there exists $\beta \neq 0$ such that

$$\left(\bar{a}_{\bar{J}^0} - \bar{A}^1\bar{a}_{J^0}; \tilde{y}^1 - \bar{A}^2\bar{a}_{J^0}\right) = \beta\left(\bar{b}_{\bar{J}^0} - \bar{A}^1\bar{b}_{J^0}; \tilde{y}^2 - \bar{A}^2\bar{b}_{J^0}\right).$$

Hence, we have $d^C = -\beta d^\gamma$ and equation (159) reduces to

$$d^u = \bar{z}^1 d^\gamma \text{ and } d^\alpha_{\Lambda_3^+} = \tau d^\gamma, \tag{163}$$

where $\bar{z}^1$, τ are defined by

$$\begin{cases} \tau := -\bar{a}^1\beta + \bar{b}^1 + \bar{P}^1\bar{z}^1, \\ \bar{z}^1 := (\bar{P}_{J^0})^{-1}(\bar{a}_{J^0}\beta - \bar{b}_{J^0}). \end{cases}$$

Note that with $\bar{a}^1$, $\bar{b}^1$ and $\bar{P}^1$ in (158), we obtain that the above τ takes the form in (69). Let V be defined by

$$V := \begin{bmatrix} -\beta \\ 1 \\ \tau \\ k\mathbf{e}_{|\Lambda_3^c|} \end{bmatrix}. \tag{164}$$

It is easy to calculate that $Vd^\gamma = (d^C; d^\gamma; d^\alpha_{\Lambda_3^+}; d^\alpha_{\Lambda_3^c})$. Therefore, for any $d \in \mathcal{C}(\bar{v})$,

$$d^\top \nabla^2_{vv}\mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta}))d = d^\gamma V^\top D V d^\gamma,$$

where $D := D^1 - D^2$, D^1 and D^2 are given by (165) and (155).

Moreover, if $f(v)$ satisfies Assumption 11, D^1 reduces to the following form

$$D^1 = \begin{bmatrix} 0 & 0 & \mathbf{0}_{(1,|\Lambda_3|)} \\ 0 & \nabla^2_{\gamma\gamma} f(\bar{v}) & \nabla^2_{\gamma\alpha_{\Lambda_3}} f(\bar{v}) \\ \mathbf{0}_{(|\Lambda_3|,1)} & \nabla^2_{\gamma\alpha_{\Lambda_3}} f(\bar{v})^\top & \nabla^2_{\alpha_{\Lambda_3}\alpha_{\Lambda_3}} f(\bar{v}) \end{bmatrix}. \tag{165}$$

It is easy to check that $d^\gamma V^\top D V d^\gamma = d^\gamma (\dot{V}^f)^\top \dot{D}^f \dot{V}^f d^\gamma$, where $\dot{V}^f$, $\dot{D}^f$ are defined by (68). This gives (i).

(ii) If $rank(\widehat{M}) < 1+ \mid \Lambda_3^+ \mid$, by Proposition 44 (ii), for $d \in \mathcal{C}(\bar{v})$, we have

$$d^\top \nabla^2_{vv}\mathbb{L}(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{w})d = \bar{d}^\top V^\top D V \bar{d},$$

where $\bar{d}$ is given by (156), D is given by (157), and $d^C, d^\alpha_{\Lambda_3^+}, d^u$ satisfy (70). Moreover, if $f(v)$ satisfies Assumption 11, D in (157) reduces to the following form

$$D = \begin{bmatrix} 0 & \mathbf{0}_{(1,|\Lambda_3|)} \\ \mathbf{0}_{(|\Lambda_3|,1)} & \nabla^2_{\alpha_{\Lambda_3}\alpha_{\Lambda_3}} f(\bar{v}) \end{bmatrix}.$$

By calculation, it is easy to obtain that $\bar{d}^\top V^\top D V \bar{d} = \hat{d}^\top (V^f)^\top D^f V^f \hat{d}$ where $\hat{d}$, V^f and D^f are given in Assumption 13, $d^C, d^\alpha_{\Lambda_3^+}, d^u$ satisfy the linear system in (70). ■

Proof of Theorem 30

The results follow directly from Proposition 45. □

Appendix F. Proofs for Section 7

To show Theorem 32, we need to define the following index sets for problem (NLP-t_j). Recall that $r = 2n$ and define

$$\begin{array}{rcl} I_G(v) & := & \{i \in [r] \ : \ G_i(v) = 0\}, \\ I_H(v) & := & \{i \in [r] \ : \ H_i(v) = 0\}, \\ I_{GH}(v, t_j) & := & \{i \in [r] \ : \ G_i(v)H_i(v) - t_j = 0\}. \end{array}$$

Proof of Theorem 32

Obviously, $\bar{v}$ is feasible for (MPEC-BHO), and for all $j \in \mathbb{N}$ sufficiently large,

$$I_g(v^j) \subseteq I_g(\bar{v}), \ I_G(v^j) \subseteq I_{00}(\bar{v}) \bigcup I_{0+}(\bar{v}), \ I_H(v^j) \subseteq I_{00}(\bar{v}) \bigcup I_{+0}(\bar{v}). \tag{166}$$

Moreover, we have the following relationship between different index sets. For $i \in I_G(v^j)$, we have $G(v^j) = 0$, implying that $G_i(v^j)H_i(v^j) - t_j < 0$ and $\delta_i^j = 0$; i.e.,

$$I_G(v^j) \bigcap I_{GH}(v^j, t_j) = \emptyset \text{ and } \delta_i^j = 0 \text{ for all } i \in I_G(v^j). \tag{167}$$

Similarly, we have

$$I_H(v^j) \bigcap I_{GH}(v^j, t_j) = \emptyset \text{ and } \delta_i^j = 0 \text{ for all } i \in I_H(v^j). \tag{168}$$

Since all $(v^j, \lambda^j, \mu^j, \gamma^j, w^j, \delta^j)$ are KKT points of (NLP-t_j), we have

$$\begin{array}{rcl} 0 & = & \nabla f(v^j) + \sum_{i=1}^p \lambda_i^j g_i(v^j) + \sum_{i=1}^j \mu_i^j \nabla h_i(v^j) \\ & & \quad - \sum_{i=1}^r \eta_i^j \nabla G_i(v^j) - \sum_{i=1}^r w_i^j \nabla H_i(v^j) \\ & & \quad - \sum_{i=1}^r \delta_i^j \left(-H_i(v^j)\nabla G_i(v^j) - G_i(v^j)\nabla H_i(v^j)\right) \end{array} \tag{169}$$

with

$$\begin{array}{rcl} \lambda_i^j = 0, \ \forall i \notin I_g(v^j) & \text{and} & \lambda_i^j \geq 0, \ \forall i \in I_g(v^j), \\ \eta_i^j = 0, \ \forall i \notin I_G(v^j) & \text{and} & \eta_i^j \geq 0, \ \forall i \in I_G(v^j), \\ \zeta_i^j = 0, \ \forall i \notin I_H(v^j) & \text{and} & \zeta_i^j \geq 0, \ \forall i \in I_H(v^j), \end{array} \tag{170}$$

$$\delta_i^j = 0, \ \forall i \notin I_{GH}(v^j, t_j) \quad \text{and} \quad \delta_i^j \geq 0, \ \forall i \in I_{GH}(v^j, t_j). \tag{171}$$

Respectively define $\delta^{G,j}$ and $\delta^{H,j}$ as

$$\delta_i^{G,j} := \begin{cases} \delta_i^j H_i(v^j) & i \in I_{GH}(v^j, t_j), \\ 0 & \text{otherwise}, \end{cases} \quad \text{and } \delta_i^{H,j} := \begin{cases} \delta_i^j G_i(v^j) & i \in I_{GH}(v^j, t_j), \\ 0 & \text{otherwise}. \end{cases} \tag{172}$$

We can rewrite (169) as follows

$$\begin{array}{rcl} 0 & = & \nabla f(v^j) + \sum_{i=1}^p \lambda_i^j g_i(v^j) + \sum_{i=1}^k \mu_i^j \nabla h_i(v^j) - \sum_{i=1}^r \eta_i^j \nabla G_i(v^j) \\ & & - \sum_{i=1}^r \zeta_i^j \nabla H_i(v^j) + \sum_{i=1}^r \delta_i^{G,j} \nabla G_i(v^j) + \sum_{i=1}^r \delta_i^{H,j} \nabla H_i(v^j). \end{array} \tag{173}$$

Note that the multipliers $\delta^{G,j}$ and $\delta^{H,j}$ are nonnegative too. Our next step is to prove that the sequence $\{(\lambda^j, \mu^j, \eta^j, w^j, \delta^{G,j}, \delta^{H,j})\}$ is bounded. If we assume the contrary, we can find a subsequence $\mathbb{J}$ such that

$$\frac{(\lambda^j, \mu^j, \eta^j, \zeta^j, \delta^{G,j}, \delta^{H,j})}{\|(\lambda^j, \mu^j, \eta^j, \zeta^j, \delta^{G,j}, \delta^{H,j})\|} \to_{\mathbb{J}} (\hat{\lambda}, \hat{\mu}, \hat{\eta}, \hat{\zeta}, \hat{\delta}^G, \hat{\delta}^H) \neq 0.$$

Dividing by $\|(\lambda^j, \mu^j, \eta^j, \zeta^j, \delta^{G,j}, \delta^{H,j})\|$ and taking this limit in equation (173) yields (note that the functions g, h, G and H are continuously differentiable):

$$\begin{aligned} 0 \;\; = \;\; & \textstyle\sum_{i=1}^p \hat{\lambda}_i g_i(\bar{v}) + \sum_{i=1}^k \hat{\mu} \nabla h_i(\bar{v}) - \sum_{i=1}^r \hat{\eta}_i \nabla G_i(\bar{v}) - \sum_{i=1}^r \hat{w}_i \nabla H_i(\bar{v}) \\ & + \textstyle\sum_{i=1}^r \hat{\delta}_i^G \nabla G_i(\bar{v}) + \sum_{i=1}^r \hat{\delta}_i^H \nabla H_i(\bar{v}), \end{aligned}$$

i.e., the gradients

$$\begin{array}{l} \left\{\nabla g_i\,(\bar{v}) \;:\; i \in \operatorname{supp}(\hat{\lambda})\right\} \cup \left\{\nabla h_i\,(\bar{v}) \;:\; i \in \operatorname{supp}(\hat{\mu})\right\} \cup \\ \left\{\nabla G_i\,(\bar{v}) \;:\; i \in \operatorname{supp}(\hat{\eta}) \cup \operatorname{supp}(\hat{\delta}^G)\right\} \cup \left\{\nabla H_i\,(\bar{v}) \;:\; i \in \operatorname{supp}(\hat{\zeta}) \cup \operatorname{supp}(\hat{\delta}^H)\right\} \end{array} \tag{174}$$

are linearly dependent, which is a contradiction to MPEC-LICQ. Here we used the fact that $I_G(v^j) \bigcap I_{GH}(v^j, t_j) = \emptyset$ in (167) and $I_H(v^j) \bigcap I_{GH}(v^j, t_j) = \emptyset$ in (168), which implies[1]

$$\operatorname{supp}(\hat{\eta}) \cap \operatorname{supp}(\hat{\delta}^G) = \emptyset \;\text{ and }\; \operatorname{supp}(\hat{w}) \cap \operatorname{supp}(\hat{\delta}^H) = \emptyset. \tag{175}$$

We also used the observations that[2] $\operatorname{supp}(\bar{\lambda}) \subseteq I_g(\bar{v})$ and

$$\operatorname{supp}(\hat{\eta}) \cup \operatorname{supp}(\hat{\delta}^G) \subseteq I_{00}(\bar{v}) \cup I_{0+}(\bar{v}) \;\text{ and }\; \operatorname{supp}(\bar{\zeta}) \cup \operatorname{supp}(\bar{\delta}^H) \subseteq I_{00}(\bar{v}) \cup I_{+0}(\bar{v}).$$

Consequently, the sequence $\{(\lambda^j, \mu^j, \eta^j, \zeta^j, \delta^{G,j}, \delta^{H,j})\}$ is bounded. Therefore, it is convergent to some limit $(\bar{u}, \bar{\lambda}, \hat{\eta}, \hat{\zeta}, \hat{\delta}^G, \hat{\delta}^H)$. In fact, convergence holds on the whole sequence

1. For contradiction, assume that there exists $i_0 \in \operatorname{supp}(\hat{\eta}) \cap \operatorname{supp}(\hat{\delta}^G)$. For j sufficiently large and $j \in J$, it holds that $\eta^j_{i_0} > 0$ and $\delta^{G,j}_{i_0} > 0$. By (170), we have $i_0 \in I_G(v^j)$, i.e., $G_{i_0}(v^j) = 0$. On the other hand, by the definition of $\delta^{G,j}$ in (172), we have $\delta^j_{i_0} H_{i_0}(v^j) > 0$, implying that $\delta^j_{i_0} > 0$. Again by (171), we have $i_0 \in I_{GH}(v^j)$, meaning that $G_{i_0}(v^j) H_{i_0}(v^j) - t_j = 0$. However, this contradicts to $G_{i_0}(v^j) = 0$. Therefore, we have $\operatorname{supp}(\bar{\eta}) \cap \operatorname{supp}(\bar{\delta}^G) = \emptyset$. Similarly, we have $\operatorname{supp}(\hat{w}) \cap \operatorname{supp}(\hat{\delta}^H) = \emptyset$.
2. First, note that for $j \in \mathbb{J}$ sufficiently large, it holds that $\operatorname{supp}(\hat{\lambda}) \subseteq \operatorname{supp}(\lambda^j) \subseteq I_g(v^j) \subseteq I_g(\bar{v})$. Therefore, $\operatorname{supp}(\hat{\lambda}) \subseteq I_g(\bar{v})$ holds. Next we show that $\operatorname{supp}(\hat{\eta}) \subseteq I_{00}(\bar{v}) \cup I_{0+}(\bar{v})$. Note that for $i \in \operatorname{supp}(\hat{\eta})$, there exists $\epsilon_1 > 0$ such that for j sufficiently large and $j \in J$, $\eta^j_i > \epsilon_1$. Therefore, it holds that $G_i(v^j) = 0$, i.e., $i \in I_G(v^j)$. By (166), it holds that $\operatorname{supp}(\hat{\eta}) \subseteq I_{00}(\bar{v}) \cup I_{0+}(\bar{v})$. To show $\operatorname{supp}(\hat{\delta}^G) \subseteq I_{00}(\bar{v}) \cup I_{+0}(\bar{v})$, we will show $\operatorname{supp}(\hat{\delta}^G) \subseteq I_{+0}(\bar{v})$. Denote $\chi^k = (\lambda^j, \mu^j, \eta^j, w^j, \delta^{G,j}, \delta^{H,j})$. Since $\{\frac{\chi^j}{\|\chi^j\|}\}$ is bounded, we have $\{\frac{\eta^j}{\|\chi^j\|}\}$ and $\{\frac{\delta^{G,j}}{\|\chi^j\|}\}$ are bounded as well. For $i \in \operatorname{supp}(\delta^G)$, there exists $\epsilon_2 > 0$ such that for j sufficiently large and $j \in J$, $\frac{\delta_i^{G,j}}{\|\chi^j\|} > \epsilon_2$. By the definition of $\delta_i^{G,j}$ in (172), it holds that $\frac{\delta^j H_i(v^j)}{\|\chi^j\|} > \epsilon_2$. Now we claim that for such i, $H_i(\bar{v}) > 0$. Indeed, if $H_i(\bar{v}) = 0$, it holds that $H_i(v^k) \to_J 0$. By the boundedness of $\{\frac{\delta^{G,j}}{\|\chi^k\|}\}$, we obtain that $\frac{\delta^j H_i(v^j)}{\|\chi^j\|} \to_K 0$, which contradicts to $\frac{\delta_i^{G,j}}{\|\chi^j\|} > \epsilon_2$. Therefore, we obtain that for $i \in \operatorname{supp}(\delta^G)$, $i \in I_{0+}(\bar{v})$, i.e., $\operatorname{supp}(\bar{\delta}^G) \subseteq I_{0+}(\bar{v})$. Therefore, $\operatorname{supp}(\bar{\eta}) \cup \operatorname{supp}(\bar{\delta}^G) \subseteq I_{00}(\bar{v}) \cup I_{0+}(\bar{v})$. Similarly, we can show that $\operatorname{supp}(\bar{w}) \cup \operatorname{supp}(\bar{\delta}^H) \subseteq I_{00}(\bar{v}) \cup I_{+0}(\bar{v})$.

since the existence of two different accumulation points would again result in a contradiction to MPEC-LICQ. Due to (175), we respectively define the multipliers $\bar{\eta}$ and $\bar{\zeta}$ as follows

$$\bar{\eta}_i := \begin{cases} \hat{\eta}_i, & i \in \operatorname{supp}(\hat{\eta}), \\ -\hat{\delta}_i^G, & i \in \operatorname{supp}(\hat{\delta}^G), \\ 0, & \text{otherwise}, \end{cases} \qquad \bar{\zeta}_i := \begin{cases} \hat{\zeta}_i, & i \in \operatorname{supp}(\hat{\zeta}), \\ -\hat{\delta}_i^H, & i \in \operatorname{supp}(\hat{\delta}^H), \\ 0, & \text{otherwise}. \end{cases}$$

By taking the limit in (169), we obtain that

$$0 = \nabla f(\bar{v}) + \bar{\lambda}_i g_i(\bar{v}) + \sum_{i=1}^{p} \bar{\lambda}_i g_i(\bar{v}) + \sum_{i=1}^{k} \bar{\mu} \nabla h_i(\bar{v}) - \sum_{i=1}^{r} \bar{\eta}_i \nabla G_i(\bar{v}) - \sum_{i=1}^{r} \bar{\zeta}_i \nabla H_i(\bar{v}).$$

Here, $\bar{\lambda} \geq 0$ and

$$\begin{aligned} \operatorname{supp}(\bar{\lambda}) &\subseteq I_g(v^j) \subseteq I_g(\bar{v}), \\ \operatorname{supp}(\bar{\eta}) &= \operatorname{supp}(\hat{\eta}) \cup \operatorname{supp}(\hat{\delta}^G) \subseteq I_{00}(\bar{v}) \cup I_{0+}(\bar{v}), \\ \operatorname{supp}(\bar{\zeta}) &= \operatorname{supp}(\hat{\zeta}) \cup \operatorname{supp}(\hat{\delta}^H) \subseteq I_{00}(\bar{v}) \cup I_{+0}(\bar{v}). \end{aligned}$$

Consequently, we have

$$\bar{\eta}_i = 0 \text{ for all } i \in I_{+0}(\bar{v}) \text{ and } \bar{\zeta}_i = 0 \text{ for all } i \in I_{0+}(\bar{v}). \tag{176}$$

That is, $(\bar{v}, \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\zeta})$ is at least a weakly stationary point of (MPEC-BHO). To prove M-stationarity, assume that it holds that an $i \in I_{00}(\bar{v})$ with $\bar{\eta}_i < 0$ and $\bar{\zeta}_i \neq 0$ (the case $\bar{\eta}_i \neq 0$ and $\bar{w}_i < 0$ can be treated in a symmetric way). The condition $\bar{\eta}_i < 0$ implies $i \in \operatorname{supp}(\hat{\delta}^G) \subseteq I_{0+}(\bar{v})$ for all j sufficiently large[3]. We have $\bar{\zeta}_i = 0$ by (176), which is a contradiction. Hence, the result. □

References

K. P. Bennett, J. Hu, X. Y. Ji, G. Kunapuli, and J.-S. Pang. Model selection via bilevel optimization. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pages 1922–1929, 2006.

K. P. Bennett, G. Kunapuli, J. Hu, and J.-S. Pang. Bilevel optimization and machine learning. In *IEEE World Congress on Computational Intelligence*, pages 25–47, 2008.

C. M. Bishop and N. M. Nasrabadi. *Pattern recognition and machine learning.* Springer, New York, 2006.

K.-M. Chung, W.-C. Kao, C.-L. Sun, L.-L. Wang, and C.-J. Lin. Radius margin bounds for support vector machines with the RBF kernel. *Neural Computation*, 15(11):2643–2681, 2003.

3. To show $\operatorname{supp}(\hat{\delta}^G) \subseteq I_{0+}(\bar{v})$, for $i \in \operatorname{supp}(\delta^G)$, there exists $\epsilon > 0$ such that for j sufficiently large, $\delta_i^{G,j} = \delta^j H_i(v^j) > \epsilon$. Now we claim that for such i, $H_i(\bar{v}) > 0$. Indeed, if for contradiction, $H_i(\bar{v}) = 0$, it holds that $H_i(v^j) \to 0$. By the boundedness of $\delta^{G,j}\}$, we obtain that $\xrightarrow{\delta^j H_i(v^j)} 0$, which contradicts to $\delta_i^{G,j} > \epsilon$. Therefore, we obtain that for $i \in \operatorname{supp}(\delta^G)$, $i \in I_{0+}(\bar{v})$, i.e., $\operatorname{supp}(\hat{\delta}^G) \subseteq I_{0+}(\bar{v})$.

C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

S. Dempe and J. Dutta. Is bilevel programming a special case of a mathematical program with complementarity constraints? *Mathematical programming*, 131:37–48, 2012.

S. Dempe and A. B. Zemkoho. On the Karush–Kuhn–Tucker reformulation of the bilevel optimization problem. *Nonlinear Analysis: Theory, Methods & Applications*, 75(3):1202–1218, 2012.

S. Dempe and A. B. Zemkoho. The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. *Mathematical Programming*, 138:447–473, 2013.

B. Feizizadeh, M. S. Roodposhti, T. Blaschke, and J. Aryal. Comparing gis-based support vector machine kernel functions for landslide susceptibility mapping. *Arabian Journal of Geosciences*, 10(5):1–13, 2017.

M. L. Flegel. *Constraint qualifications and stationarity concepts for mathematical programs with equilibrium constraints.* PhD thesis, Universität Würzburg, 2005.

D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut. Comparison of linear, nonlinear, and feature selection methods for eeg signal classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):141–144, 2003.

L. Guo, G.-H. Lin, and J. J. Ye. Second-order optimality conditions for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 158:33–64, 2013.

M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. Support vector machines. *IEEE Intelligent Systems and their Applications*, 13(4):18–28, 1998.

T. Hoheisel, C. Kanzow, and A. Schwartz. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Mathematical Programming*, 137(1):257–288, 2013.

H. Hong, B. Pradhan, D. T. Bui, C. Xu, A. M. Youssef, and W. Chen. Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: a case study at Suichuan area (China). *Geomatics, Natural Hazards and Risk*, 8(2):544–569, 2017.

C. Kanzow and A. Schwartz. A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. *SIAM Journal on Optimization*, 23(2):770–798, 2013.

G. Kunapuli, K. Bennett, J. Hu, and J.-S. Pang. Bilevel model selection for support vector machines. In *CRM Proceedings and Lecture Notes*, volume 45, pages 129–158, 2008a.

G. Kunapuli, K. P. Bennett, J. Hu, and J.-S. Pang. Classification model selection via bilevel programming. *Optimization Methods & Software*, 23(4):475–489, 2008b.

Q.-N. Li, Z. Li, and A. B. Zemkoho. Bilevel hyperparameter optimization for support vector classification: theoretical analysis and a solution method. *Mathematical Methods of Operations Research*, 96(3):315–350, 2022a.

Z. Li, Y. Qian, and Q. Li. A unified framework and a case study for hyperparameter selection in machine learning via bilevel optimization. In *5th International Conference on Data Science and Information Technology*, 2022b.

G. Moore, C. Bergeron, and K. P. Bennett. Nonsmooth bilevel programming for hyperparameter selection. In *2009 IEEE International Conference on Data Mining Workshops*, pages 374–381, 2009.

M. A. Nanda, K. B. Seminar, D. Nandika, and A. Maddu. A comparison study of kernel functions in the support vector machine and its application for termite detection. *Information*, 9(1):5, 2018.

C. T. Nguyen, J. H. Alcantara, T. Okuno, A. Takeda, and J.-S. Chen. Unified smoothing approach for best hyperparameter selection problem using a bilevel optimization strategy. *arXiv:2110.12630v2*, 2023.

T. Okuno, A. Takeda, A. Kawana, and M. Watanabe. On ℓ_p-hyperparameter learning via bilevel nonsmooth optimization. *Journal of Machine Learning Research*, 22:245–1, 2021.

G. L. Prajapati and A. Patle. On performing classification using SVM with radial basis and polynomial kernel functions. In *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, pages 512–515, 2010.

S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 11(4):918–936, 2001.

K. Tbarki, S. B. Said, R. Ksantini, and Z. Lachiri. RBF kernel based SVM classification for landmine detection and discrimination. In *2016 International Image Processing, Applications and Systems (IPAS)*, pages 1–6, 2016.

Y. Wang and Q.-N. Li. A fast smoothing newton method for bilevel hyperparameter optimization for SVC with logistic loss. *arXiv:2308.07734v2*, 2023.

J. J. Ye, D. L. Zhu, and Q. J. Zhu. Exact penalization and necessary optimality conditions for generalized bilevel programming problems. *SIAM Journal on optimization*, 7(2):481–507, 1997.

B. Yekkehkhany, A. Safari, S. Homoyouni, and M. Hasanlou. A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry sar data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(2):281, 2014.