Noname manuscript No.
(will be inserted by the editor)

Bilevel optimization with a multiobjective problem in the lower level

Roberto Andreani · **Viviana A. Ramirez** ·
Sandra A. Santos · **Leonardo D. Secchin**

Received: date / Accepted: date

Abstract Bilevel problems model instances with a hierarchical structure. Aiming at an efficient solution of a constrained multiobjective problem according with some pre-defined criterion, we reformulate this optimization but non standard problem as a classic bilevel one. This reformulation intents to encompass all the objectives, so that the properly efficient solution set is recovered by means of a convenient weighted-sum scalarization approach. Inexact restoration strategies potentially take advantage of the structure of the problem under consideration, being employed as an alternative to the Karush-Kuhn-Tucker reformulation of the bilevel problem. Genuine multiobjective problems possess inequality constraints in their modelling, and these constraints generate theoretical and practical difficulties to our lower level problem. We handle these difficulties by means of a perturbation strategy, providing the convergence analysis, together with enlightening examples and illustrative numerical tests.

This work has been partially supported by the Brazilian funding agencies Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (grants 2013/05475-7 and 2013/07375-0) and Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (grants 303013/2013-3 and 302915/2016-8).

Roberto Andreani
Institute of Mathematics, University of Campinas, Rua Sergio Buarque de Holanda, 651, 13083-859, Campinas, São Paulo, Brazil.
E-mail: andreani@ime.unicamp.br

Viviana A. Ramirez
Academic Regional Center of Bariloche, National University of Comahue, Quintral 1250, San Carlos de Bariloche, Rio Negro, Argentina.
E-mail: viviramirezana@gmail.com

Sandra A. Santos (corresponding author)
Institute of Mathematics, University of Campinas, Rua Sergio Buarque de Holanda, 651, 13083-859, Campinas, São Paulo, Brazil.
E-mail: sandra@ime.unicamp.br

Leonardo D. Secchin
Department of Applied Mathematics, Federal University of Espírito Santo, São Mateus, Espírito Santo, Brazil.
E-mail: leonardo.secchin@ufes.br

Keywords Bilevel optimization · inexact restoration · multiobjective optimization · KKT reformulation · numerical experiments

Mathematics Subject Classification (2000) 90C29 · 65K05 · 49M37

1 Introduction

This work addresses a bilevel mathematical programming problem in which the lower level problem is a multiobjective one. An example that models this situation is as follows: given a certain decision, that one considers as acceptable, what is the closest and feasible decision that takes into consideration other objectives, also of interest? The general problem is formulated below

$$\min\ F(x) \text{ such that } x \text{ belongs to the set of efficient solutions of problem MOP} \quad (1)$$

where MOP is the following multiobjective problem

$$\begin{array}{ll} \min & \{f_1(x),\dots,f_p(x)\} \\ \text{s.t.} & h(x)=0 \\ & g(x)\le 0 \\ & x\in X. \end{array} \quad (2)$$

We assume that $X \subset \mathbb{R}^n$ is compact, $F:\mathbb{R}^n \to \mathbb{R}$, $f_i:\mathbb{R}^n \to \mathbb{R}$, $i=1,\dots,p$, $h:\mathbb{R}^n \to \mathbb{R}^m$ and $g:\mathbb{R}^n \to \mathbb{R}^q$ are twice continuously differentiable functions.

As (1)-(2) do not have the classic mathematical programming problem format, given tolerances $\varepsilon_i \in (0,1)$, $i=1,\dots,p$, we first adopt a weighted-sum scalarization reformulation for (2) as follows

$$\begin{array}{ll} \min\limits_{w,x} & \sum_{i=1}^p w_i f_i(x) \\ \text{s.t.} & \sum_{i=1}^p w_i = 1 \\ & w_i \ge \varepsilon_i, i=1,\dots,p \\ & h(x)=0 \\ & g(x)\le 0 \\ & x\in X. \end{array} \quad (3)$$

Such a reformulation depends on the positive parameters ε_i, which guarantee that all the objectives will play some role in the optimization process. In the literature, such solutions are related to the concept of *properly efficient* solutions, a nomenclature that dates back to Kuhn and Tucker [17] and Geoffrion [13] (see also [2,4]). These notion eliminates efficient points of anomalous type, in which a small gain for one function provides a large loss for another. Moreover, for any fixed w, global solutions of (3) are efficient points of (2). Reciprocally, in case all the functions f_i are convex and the feasible set of problem (2) is convex, then any properly efficient point of (2), in Geoffrion's sense, is a global solution of (3) for some $w > 0$ [13]. Hence, at least in the convex case, we have a good description of the efficient set of (2) by means of the weighted problem (3).

Therefore, (1)-(2) turns into

$$
\begin{aligned}
&\min_{w,x} \; F(x) \\
&\text{s.t.} \; \textstyle\sum_{i=1}^{p} w_i = 1 \\
&\qquad w_i \geq \varepsilon_i, i = 1, \ldots, p \\
&\qquad x \in \operatorname{argmin} \left\{ \begin{aligned} &\min \textstyle\sum_{i=1}^{p} w_i f_i(x) \\ &\text{s.t.} \;\; h(x) = 0 \\ &\qquad g(x) \leq 0 \\ &\qquad x \in X. \end{aligned} \right\}
\end{aligned} \tag{4}
$$

A well-known strategy for addressing problem (4) is to replace its lower level problem by the associate Karush-Kuhn-Tucker (KKT) conditions. This strategy, however, does not take into account the minimization structure inherent to the lower level problem. Alternatively, we have adopted an Inexact Restoration (IR) based strategy in which the optimization problem is addressed in two phases, first pursuing feasibility and then optimality, keeping a certain control on the feasibility already achieved. Consequently, our approach takes advantage of the intrinsic minimization structure of the problem, particularly at the feasibility phase, so that better solutions might be expected. Moreover, at the feasibility phase of the IR strategy, the user is free to adopt the method of his choice, as long as the restored iterates fulfill some mild assumptions [18,19].

Concerning the resolution of the original bilevel problem (1)–(2) by its weighted reformulation (4), when problem (3) is convex, it is well known that the KKT conditions are sufficient for its optimality. Thus, we can expect that an optimization algorithm applied to (3) always recover properly efficient solutions of problem (2) (at least if a constraint qualification holds). In the general case, only global solutions of the weighted problem (3) are guaranteed to be properly efficient solutions of (2) [13]. Thus, optimization algorithms may fail to get efficient solutions of (2) when they converge to a nonglobal stationary point of the weighted problem. In this sense, we reinforce that IR strategies can benefit from the fact that they treat the lower level weighted problem directly, without reformulating it, because in this case algorithms have more chances to reach global solutions (for instance, when they use second order information as sequential quadratic programming techniques do). In the worst case, the IR strategy is likely to obtain local solutions [21], and hence, properly local efficient solutions [14], whereas the KKT strategy does not distinguish minimizers from maximizers.

One should also notice that, at the optimality phase of the IR strategy, the KKT reformulation takes place as an auxiliary tool. The KKT reformulation of the lower level problem within the bilevel context, on the other hand, is a basic tool for addressing the problem, as it is the IR strategy. Along this perspective, it is worth mentioning a software environment for solving bilevel problems using inexact restoration, presented in [22].

The optimization over efficient sets has been previously analysed, but mainly considering linear optimization, linear bilevel problems and multicriteria optimization with polyhedral feasible sets. Benson [3] has studied the problem of optimizing a linear function over the set of efficient solutions for a vector optimization problem. The linear objective function measures the importance of, or discriminates, the efficient alternatives that are available. He established necessary and sufficient conditions for efficient and for arbitrary solutions of the underlying vector optimization problem to be optimal solutions for the problem upon analysis. By taking advantage of the equivalence between efficient points and optimal solutions of a linear programming problem, Fulop [12] pointed out that a pure linear bilevel problem might be addressed instead. In the general nonlinear case, a similar equivalence is not expected. A related and recent work is [6], in which the authors have addressed linearly equality constrained and bound constrained multiobjective instances in the lower level by means of a flexible IR strategy.

By addressing genuine multiobjective problems with inequality constraints, we have obtained undesired KKT points, particularly if the Lagrange multiplier of an inequality constraint of the lower level problem is strictly greater than zero, what forces the constraint to be active at the corresponding primal point. In such a case, the KKT point may be a maximizer of the problem (see Example 2). To overcome this inconvenient, we have devised an IR strategy in which the inequality constraints of the lower level problem are perturbed. The associated convergence analysis is provided, together with a set of illustrative numerical results.

The text is organized as follows. In Section 2 we provide an overview of the IR strategy employed in this work, with examples that highlight our contributions. The convergence analysis is provided in Section 3. Section 4 contains the numerical results, and in Section 5 we present final remarks and future work. The test problems are described in the Appendix A.

Notation. The symbols $\|\cdot\|$, $\|\cdot\|_\infty$ and $|\cdot|$ denote, respectively, the Euclidean, the supremum and an arbitrary norm. We also use the notation $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_i \geq 0\}$.

2 Inexact Restoration for bilevel programming

For the sake of simplicity, in this section we deal with the problem

$$\begin{aligned} \text{P: } & \min_x \; F(x) \\ & \text{s.t. } \; x \in \mathscr{X}^* \end{aligned}$$

where $F : \mathbb{R}^n \to \mathbb{R}$ and $\mathscr{X}^*$ is the set of efficient solutions of the lower level multiobjective problem

$$\begin{aligned} \min \; & \{f_1(x), \ldots, f_p(x)\} \\ \text{s.t. } \; & h(x) = 0 \\ & 0 \leq x \in X \end{aligned}$$

with $f_1, \ldots, f_p : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$. We assume that all functions are twice continuously differentiable.

Given tolerances $\varepsilon_i \in (0,1)$, $i = 1,\dots,p$, we can state P as

$$\begin{aligned} \min_{w,x}\ & F(x) \\ \text{s.t.}\ & w \in W \\ & x \in \operatorname{argmin} \left\{ \begin{array}{l} \min_x \sum_{i=1}^p w_i f_i(x) \\ \text{s.t. } h(x) = 0 \\ \qquad 0 \le x \in X \end{array} \right\} \end{aligned} \tag{5}$$

where

$$W = \left\{ w \in \mathbb{R}^p \mid \sum_{i=1}^p w_i = 1, \quad \varepsilon_i \le w_i \le 1 - \varepsilon_i, \quad i = 1,\dots,p \right\}.$$

Note that W is a convex and compact set. Here, we have $X \equiv \mathbb{R}^n$, so that $x \in X$ trivially holds. Nevertheless, such a pertinence is kept because, in the general case, a bounding compact set X is required. We emphasize that the feasible set of the lower level problem of (5) was described with nonlinear equalities and simple nonegative variables for simplifying the presentation. Compared with the feasible set of problem (2), one might assume that slack variables were added to the nonlinear inequality constraints, and all variables are nonnegative, so that the technicalities of dealing with some free variables and the associated cumbersome notation are avoided.

We define

$$C(w,x,\mu,\gamma) = \begin{bmatrix} \sum_{i=1}^p w_i \nabla_x f_i(x) + \nabla_x h(x)\mu - \gamma \\ h(x) \\ \gamma_1 x_1 \\ \vdots \\ \gamma_n x_n \end{bmatrix} \in \mathbb{R}^{2n+m} \tag{6}$$

as in [1], where $(\mu,\gamma) \in \Delta \subset \mathbb{R}^m \times \mathbb{R}^n_+$. The KKT conditions for the lower level problem, parametrized in w, are

$$C(w,x,\mu,\gamma) = 0, \quad 0 \le x \in X, \quad \gamma \ge 0 \tag{7}$$

(thus Δ plays the role of the set of multipliers for this parametrized problem, which in principle, we can take as $\mathbb{R}^m \times \mathbb{R}^n_+$). We also define

$$L(w,x,\mu,\gamma,\lambda) = F(x) + C(w,x,\mu,\gamma)^T \lambda$$

for $\lambda \in \mathbb{R}^{2n+m}$. We write shortly $s = (w,x,\mu,\gamma) \in W \times X \times \Delta$, and the Jacobian of C is given by

$$C'(s) = \begin{bmatrix} \nabla_x f_1 & \cdots & \nabla_x f_p & \sum_{i=1}^p w_i \nabla^2_{xx} f_i + \sum_{i=1}^m \mu_i \nabla^2_{xx} h_i & \nabla h & -I_n \\ 0 & \cdots & 0 & \nabla h^T & 0 & 0 \\ 0 & \cdots & 0 & \operatorname{diag}(\gamma) & 0 & \operatorname{diag}(x) \end{bmatrix}. \tag{8}$$

In the following, we adapt the inexact restoration method presented in [1] to our purposes. The parameters $\eta > 0$, $M > 0$, $\theta_{-1} \in (0,1)$, $\delta_{\min} > 0$, $\tau_1 > 0$, $\tau_2 > 0$ are given, as well as initial approximations $s^0 \in W \times X \times \Delta$, $\lambda^0 \in \mathbb{R}^{2n+m}$, and a sequence $\{\omega^k\}$ such that $\sum_{k=0}^{\infty} \omega^k < \infty$.

The steps for obtaining $s^{k+1} = (w^{k+1}, x^{k+1}, \mu^{k+1}, \gamma^{k+1})$ and λ^{k+1} are as follows.

Algorithm 1 Inexact Restoration

1. Define $\theta_k^{\min} = \min\{1, \theta_{k-1}, \ldots, \theta_{-1}\}$, $\theta_k^{\text{large}} = \min\{1, \theta_k^{\min} + \omega^k\}$ and $\theta_{k,-1} = \theta_k^{\text{large}}$.
2. **(Restoration phase)** Find an approximate minimizer $\bar{x}$ and multipliers $(\overline{\mu}, \overline{\gamma}) \in \Delta$ for the problem
$$\begin{array}{ll} \min\limits_{x} & \sum_{i=1}^{p} w_i^k f_i(x) \\ \text{s.t.} & h(x) = 0 \\ & 0 \leq x \in X, \end{array}$$
and define $z^k = (w^k, \bar{x}, \overline{\mu}, \overline{\gamma})$.
3. **(Tangent direction)** Compute
$$d_{\tan}^k = P_k[z^k - \eta \nabla_s L(z^k, \lambda^k)] - z^k$$
where $P_k[\cdot]$ is the orthogonal projection on
$$\pi_k = \{s \in W \times X \times \Delta \mid C'(z^k)(s - z^k) = 0\}.$$
$P_k[z^k - \eta \nabla_s L(z^k, \lambda^k)]$ is a solution of the problem
$$\begin{array}{ll} \min\limits_{y \in W \times X \times \Delta} & \dfrac{1}{2} \left\| y - [z^k - \eta \nabla_s L(z^k, \lambda^k)] \right\|^2 \\ \text{s.t.} & C'(z^k)(y - z^k) = 0 \end{array}$$
(this solution is unique if X and Δ are convex). If $z^k = s^k$ and $d_{\tan}^k = 0$ stop and return x^k as a solution of P. Otherwise, set $i \leftarrow 0$, choose $\delta_{k,0} \geq \delta_{\min}$.
4. **(Minimization phase)** If $d_{\tan}^k = 0$ take $v^{k,i} = z^k$.
Otherwise, take $t_{\text{break}}^{k,i} = \min\{1, \delta_{k,i}/\|d_{\tan}^k\|\}$ and find $v^{k,i} \in \pi_k$ such that, for some $0 < t \leq t_{\text{break}}^{k,i}$, we have
$$L(v^{k,i}, \lambda^k) \leq \max\{L(z^k + t d_{\tan}^k, \lambda^k), L(z^k, \lambda^k) - \tau_1 \delta_{k,i}, L(z^k, \lambda^k) - \tau_2\}$$
with $\gamma \geq 0$ (γ of $v^{k,i}$) and $\|v^{k,i} - z^k\|_\infty \leq \delta_{k,i}$.
5. If $d_{\tan}^k = 0$ define $\lambda^{k,i} = \lambda^k$. Otherwise, take $\lambda^{k,i} \in \mathbb{R}^{2n+m}$ such that $|\lambda^{k,i}| \leq M$.
6. For all $\theta \in [0,1]$ define
$$\begin{aligned} Pred_{k,i}(\theta) = \theta \left[L(s^k, \lambda^k) - L(v^{k,i}, \lambda^k) - C(z^k)^T (\lambda^{k,i} - \lambda^k) \right] + \\ (1 - \theta) \left[\|C(s^k)\| - \|C(z^k)\| \right]. \end{aligned}$$

Take $\theta_{k,i}$ as the maximum $\theta \in [0, \theta_{k,i-1}]$ that satisfies

$$Pred_{k,i}(\theta) \geq \frac{1}{2}\left[\|C(s^k)\| - \|C(z^k)\|\right],$$

and define $Pred_{k,i} = Pred_{k,i}(\theta_{k,i})$.

7. Calculate

$$Ared_{k,i} = \theta_{k,i}\left[L(s^k,\lambda^k) - L(v^{k,i},\lambda^{k,i})\right] + (1-\theta_{k,i})\left[\|C(s^k)\| - \|C(v^{k,i})\|\right].$$

If

$$Ared_{k,i} \geq 0.1 Pred_{k,i}$$

set

$$s^{k+1} = v^{k,i}, \quad \lambda^{k+1} = \lambda^{k,i}, \quad \theta_k = \theta_{k,i}, \quad \delta_k = \delta_{k,i},$$
$$Ared_k = Ared_{k,i}, \quad Pred_k = Pred_{k,i}$$

and finish the current k-th iteration.
Otherwise, choose $\delta_{k,i+1} \in [0.1\delta_{k,i}, 0.9\delta_{k,i}]$, set $i \leftarrow i+1$ and go to step 4.

An interesting issue in multiobjective optimization is the effect that constraints may have on the efficient set of a problem. Contrary to what we can expect, the efficient set of a constrained multiobjective problem

$$\min\{f_1(x),\ldots,f_p(x)\} \quad \text{s.t.} \quad x \in X$$

is not necessarily composed only by the efficient points of the unconstrained problem

$$\min\{f_1(x),\ldots,f_p(x)\}$$

that lies in the feasible region X. There may be additional efficient points at the border of X. In fact, a border point such that all non-ascending paths for all objective f_i are infeasible is efficient (see Figure 1). The next example ilustrates an instance for which the optimal solution of the bilevel problem (1) is located at the border of the lower level feasible set. It is based on a problem of [23], where the authors have not addressed this issue, missing efficient points from their analysis.

Example 1 Let us consider the nonconvex constrained multiobjective problem F3 of [23]

$$\begin{array}{ll} \min & \{(\tilde{x}_1-2)^2 + (\tilde{x}_2-1)^2 + 2,\, 9\tilde{x}_1 - (\tilde{x}_2-1)^2\} \\ \text{s.t.} & \tilde{x}_1^2 + \tilde{x}_2^2 \leq 225 \\ & \tilde{x}_1 - 3\tilde{x}_2 \leq -10. \end{array}$$

The search space adopted by the authors is the box $[-20,20]^2$. In order to turn all variables into nonnegative, we make the transformation $x_i = \tilde{x}_i + 20$, $i = 1,2$, and define the problem

$$\begin{array}{ll} \min & F(x_1,x_2) = (x_1-20)^2 + (x_2-20)^2 \\ \text{s.t.} & x \in \mathscr{X}^*, \end{array} \tag{9}$$

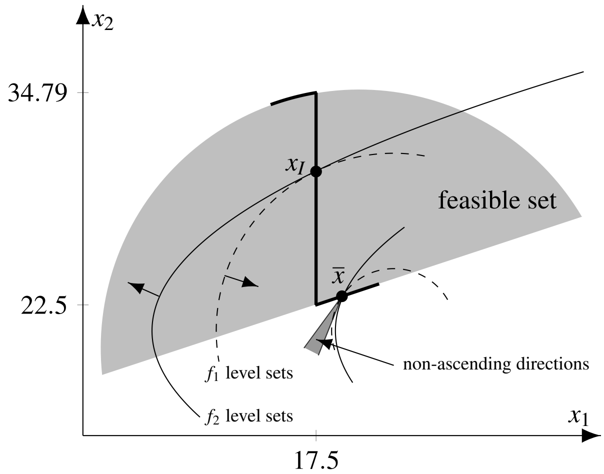

Fig. 1 Geometry of Example 1. The arrows at the level sets of the objective functions indicate the corresponding decreasing directions. The vertical continuous line is the set of efficient points in which the level sets of f_1 and f_2 have the same slope. The continuous lines at the border of the feasible set represent the efficient points such that all non-ascending directions for both objectives are infeasible. In particular, x_I and $\bar{x}$ are efficient points of each one of these two types, respectively.

where $\mathscr{X}^*$ is the set of efficient solutions of

$$\begin{array}{ll} \min & \{(x_1-22)^2+(x_2-21)^2+2,\, 9x_1-(x_2-21)^2-180\} \\ \text{s.t.} & (x_1-20)^2+(x_2-20)^2 \leq 225 \\ & x_1-3x_2 \leq -50 \\ & 0 \leq x_1, x_2 \leq 40. \end{array} \tag{10}$$

In the interior of the feasible region, we analyse the slopes of the two objective functions of (10) to conclude that the interval $x_1 = 17.5$, $22.5 \leq x_2 \leq 34.79$ is efficient. This interval corresponds to the intersection of the efficient set of the unconstrained problem with the feasible region of (10), and the authors of [23] claim that it is the entire efficient set of (10). However, there are other efficient points at the border of feasible set. Geometrically, these points are those in which there is no feasible non-ascending direction that decreases at least one objective function of (10) (see Figure 1). With a simple geometric argument, it is easy to see that $\bar{x} = (19,23)$ is the optimal solution of (9), for which $F(\bar{x}) = 10$ and the associated vector of weights $\overline{w} = (23/37, 14/37)$.

It is worth mentioning that, in our computational tests, the Inexact Restoration strategy was able to reach the optimal solution $\bar{x}$ located at the border of the feasible set, whereas the KKT reformulation does not converge (see the results of instance P1.1 in Section 4).

2.1 A perturbed inexact restoration strategy

We start this section with an example showing that Algorithm 1 can converge to an undesirable feasible point of the multiobjective problem. The key to the failure lies in the fact that, for certain weights, the restoration phase can return a (first order) solution that is undesirable for the upper level problem. Despite the fact that the weighted-sum scalarized version of the lower level problem is a strictly convex quadratic problem, the convexity does not avoid this potential undesired situation.

Example 2 Consider the bidimensional problem

$$\begin{aligned} &\min\ F(x_1,x_2) = x_1^2 + x_2^2 \\ &\text{s.t.}\ \ (x_1,x_2) \in \mathscr{X}^*, \end{aligned} \tag{11}$$

where $\mathscr{X}^*$ is the set of efficient solutions of the lower level multiobjective problem

$$\begin{aligned} &\min\ \{f_1(x),\dots,f_p(x)\} \\ &\text{s.t.}\ \ (2/5)x_1 + x_2 \leq 2 \\ &\qquad 5x_1 - x_2 \leq 10 \\ &\qquad x_1,x_2 \geq 0, \end{aligned} \tag{12}$$

and, for $1 \leq i \leq p$,

$$\begin{aligned} f_i(x_1,x_2) = (x_1-a_i)^2 + (x_2-b_i)^2&, \\ a_i = 1 + 2\cos\theta_i&, \\ b_i = 1 + 2\sin\theta_i&, \\ \theta_i = \frac{2\pi}{p}(i-1)&. \end{aligned}$$

We can take X as a bounding box that strictly contains the compact feasible region of (12), and thus we will omit it. We consider the case $p=3$, for which $a=(3,0,0)^T$ and $b=(1,1+\sqrt{3},1-\sqrt{3})^T$. It is possible to show that the efficient solutions of the unconstrained multiobjective problem such that all objective functions have the form $\|x-c\|^2$ is the convex hull of the centers c's. Then, the set of efficient solutions $\mathscr{X}^*$ of (12) (i.e., the feasible region of (11)) is the intersection of the convex hull of (a_i,b_i), $i=1,2,3$, with the region delimited by the constraints of the lower level problem (12) (see Figure 2). Hence, the origin is the unique global optimal solution of (11).

Adding slack variables, we write the weighted-sum scalarized version of problem (11) as

$$\begin{aligned} &\min\ F(x_1,x_2) = x_1^2 + x_2^2 \\ &\text{s.t.}\ \ \textstyle\sum_{i=1}^3 w_i = 1 \\ &\qquad w_i \geq \varepsilon_i = \varepsilon, \quad i=1,2,3 \\ &\qquad x \in \operatorname{argmin}\left\{ \begin{aligned} &\min\ \textstyle\sum_{i=1}^3 w_i f_i(x_1,x_2) \\ &\text{s.t.}\ \ (2/5)x_1 + x_2 + x_3 = 2 \\ &\qquad 5x_1 - x_2 + x_4 = 10 \\ &\qquad x_i \geq 0, i=1,\dots,4. \end{aligned} \right\}. \end{aligned} \tag{13}$$

We have that $\bar{x} = (20/29, 50/29, 0, 240/29)^T$ is a KKT point of the lower level problem for certain feasible weights w for problem (13). In fact, the KKT conditions of the lower level problem are

$$\sum_{i=1}^{3} 2w_i \begin{bmatrix} x_1 - a_i \\ x_2 - b_i \\ 0 \\ 0 \end{bmatrix} + \mu_1 \begin{bmatrix} 2/5 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \mu_2 \begin{bmatrix} 5 \\ -1 \\ 0 \\ 1 \end{bmatrix} - \gamma = 0, \quad (14)$$

$$(2/5)x_1 + x_2 + x_3 = 2, \quad (15)$$

$$5x_1 - x_2 + x_4 = 10, \quad (16)$$

$$\gamma_i x_i = 0,\ i = 1, \ldots, 4, \quad (17)$$

$$x \geq 0,\ \mu \in \mathbb{R}^2,\ \gamma \in \mathbb{R}^4_+. \quad (18)$$

At $\bar{x}$, (15)-(16) hold, $\gamma_1 = \gamma_2 = \gamma_4 = 0$ by (17), and $\mu_1 = \gamma_3 \geq 0$, $\mu_2 = \gamma_4 = 0$ by (14) and (18). Furthermore, if $w_1 \neq 20/87$, the expression (14) implies $\gamma_3 = \mu_1 > 0$, since $\mu_2 = \gamma_1 = 0$ and

$$\sum_{i=1}^{3} w_i(\bar{x}_1 - a_i) = \bar{x}_1 - w_1 a_1 \neq 0.$$

Thus, by expression (14), $\bar{x}$ is a KKT point of the lower level problem for all feasible w such that

$$\begin{cases} 15w_1 - 2\sqrt{3}w_2 + 2\sqrt{3}w_3 = 2 \\ w_1 \neq 20/87. \end{cases} \quad (19)$$

Evidently, the system (19) has feasible solutions for all $\varepsilon < 20/87 \approx 0.23$.

Now, take $\overline{w} \in \mathbb{R}^3$, feasible for (13) and satisfying (19). We will show that the inexact restoration method can converge to $\bar{x}$, an undesirable solution for problem (11) (see Figure 2). In fact, since $\bar{x}$ is a KKT point of the lower level problem with $w = \overline{w}$, there are multipliers $(\overline{\mu}, \overline{\gamma})$ satisfying (14)-(18). As we have already mentioned, $\overline{\gamma}_3 > 0$. We denote $\bar{z} = (\overline{w}, \bar{x}, \overline{\mu}, \overline{\gamma})$. Starting Algorithm 1 with $s^0 = \bar{z}$ and $\lambda^0 = 0$, we can assume that step 2 returns $z^0 = \bar{z}$. Step 3 gives

$$\overline{\gamma}_3(x_3 - \bar{x}_3) + \bar{x}_3(\gamma_3 - \overline{\gamma}_3) = 0 \Rightarrow \overline{\gamma}_3 x_3 = 0,$$

that implies $x_3 = 0$, since $\overline{\gamma}_3 > 0$. Thus, the x-space of π_0 is contained in the line $(2/5)x_1 + x_2 = 2$ (indeed, it is exactly this line, see Figure 2). It is easy to verify that $\nabla_x F(\bar{x})$ is orthogonal to the line $(2/5)x_1 + x_2 = 2$, and hence

$$d^0_{\tan} = P_0[\bar{z} - \eta \nabla_s L(\bar{z}, \lambda^0)] - \bar{z} = P_0[\bar{z} - \eta \nabla_s F(\bar{x})] - \bar{z} = 0.$$

Therefore, Algorithm 1 stops declaring $\bar{x}$ as a solution of problem (11).

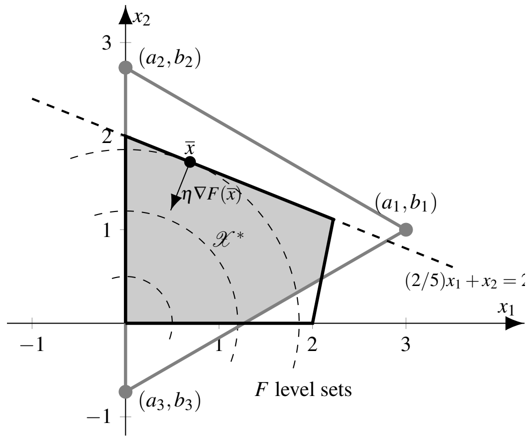

Fig. 2 Geometry of Example 2. The feasible set $\mathscr{X}^*$ of (11) is the intersection of the convex hull of (a_i,b_i), $i=1,2,3$, with the region delimited by the constraints of the lower level problem. The lower level KKT point $\bar{x}$ is an undesired solution for problem (11).

In Example 2, the x-space of π_0 at $\bar{x}$ is a border line of the lower level feasible region which is orthogonal to $\nabla_x F(\bar{x})$, resulting in a premature convergence. If we "balance" complementarity by requiring that $\gamma_i x_i = \xi > 0$, the current point x will be in the interior of the lower level feasible region. In particular, we avoid that Algorithm 1 converges to $\bar{x}$ in the example. Evidently, during the algorithm, ξ must vanish in order to recover the original KKT system. We call this strategy *perturbed inexact restoration*, clearly inspired by interior point methods.

More specifically, at iteration $k \geq 1$ we make, for all i, $\gamma_i^k x_i^k = \xi^k$ where

$$\xi^k = \min\left\{\tilde{\xi}^k\,,\,\xi_{\max}\,,\,\xi_{\text{dec}}\cdot\xi^{k-1}\right\}, \tag{20}$$

$$\tilde{\xi}^k = \frac{1}{k^2}\left(\frac{\gamma^{k-1}\cdot x^{k-1}}{n}\right), \tag{21}$$

being $\xi_{\max} > 0$ the maximum allowed perturbation and $\xi_{\text{dec}} \in (0,1)$ a parameter to ensure that the perturbation decreases between consecutive iterations. We also define $\xi^0 = \xi_{\max}$. The expression (21), motivated by convergence results in [8], has shown better numerical performance for our tests than the update $\tilde{\xi}^k = \gamma^{k-1}\cdot x^{k-1}/n$, usually employed in interior-point methods. The restoration phase of Algorithm 1 turns into the resolution of

$$C_\xi(w,x,\mu,\gamma) = \begin{bmatrix} \sum_{i=1}^p w_i \nabla_x f_i(x) + \nabla_x h(x)\mu - \gamma \\ h(x) \\ \gamma_1 x_1 - \xi \\ \vdots \\ \gamma_n x_n - \xi \end{bmatrix} = 0, \quad 0 \leq x \in X, \quad \gamma \geq 0 \tag{22}$$

with $\xi = \xi^k$ and $w = w^k$ (actually, $(\mu, \lambda) \in \Delta$). We note that $C'_{\xi^k} = C'$ and hence the optimization problem in step 3 of Algorithm 1 remains the same, whereas steps 4 to 7 are adapted in an obvious way.

A solution of (22) can be recovered if the optimal functional value of the problem

$$\begin{array}{ll} \min\limits_{x,\mu,\gamma} & \left\|\sum_{i=1}^p w_i \nabla_x f_i(x) + \nabla_x h(x)\mu - \gamma\right\|^2 \\ \text{s.t.} & h(x) = 0 \\ & x_i\gamma_i = \xi, \quad i = 1, \ldots, n \\ & 0 \le x \in X, \quad (\mu, \gamma) \in \Delta \end{array} \tag{23}$$

is zero. If we can not obtain such a solution, we proceed with a stationary point of this problem.

The perturbed strategy avoids convergence to an undesirable point in the example. However, when the perturbation ξ is positive we do not have a feasible point (or at least, a lower level KKT point). The next example shows that if we simply adopt the same stopping criteria "$z^k = s^k$ and $d^k_{\tan} = 0$" of Algorithm 1 to our perturbed inexact restoration strategy, it may stop with a positive perturbation.

Example 3 Given a fixed $\alpha > 0$, let us consider the problem

$$\begin{array}{ll} \min & F(x) = \frac{1}{2}x^2 - \sqrt{\alpha}x \\ \text{s.t.} & w_1 + w_2 = 1 \\ & w_1, w_2 \ge \varepsilon \\ & x \in \operatorname{argmin} \left\{ \begin{array}{l} \min \ w_1\left(\frac{1}{2}x^2\right) + w_2\left(\frac{1}{2}x^2\right) \\ \text{s.t.} \ \ x \ge 0. \end{array} \right\} \end{array}$$

The perturbed lower level KKT system $C_\xi(w_1, w_2, x, \gamma)$ is

$$(w_1 + w_2)x - \gamma = 0, \quad x, \gamma \ge 0, \quad x\gamma = \xi,$$

which yields $x = \gamma = \sqrt{\xi}$. As $C_\xi(w_1, w_2, \sqrt{\xi}, \sqrt{\xi}) = 0$ we can assume that this point is the one obtained by the restoration phase. Thus, if we initialize the method with $\xi^0 = \alpha$, $x^0 = \gamma^0 = \sqrt{\alpha}$, $w^0_1 = w^0_2 = 1/2$ (i.e., $s^0 = \left(1/2, 1/2, \sqrt{\alpha}, \sqrt{\alpha}\right)$) and $\lambda^0 = 0$ then the problem of step 3 in the first outer iteration takes the form

$$\begin{array}{ll} \min\limits_{w_1,w_2,x,\gamma} & \frac{1}{2}\left\|(w_1, w_2, x, \gamma) - \left(\frac{1}{2}, \frac{1}{2}, \sqrt{\alpha}, \sqrt{\alpha}\right)\right\|_2^2 \\ \text{s.t.} & \begin{bmatrix} \sqrt{\alpha} & \sqrt{\alpha} & 1 & -1 \\ 0 & 0 & \sqrt{\alpha} & \sqrt{\alpha} \end{bmatrix} \begin{bmatrix} w_1 - 1/2 \\ w_2 - 1/2 \\ x - \sqrt{\alpha} \\ \gamma - \sqrt{\alpha} \end{bmatrix} = 0 \\ & w_1 + w_2 = 1 \\ & \varepsilon_i \le w_i \le 1 - \varepsilon_i, \quad i = 1, 2. \end{array}$$

The unique solution is the point $z^0 = \left(1/2, 1/2, \sqrt{\alpha}, \sqrt{\alpha}\right)$, exactly the initial point s^0. Therefore, $z^0 = s^0$ and $d^0_{\tan} = 0$, which implies a premature convergence to the nonfeasible point s^0 if no new conditions are imposed.

In order to guarantee a correct convergence of our perturbed inexact restoration algorithm, we reduce the perturbation ξ whenever there is lack of progress ($z^k = s^k$ and $d^k_{\tan} = 0$) with a positive ξ. In this case, we go back to the restoration phase with a smaller ξ. This ensures that the original KKT system will be recovered as desired. We summarize the new step 3 below.

3' **(Tangent direction)** Compute

$$d^k_{\tan} = P_k[z^k - \eta \nabla_s L(z^k, \lambda^k)] - z^k$$

where $P_k[\cdot]$ is the orthogonal projection on

$$\pi_k = \{s \in W \times X \times \Delta \mid C'_{\xi^k}(z^k)(s - z^k) = 0\}.$$

If $z^k \neq s^k$ or $d^k_{\tan} \neq 0$, do $i \leftarrow 0$, choose $\delta_{k,0} \geq \delta_{\min}$.
If $z^k = s^k$, $d^k_{\tan} = 0$ and $\xi^k > 0$, do

$$\xi^{k+1} = \xi_{\text{dec}} \cdot \xi^k, \quad s^{k+1} = s^k, \quad \lambda^{k+1} = \lambda^k, \quad \theta_k = \theta_{k-1},$$

finish the current iteration k and go to step 2.
If $z^k = s^k$, $d^k_{\tan} = 0$ and $\xi^k = 0$, stop and return x^k as a solution of problem P.

There might be other ways to overcome an undesirable premature convergence, but certainly this choice is valid. Also, we did not observe this occurrence in our numerical tests, so possibly these cases are rare. Furthermore, a few modifications in the original inexact restoration method are made, and the already established convergence results are easily adapted to our perturbed version, as we discuss next.

3 Convergence results for the IR strategies

In this section, we stated convergence results for the plain and the perturbed inexact restoration strategies presented in the previous section. In [1], convergence results for an inexact restoration method applied to bilevel programming were established. The authors considered the method with Lagrangian projection, i.e., with tangent direction

$$d^k_{\tan} = P_k[z^k - \eta \nabla_s L(z^k, \lambda^k)] - z^k.$$

However, in standard nonlinear optimization, the convergence to points that satisfy the so-called *Approximate Gradient Projection* (AGP) [20], a sequential optimality condition, has been proved only with the tangent direction that uses the objective function [19], i.e. directions defined by

$$d^k_{\tan,\text{F}} = P_k[z^k - \eta \nabla_s F(z^k)] - z^k.$$

In fact, a point x^* that conforms to the AGP necessary optimality condition drives to zero the tangent direction defined by means of the objective function. In this sense, the AGP condition is closely linked to inexact restoration that uses $d^k_{\tan,\text{F}}$. The use of the Lagrangian in the tangent directions was proposed in [18] and aims to avoid Marato's effect.

We have

$$d^k_{\text{tan,F}} = P_k[z^k - \eta \nabla_s L(z^k, 0)] - z^k,$$

and the inexact restoration method with $d^k_{\text{tan,F}}$ is exactly the Algorithm 1 setting $\lambda^k \equiv 0$. Of course, in this case Step 5 can be eliminated. In the next sections, we will analyse the following variants of the inexact restoration strategy:

- IR_L: Inexact restoration with d^k_{tan} (Algorithm 1);
- IR^ξ_L: Perturbed inexact restoration with d^k_{tan};
- IR_F: Inexact restoration with $d^k_{\text{tan,F}}$. This variant corresponds to Algorithm 1 setting $\lambda^k = 0$ for all k;
- IR^ξ_F: Perturbed inexact restoration with $d^k_{\text{tan,F}}$.

3.1 General assumptions

We make assumptions on the weighted bilevel problem (5) in order to guarantee convergence for all inexact restoration variants, following [1] and previous results for standard nonlinear programming [18, 19]. For non-perturbed variants, assumptions are easily adapted making $\xi = 0$. The hypothesis are:

A1. The sets X and Δ are compact and convex.

A2. There exists $L_1 > 0$ such that, for all $(w,x), (\overline{w},\overline{x}) \in W \times X$, $(\mu,\gamma), (\overline{\mu},\overline{\gamma}) \in \Delta$ and $\xi \in [0, \xi_{\max}]$,

$$|C'_\xi(w,x,\mu,\gamma) - C'_\xi(\overline{w},\overline{x},\overline{\mu},\overline{\gamma})| \le L_1 |(w,x,\mu,\gamma) - (\overline{w},\overline{x},\overline{\mu},\overline{\gamma})|.$$

A3. There exists $L_2 > 0$ such that, for all $x, \overline{x} \in X$,

$$|\nabla F(x) - \nabla F(\overline{x})| \le L_2 |x - \overline{x}|.$$

A4. There exists $r \in [0,1)$, independently of k, such that the point $z^k = (w^k, \overline{x}, \overline{\mu}, \overline{\gamma})$ obtained at the restoration phase satisfies

$$|C_{\xi^k}(z^k)| \le r |C_{\xi^k}(s^k)|.$$

where $s^k = (w^k, x^k, \mu^k, \gamma^k)$ is the current iterate. Moreover, if $C_{\xi^k}(s^k) = 0$ then $z^k = s^k$.

The first three assumptions (A1 to A3) correspond exactly to those made in [18]. As $C'_\xi = C'$ for all $\xi \ge 0$ (see expressions (6) and (8)), assumption A2 is independent of the perturbation. We observe that a way to guarantee A2 is imposing, together with A1, the Lipschitz continuity of the functions $\nabla^2 f_i$ and $\nabla^2 h_i$. This is the assumption adopted in [1].

3.2 Well-definiteness

Theorem 1 *Under assumptions A1, A2, A3 and A4, all inexact restoration algorithms are well defined.*

Proof In the $\mathrm{IR}_{\mathrm{L}}^{\xi}$ variant, if $C_{\xi^k}(s^k) \neq 0$ or ($C_{\xi^k}(s^k) = 0$ and $d_{\tan}^k \neq 0$) the proof follows Theorem 4.1 of [18]. If $C_{\xi^k}(s^k) = 0$ and $d_{\tan}^k = 0$, we have $\xi^k > 0$ (otherwise the algorithm stops). In this case, the current iteration finishes, going back to the restoration phase.

The other algorithms are also well defined since they correspond to $\mathrm{IR}_{\mathrm{L}}^{\xi}$ setting $\lambda^k = 0$ and/or $\xi^k = 0$. □

3.3 Convergence to feasible points

If the $\mathrm{IR}_{\mathrm{L}}^{\xi}$ method stops at iteration k, then $C_{\xi^k}(s^k) = 0$ with $\xi^k = 0$, i.e., s^k is a feasible point of the KKT reformulation of the bilevel problem (5). The case of possibly infinite number of iterations are treated in the next theorem, which is an obvious adaptation of Theorems 3.4 and 3.5 of [19].

Theorem 2 *Suppose that assumptions A1, A2, A3 and A4 are valid. If the IR_L^{ξ} algorithm generates an infinite sequence, then*

$$\lim_{k\to\infty} Ared_k = 0 \quad \textit{and} \quad \lim_{k\to\infty} |C_{\xi^k}(s^k)| = 0.$$

The same occurs when $\xi^k = 0$ and/or $\lambda^k = 0$ for all k, that is, analogous results are valid for each variant of the inexact restoration algorithm.

Observe that a feasible limit point is obtained in the perturbed variants by Theorem 2, $\xi^k \to 0$ and assumption A1.

3.4 Convergence to optimality

As we have already mentioned, the classical inexact restoration method for standard nonlinear programming converges to AGP points [19] when the tangent directions are defined by the upper level objective function. In our case, the AGP condition is related to the vector $d_{\tan,\mathrm{F}}(s)$ defined by

$$d_{\tan,\mathrm{F}}(s) = P_{\pi}[s - \eta \nabla_s F(s)] - s$$

for any $s = (w,x,\mu,\gamma) \in W \times X \times \Delta$ (the domain of F is extended to the s-space in a natural way). Clearly, concepts based on AGP will be closely related to the IR_{F} and $\mathrm{IR}_{\mathrm{F}}^{\xi}$ variants. First, let us consider the KKT reformulation of the weighted-sum reformulation (5) of the bilevel problem P,

$$\begin{array}{ll} \min\limits_{w,x,\mu,\gamma} & F(x) \\ \text{s.t.} & C(w,x,\mu,\gamma) = 0, \quad x \geq 0 \\ & (w,x,\mu,\gamma) \in W \times X \times \Delta. \end{array} \tag{24}$$

Theorem 2 ensures that all inexact restoration strategies converge to a feasible point of (24). We then define two optimality conditions for the bilevel programming based on the AGP condition as follows [1]:

- **Weak AGP**. A feasible point s^* of problem (24) satisfies the *weak AGP* condition of problem (5) if there exists a sequence $\{s^k\}$ converging to s^* such that $d_{\tan,\mathrm{F}}(s^k) \to 0$;
- **Strong AGP**. A point $s^* = (w^*, x^*, \mu^*, \gamma^*)$ satisfies the *strong AGP* condition of problem (5) if it conforms to the weak AGP condition, $C(s^*) = 0$ and (w^*, x^*) is feasible for (5).

Weak AGP consists exactly in the classical AGP condition for problem (24). Strong AGP requires additionally that the primal vector (w^*, x^*) is a global optimal solution of the lower level problem of the original weighted bilevel problem (5). As the AGP condition is a legitimate necessary optimality condition [20] for standard nonlinear programming, weak and strong AGP are necessary optimality conditions for the bilevel problem (5).

A usual sufficient assumption to obtain convergence results for inexact restoration methods, incorporating the pertubation in a natural way, is the following [1, 18, 19]:

A5. There exists $\beta > 0$, independently of k, such that

$$\|s^k - z^k\| \le \beta |C_{\xi^k}(s^k)|.$$

The next theorem is valid for all inexact restoration strategies, and it is analogous to Theorem 4.3 of [1]. We only state the result for the $\mathrm{IR}_\mathrm{L}^\xi$ variant, since it encompasses the other ones.

Theorem 3 *Suppose that assumptions A1, A2, A3, A4 and A5 hold. If $\{s^k\}$ is an infinite sequence generated by IR_L^ξ and $\{z^k\}$ is the sequence defined at the restoration phase, then*

1. $\left|C_{\xi^k}(s^k)\right| \to 0$;
2. *There exists a limit point s^* of $\{s^k\}$;*
3. *Every limit point of $\{s^k\}$ is a feasible point of the KKT reformulation* (24)*;*
4. *If, for all w, a global solution of the lower level problem is found then any limit point (w^*, x^*) is feasible for the weighted bilevel problem* (5)*;*
5. *If s^* is a limit point of $\{s^k\}$, there exists an infinite set $K \subset \mathbb{N}$ such that*

$$\lim_{k \in K} s^k = \lim_{k \in K} z^k = s^*, \quad C(s^*) = 0 \quad \textit{and} \quad \lim_{k \in K} d^k_{tan} = 0.$$

Proof The first two items follow from Theorem 2 and the assumption A1. Consequentely, the third and fourth items are valid. The fifth item follows from assumption A5, from the first item and from arguments similar to those of Theorem 5.2 of [18]. □

Item 5 of Theorem 3 implies the fulfillment of AGP-type necessary optimality conditions in the IR_F and $\mathrm{IR}_\mathrm{F}^\xi$ variants. We summarize this in the next result.

Corollary 1 *Suppose that assumptions A1, A2, A3, A4 and A5 hold. If $\{s^k\}$ is an infinite sequence generated by the IR_F or the IR_F^{ξ} method, then*

1. *Every limit point s^* is a weak AGP point;*
2. *If, for all w, a global solution of the lower level problem is found then any limit point s^* is a strong AGP point.*

3.5 Sufficient conditions for the assumptions

Compactness of X is a natural way to ensure the well-definiteness of optimization algorithms in the literature. In this sense, assumption A1 is a slightly stringent condition. Assumptions A2 and A3 are natural in the context of inexact restoration methods [19, 18] (and hence, we assume assumptions A1 to A3 from now on). However, assumptions A4 and A5 are related to sequences generated by the IR algorithms. Thus, it is interesting to establish sufficient conditions that ensure their validity and that do not depend on the progress of the methods. To verify A4, it is convenient that the lower level KKT perturbed system (22) has a solution. Particularly in the case of $\xi = 0$, it is equivalent to require that the lower level has a first order stationary point. Following [1], we consider the next hypotheses throughout this subsection:

A6. For each $\xi \in [0, \xi_{\max}]$, every solution $s = (w, x, \mu, \gamma)$ of (22) is such that the gradients of the active lower level constraints

$$\nabla h_i(x),\ i = 1, \ldots, m, \quad \text{and} \quad -e_j, \ \text{for } j \text{ such that } x_j = 0,$$

where e_j is the j-th canonical vector of $\mathbb{R}^n$, are linearly independent.

A7. For each $\xi \in [0, \xi_{\max}]$, every solution $s = (w, x, \mu, \gamma)$ of (22) is such that the matrix

$$H(w, x, \mu) = \sum_{i=1}^{p} w_i \nabla^2_{xx} f_i(x) + \sum_{i=1}^{m} \mu_i \nabla^2_{xx} h_i(x)$$

is positive definite in the set

$$Z(x, \gamma) = \{d \in \mathbb{R}^n \mid \nabla h(x)^T d = 0,\ d_j = 0 \text{ for all } j \text{ such that } \gamma_j = 0\}.$$

A8. For $\xi = 0$, every solution $s = (w, x, \mu, \gamma)$ of (22) satisfies $x_i + \gamma_i > 0$ for all i (note that this is trivially satisfied if $\xi > 0$).

When $\xi = 0$, assumption A6 implies the well known Linear Independence Constraint Qualification (LICQ) at every solution of the lower level problem. Thus, Δ may be naturally assumed a compact set, as in assumption A1.

We proceed analogously to the discussion made in [1]. Let us consider the matrix

$$D'(\xi, w, x, \mu, \gamma) = \begin{bmatrix} H(w, x, \mu) & \nabla h(x) & -I_n \\ \nabla h(x)^T & 0 & 0 \\ \text{diag}(\gamma) & 0 & \text{diag}(x) \end{bmatrix}.$$

We observe that D' is a submatrix of C' (or C'_{ξ}, see expression (8)), and does not depend on ξ.

Lemma 1 *The matrix $D'(\xi,w,x,\mu,\gamma)$ is non-singular for any solution (ξ,w,x,μ,γ) of* (22).

Proof For $\xi = 0$, this is exactly Lemma 4.4 of [1]. For $\xi > 0$, the proof is the same since D' is constant with respect to ξ. □

Let $D(\xi,w,x,\mu,\gamma) = C_\xi(w,x,\mu,\gamma)$ be defined on $[0,\xi_{\max}] \times W \times X \times \Delta$. As $D(\xi,w,x,\mu,\gamma)$ is continuous with respect to (ξ,x,μ,γ), it is possible to choose, for each $w \in W$, a solution $u_\xi(w) = (x_\xi(w),\mu_\xi(w),\gamma_\xi(w))$ of (22) such that the function $v(\xi,w) = u_\xi(w)$ is continuous on $[0,\xi_{\max}] \times W$. From now on, let the function $v(\xi,w)$ be fixed. By Lemma 1, we can define a function Υ over the set $[0,\xi_{\max}] \times W$ by

$$\Upsilon(\xi,w) = \left[D'(\xi,w,v(\xi,w))\right]^{-1}. \tag{25}$$

Let $V(v(\xi,w),\alpha) = \{v \in [0,\xi_{\max}] \times X \times \Delta \mid \|v - v(\xi,w)\| \le \alpha\}$.

Lemma 2
1. *There exists $\beta > 0$ such that $|\Upsilon(\xi,w)| \le \beta$ for all $(\xi,w) \in [0,\xi_{\max}] \times W$.*
2. *There exists $\alpha > 0$ such that Υ coincides with the local inverse operator of $D'(\xi,w,\cdot)$ for all $v \in V(v(\xi,w),\alpha)$.*

Proof As $D'(\xi,w,v)$ is continuous on (w,v) and $v(\xi,w)$ on (ξ,w), $\Upsilon(\xi,w)$ is continuous with respect to $(\xi,w) \in [0,\xi_{\max}] \times W$. As this set is compact, there exists $\beta > 0$ such that $|\Upsilon(\xi,w)| \le \beta$ for all $(\xi,w) \in [0,\xi_{\max}] \times W$. Thus, the first statement was proved.

The second statement can be obtained in the same way as in Lemma 4.5 of [1]. □

It is worth noticing that to adapt the analysis previously presented in [1] to our context, in which the perturbation is considered, the parameter ξ had to be incorporated as a new variable in (25). This is enough to ensure that the radius α, cf. the second item of Lemma 2, does not depend on ξ. This independence allows us to obtain local and global error bounds on the function D, analogous to Lemmas 4.6 and 4.7 of [1], respectively.

Finally, we state that A4 and A5 hold under assumptions A6 to A8 of this subsection. The next theorem summarizes this fact, and it can be proved by obvious adaptations on Theorem 4.8 of [1].

Theorem 4 *Let $\xi \in [0,\xi_{\max}]$ and $r \in [0,1)$. Let $(w,u_\xi) \in W \times X \times \Delta$ be such that $C_\xi(w,u_\xi) \neq 0$. If the assumptions A6–A8 hold, there exist $\beta > 0$, independent of w and ξ, and $\overline{u}_\xi = (\overline{x},\overline{\mu},\overline{\gamma}) \in X \times \Delta$ such that*

$$|C_\xi(w,\overline{u}_\xi)| \le r|C_\xi(w,u_\xi)|$$

and

$$\|(w,u_\xi) - (w,\overline{u}_\xi)\| \le \beta|C_\xi(w,u_\xi)|.$$

4 Numerical experiments

First of all, we describe details of our implementation of Algorithm 1, namely the strategy $\mathrm{IR_L}$. The same observations are valid for the other variants.

- We use the stabilized sequential quadratic programming (SSQP) method implemented in the package WORHP [5] to solve the problem of step 2 (including (23)) and to compute the projection in step 3. WORHP is a robust and efficient software library, especially designed for large problems;
- Theoretical convergence results require assumption A5 whenever s^k is not a lower level stationary point. In [1], the authors proposed to monitor the growth of the quotient $\|z^k - s^k\|/\|C(s^k)\|$, comparing it to the quotient of the previous iteration (note that the theory does not require any relationship between iterations). Instead, we simply verify if this quotient exceeds 10^8. If this is the case, we stop the algorithm declaring failure. In our experiments this situation never happened, and we will not make further comments along this line;
- We consider that the condition $d^k_{\tan} = 0$ in step 4 was satisfied if $\|d^k_{\tan}\|_\infty \leq \varepsilon_{\mathrm{opt}}$ for a given tolerance to optimality $\varepsilon_{\mathrm{opt}} > 0$. If this condition does not hold, we should compute $v^{k,i} \in \pi_k$ such that, for some $0 < t \leq t^{k,i}_{\mathrm{break}}$,

  $$L(v^{k,i}, \lambda^k) \leq \max\{L(z^k + t d^k_{\tan}, \lambda^k), L(z^k, \lambda^k) - \tau_1 \delta_{k,i}, L(z^k, \lambda^k) - \tau_2\}$$

  with $\gamma \geq 0$ and $\|v^{k,i} - z^k\|_\infty \leq \delta_{k,i}$. Initially, we calculate

  $$L^* = \max\{\, L(z^k + t^{k,i}_{\mathrm{break}} d^k_{\tan}, \lambda^k)\,, L(z^k, \lambda^k) - \tau_1 \delta_{k,i}\,, L(z^k, \lambda^k) - \tau_2 \,\}$$

  and we consider the optimization problem

  $$\begin{array}{ll} \displaystyle\min_{v=(w,x,\mu,\gamma)} & L(v, \lambda^k) \\ \text{s.t.} & C'(z^k)(v - z^k) = 0, \quad x \geq 0 \\ & v \in W \times X \times \Delta \\ & \|v - z^k\|_\infty \leq \delta_{k,i}. \end{array}$$

  A stationary point v^* of this problem is potentially a good point for step 4. We then solve it by the SSQP method of the WORHP package, initialized at z^k as proposed in [1]. If the SSQP method fails, or if the "optimal" value $L(v^*, \lambda^k)$ found is not sufficiently small (i.e. $L(v^*, \lambda^k) > L^*$), then we reduce the neighborhood size $\delta_{k,i}$ by the factor 0.8. This reduction anticipates the one that will be done in step 7. So we recalculate $t^{k,i}_{\mathrm{break}}$ and test if $L(v^{k,i} + t^{k,i}_{\mathrm{break}} d^k_{\tan}, \lambda^k) \leq L^*$ (this represents a step along the descent direction $d^k_{\tan}$). We make these reductions until $\delta_{k,i} \geq \delta_{\min}$ for a given parameter $\delta_{\min} > 0$. In case $\delta_{k,i} < \delta_{\min}$ we stop, declaring failure. It should be mentioned that if the SSQP strategy fails, we never reapply it in the same iteration of the inexact restoration method because there is no reason to expect that this strategy can reach a good point with a smaller neighborhood radius;
- In step 5, it is important that $\lambda^{k,i}$ is a good estimative for a solution of $\nabla_v L(v^{k,i}, \lambda) = 0$ (or equivalently, a good approximation for a solution of the linear system $C'(v^{k,i})\lambda = -\nabla_v F(x)$). If the SSQP method has success in step 4, we take $\lambda^{k,i}$

as the computed multipliers. Otherwise, i.e. if $v^{k,i}$ was obtained by backtracking along direction $d^k_{\tan}$, we try to find a least squares solution of the overdetermined linear system

$$C'(v^{k,i})\lambda = -\nabla_v F(x)$$

by the orthogonal factorization method implemented in the MA49 of HSL [15]. In case of failure, we set $\lambda^{k,i}$ as the last $\lambda^{k,i}$ computed throughout the process. In all cases we project $\lambda^{k,i}$ onto the box $[-M,M]^{2n+m}$, and thus $\|\lambda^{k,i}\|_\infty \leq M$;
- Finally, we make double-precision floating point corrections for the sums $\sum_{i=1}^{p} w_i$ of s^k, z^k, $v^{k,i}$ (to 1) and $d^k_{\tan}$ (to 0). These corrections aim to avoid numerical errors.

We compared the inexact restoration strategy (with and without perturbation) with the classical KKT reformulation (24). The resulting problem is a special case of mathematical programming problem with complementarity constraints, in which the SSQP strategy is suitable [9–11]. Thus we solve it by the WORHP package.

The WORHP package permits several configurations; among them we highlight:

- The globalization strategy is line search with an augmented Lagrangian merit function. This is standard in WORHP;
- Quadratic subproblems are solved by an interior point method;
- The BFGS updates for Hessian approximations are employed. This is standard in WORHP;
- Symmetric linear systems are solved by the multifrontal strategy MA97 of HSL [15].

We have addressed 13 test problems. Some of them were built using existing techniques or multiobjective instances in the literature, and some are completely new. Details are in the Appendix A. Varying the starting point x^0, the initial weights w^0 and the dimensions p and n, we define 69 different instances, that are summarized in Tables 1 and 2.

The experiments were run in a computer equiped with Intel(R) Core(TM) i7-6600U 2.60Ghz processor, 8Gb RAM, and GNU/Linux Ubuntu 64 bits operational system. The main part of the code was written in Fortran 90 language and compiled with gfortran 4.8 version. Another small part was written in C language, especially the AMPL interface, which is able to model multiobjective problems. For the inexact restoration strategies, we set the tolerance for optimality $\varepsilon_{\text{opt}} = 10^{-4}$ (step 3 of Algorithm 1) and the parameters $\eta = 0.1$, $\theta_{-1} = 0.5$, $r = 0.5$, $\varepsilon_i = 10^{-4}$ $\forall i$, $M = 10^{16}$, $\delta_{\min} = \varepsilon_{\text{opt}}/2$, $\tau_1 = 10^{-2}$, $\tau_2 = 10^{-3}$, $\omega^k = 1/k^2$, $\lambda^0 = 0$. For all subproblems (steps 2, 3 and 4) we set the optimality tolerance of the SSQP method equal to $\varepsilon_{\text{SQP}} = 10^{-6}$. We do not perturb the complementarity for more than $\xi_{\max} = 10^{-2}$ in $\text{IR}^{\xi}_{\text{F}}$ and $\text{IR}^{\xi}_{\text{L}}$ algorithms, while the forced decrease factor ξ_{dec} is set to 0.8 (see expression (20)). We initialize step 3 with $\delta_{k,0} = 10$ (according with [1]) and in step 7 we make $\delta_{k,i+1} = \delta_{k,i}/2$.

We have compared the inexact restoration strategies with the KKT reformulation. Tables 3 and 4 show the results. The column "F" contains the final functional value obtained by the method. Bold values represent the best value of F, while italic values indicate that the method reached the optimal value. We decide that two methods reached the same functional value or a method attains the optimal value whenever

Table 1 Instances of Problems 1 to 8.

Instance	p	n	x^0	w^0	F^*
P1.1	2	2	$(10,10)$	$(0.5,0.5)$	1.00e+01
P2.1	2	2	$(10,10)$	$(0.5,0.5)$	9.50e−01
P3.1	2	2	$(-2,20)$	$(0.8,0.2)$	0.00e+00
P3.2	2	2	$(-10,10)$	$(0.4,0.6)$	0.00e+00
P4.1	2	2	$(-10,10)$	$(0.5,0.5)$	0.00e+00
P4.2	2	2	$(-10,10)$	$(0.4,0.6)$	0.00e+00
P5.1	2	2	$(-10,10)$	$(0.5,0.5)$	5.01e−01
P5.2	2	2	$(40,100)$	$(0.1,0.9)$	5.01e−01
P6.1	2	2	$(-10,10)$	$(0.5,0.5)$	0.00e+00
P6.2	2	2	$(10,20)$	$(0.73,0.27)$	0.00e+00
P6.3	2	2	$(10,20)$	$(0.23,0.77)$	0.00e+00
P7.1	3	2	$(10,10)$	$(0.2,0.3,0.5)$	0.00e+00
P7.2	3	2	$(0.68,1.72)$	$(0.2497,0.6272,0.1231)$	0.00e+00
P7.3	3	2	$(0.68,1.72)$	$(0.3,0.6,0.1)$	0.00e+00
P7.4	6	2	$(-10,-10)$	$w_i^0 = 1/6$	0.00e+00
P7.5	100	2	$(9/10,9/10)$	$w_i^0 = 1/100$	0.00e+00
P7.6	200	2	$(10,1)$	$w_i^0 = 1/200$	0.00e+00
P8.1	2	2	$(20,10)$	$(0.9,0.1)$	0.00e+00
P8.2	2	2	$(10,40)$	$(0.45,0.55)$	0.00e+00
P8.3	2	2	$(10^{-4},10)$	$(0.8,0.2)$	0.00e+00
P8.4	2	6	$x_{2i}^0 = 20, x_{2i+1}^0 = 10$	$(0.9,0.1)$	0.00e+00
P8.5	2	6	$x_{2i}^0 = 10, x_{2i+1}^0 = 40$	$(0.45,0.55)$	0.00e+00
P8.6	2	6	$x_{2i}^0 = 10^{-4}, x_{2i+1}^0 = 10$	$(0.82,0.18)$	0.00e+00
P8.7	2	20	$x_{2i}^0 = 20, x_{2i+1}^0 = 10$	$(0.9,0.1)$	0.00e+00
P8.8	2	20	$x_{2i}^0 = 10, x_{2i+1}^0 = 40$	$(0.45,0.55)$	0.00e+00
P8.9	2	20	$x_{2i}^0 = 10^{-4}, x_{2i+1}^0 = 10$	$(0.82,0.18)$	0.00e+00
P8.10	2	30	$x_{2i}^0 = 20, x_{2i+1}^0 = 10$	$(0.9,0.1)$	0.00e+00
P8.11	2	30	$x_{2i}^0 = 10, x_{2i+1}^0 = 40$	$(0.45,0.55)$	0.00e+00

the difference between the functional values is at most 10^{-4}. The column "It" contains the number of outer iterations. We do not report the number of iterations for the KKT reformulation because its not comparable with the outer iterations of inexact restoration techniques. On the other hand, the number of iterations of all IR strategies are comparable. The column "Status" reports the answer of the method for the KKT reformulation: a check mark to indicate convergence, "max it" if the maximum number of iterations (1000) is reached and "infeas" if the SSQP method converges to a point that does not conform to the first order stationary conditions of the lower level problem. We note that all IR strategies declare convergence within 150 iterations.

Analysing the highlighted best and optimal values of F of both tables, for problems 1 to 8 (Table 3), the inexact restoration strategy, particulary its perturbed version $\mathrm{IR}_{\mathrm{F}}^{\xi}$, overcame the KKT reformulation. Actually, for problems 1 to 8, the strategy $\mathrm{IR}_{\mathrm{F}}^{\xi}$ had the best performance, and $\mathrm{IR}_{\mathrm{L}}^{\xi}$ was the second best strategy. These problems have nonlinear constraints, whereas the remaining have just bound constraints. For

Table 2 Instances of Problems 9 to 13.

Instance	p	n	x^0	w^0	F^*
P9.1	2	2	$x^0_{2i}=20, x^0_{2i+1}=10$	$(0.9,0.1)$	0.00e+00
P9.2	2	2	$x^0_{2i}=10, x^0_{2i+1}=40$	$(0.45,0.55)$	0.00e+00
P9.3	2	2	$(10^{-4},10)$	$(0.82,0.18)$	0.00e+00
P9.4	2	6	$x^0_{2i}=20, x^0_{2i+1}=10$	$(0.9,0.1)$	0.00e+00
P9.5	2	6	$x^0_{2i}=10, x^0_{2i+1}=40$	$(0.45,0.55)$	0.00e+00
P9.6	2	6	$x^0_{2i}=10^{-4}, x^0_{2i+1}=10$	$(0.82,0.18)$	0.00e+00
P9.7	2	20	$x^0_{2i}=20, x^0_{2i+1}=10$	$(0.9,0.1)$	0.00e+00
P9.8	2	20	$x^0_{2i}=10, x^0_{2i+1}=40$	$(0.45,0.55)$	0.00e+00
P9.9	2	20	$x^0_{2i}=10^{-4}, x^0_{2i+1}=10$	$(0.82,0.18)$	0.00e+00
P9.10	2	30	$x^0_{2i}=20, x^0_{2i+1}=10$	$(0.9,0.1)$	0.00e+00
P9.11	2	30	$x^0_{2i}=10, x^0_{2i+1}=40$	$(0.45,0.55)$	0.00e+00
P10.1	2	2	$(20,10)$	$(0.9,0.1)$	2.12e−03
P10.2	2	2	$(10,40)$	$(0.45,0.55)$	2.12e−03
P10.3	2	2	$(10^{-4},10)$	$(0.82,0.18)$	2.12e−03
P10.4	2	6	$x^0_{2i}=20, x^0_{2i+1}=10$	$(0.9,0.1)$	2.12e−03
P10.5	2	6	$x^0_{2i}=10, x^0_{2i+1}=40$	$(0.45,0.55)$	2.12e−03
P10.6	2	6	$x^0_{2i}=10^{-4}, x^0_{2i+1}=10$	$(0.82,0.18)$	2.12e−03
P10.7	2	20	$x^0_{2i}=20, x^0_{2i+1}=10$	$(0.9,0.1)$	2.12e−03
P10.8	2	20	$x^0_{2i}=10, x^0_{2i+1}=40$	$(0.45,0.55)$	2.12e−03
P10.9	2	20	$x^0_{2i}=10^{-4}, x^0_{2i+1}=10$	$(0.82,0.18)$	2.12e−03
P10.10	2	30	$x^0_{2i}=20, x^0_{2i+1}=10$	$(0.9,0.1)$	2.12e−03
P10.11	2	30	$x^0_{2i}=10, x^0_{2i+1}=40$	$(0.45,0.55)$	2.12e−03
P11.1	2	2	$(20,10)$	$(0.9,0.1)$	0.00e+00
P11.2	2	2	$(10,40)$	$(0.45,0.55)$	0.00e+00
P11.3	2	2	$(10^{-4},10)$	$(0.82,0.18)$	0.00e+00
P11.4	2	6	$x^0_{2i}=20, x^0_{2i+1}=10$	$(0.9,0.1)$	0.00e+00
P11.5	2	6	$x^0_{2i}=10, x^0_{2i+1}=40$	$(0.45,0.55)$	0.00e+00
P11.6	2	6	$x^0_{2i}=10^{-4}, x^0_{2i+1}=10$	$(0.82,0.18)$	0.00e+00
P11.7	2	20	$x^0_{2i}=20, x^0_{2i+1}=10$	$(0.9,0.1)$	0.00e+00
P11.8	2	20	$x^0_{2i}=10, x^0_{2i+1}=40$	$(0.45,0.55)$	0.00e+00
P11.9	2	20	$x^0_{2i}=10^{-4}, x^0_{2i+1}=10$	$(0.82,0.18)$	0.00e+00
P11.10	2	30	$x^0_{2i}=20, x^0_{2i+1}=10$	$(0.9,0.1)$	0.00e+00
P11.11	2	30	$x^0_{2i}=10, x^0_{2i+1}=40$	$(0.45,0.55)$	0.00e+00
P12.1	2	2	$(10,10)$	$(0.75,0.25)$	0.00e+00
P12.2	2	2	$(-10,-10)$	$(0.35,0.65)$	0.00e+00
P12.3	2	2	$(0.5,0.5)$	$(0.5,0.5)$	0.00e+00
P12.4	6	6	$x^0_i=1/2$	$w^0_i=1/6$	0.00e+00
P13.1	2	2	$(0.7,0.7)$	$(0.5,0.5)$	0.00e+00
P13.2	2	2	$(10,10)$	$(0.75,0.25)$	0.00e+00
P13.3	2	2	$(10,10)$	$(0.25,0.75)$	0.00e+00
P13.4	6	6	$x^0_i=7/10$	$w^0_i=1/6$	0.00e+00

problems 9 to 13 (Table 4), among the IR strategies, $\mathrm{IR}^{\xi}_{\mathrm{L}}$ stood out, being the KKT reformulation the second best. Hence, we have observed that the inexact restoration strategy suits better for problems that have more complicated constraints.

All in all, the perturbed inexact restoration proved effective for problems 1 to 5, 7, 8, 12 and 13. It is worth mentioning that the set of efficient solutions of problem

Table 3 Numerical results of Problems 1 to 8.

	IR_F		IR_F^ξ		IR_L		IR_L^ξ		KKT	
Inst.	F	It	F	It	F	It	F	It	F	Status
P1.1	***1.00e+01***	7	***1.00e+01***	8	1.03e+01	5	1.44e+01	8	1.00e+20	infeas
P2.1	1.10e+00	2	**1.08e+00**	5	1.10e+00	2	**1.08e+00**	5	1.00e+20	infeas
P3.1	***7.95e−10***	5	***2.19e−15***	5	***1.06e−14***	4	***2.12e−15***	5	***4.82e−30***	✓
P3.2	***1.92e−12***	2	***2.99e−08***	5	***1.92e−12***	2	***2.99e−08***	5	***3.15e−14***	✓
P4.1	2.00e+00	3	***1.12e−18***	6	2.00e+00	3	***1.44e−20***	7	2.00e+00	✓
P4.2	2.00e+00	2	***2.50e−24***	6	2.00e+00	2	2.40e−01	6	2.00e+00	✓
P5.1	2.00e+00	2	***5.01e−01***	5	2.00e+00	2	***5.01e−01***	5	1.44e+00	✓
P5.2	***5.01e−01***	3	***5.01e−01***	5	***5.01e−01***	3	***5.01e−01***	5	1.44e+00	✓
P6.1	***3.32e−06***	3	***3.27e−06***	5	***3.32e−06***	3	4.92e−01	5	7.07e+02	✓
P6.2	***3.28e−06***	4	***3.28e−06***	6	***3.28e−06***	4	2.12e+00	5	1.00e+20	max it
P6.3	***3.25e−06***	3	***3.26e−06***	6	***3.25e−06***	3	***3.26e−06***	6	1.00e+20	infeas
P7.1	***9.00e−08***	2	***9.03e−08***	5	***9.00e−08***	2	***9.03e−08***	5	4.00e+00	✓
P7.2	3.45e+00	2	***8.97e−08***	5	3.45e+00	2	***1.10e−07***	5	4.00e+00	✓
P7.3	3.45e+00	2	***8.97e−08***	5	3.45e+00	2	***1.52e−06***	5	4.00e+00	✓
P7.4	***3.55e−12***	2	***1.19e−19***	5	***3.55e−12***	2	***1.66e−24***	5	4.00e+00	✓
P7.5	***1.21e−09***	3	***1.16e−19***	5	***1.21e−09***	3	***1.66e−24***	5	1.00e+20	infeas
P7.6	***1.08e−09***	3	***1.34e−19***	5	***1.08e−09***	3	***1.70e−24***	5	1.00e+20	infeas
P8.1	1.00e+00	2	***1.32e−20***	6	1.00e+00	2	***1.78e−20***	6	***8.92e−15***	✓
P8.2	***2.76e−21***	2	***3.72e−09***	5	***2.76e−21***	2	***3.72e−09***	5	1.00e+20	infeas
P8.3	1.00e+00	2	***9.41e−19***	5	1.00e+00	2	***9.40e−19***	5	***3.07e−11***	✓
P8.4	1.00e+00	2	***7.44e−09***	4	1.00e+00	2	***8.28e−09***	4	***5.17e−08***	✓
P8.5	***1.68e−16***	2	***6.57e−09***	5	***1.68e−16***	2	***6.57e−09***	5	1.00e+20	infeas
P8.6	1.00e+00	2	***1.49e−19***	5	1.00e+00	2	***1.46e−19***	5	***5.66e−09***	✓
P8.7	1.00e+00	2	***1.41e−20***	5	1.00e+00	2	***2.92e−18***	5	1.00e+20	infeas
P8.8	***8.96e−19***	2	***6.45e−08***	5	***8.96e−19***	2	***6.45e−08***	5	1.00e+20	infeas
P8.9	1.00e+00	2	***2.21e−20***	5	1.00e+00	2	***3.45e−16***	5	***2.96e−13***	✓
P8.10	1.00e+00	2	***2.41e−20***	5	1.00e+00	2	***7.10e−18***	5	1.00e+20	infeas
P8.11	***2.02e−20***	2	***4.82e−08***	5	***2.02e−20***	2	***4.82e−08***	5	***1.26e−09***	✓

11 is composed by several discontinuous parts (see Table 5 and the discussion in the Appendix A), what possibly justifies the poor performance of the KKT reformulation, never reaching an optimal value for any instance of this problem.

Despite the good behavior of the inexact restoration method, especially its perturbed version, we have not obtained practical evidences to affirm that the IR approach *always* overcame the KKT reformulation. Indeed, for problem 10, the KKT reformulation obtained the best results in 9 out of the 11 instances. Nevertheless, for 19 out of the 69 total tested instances, the KKT reformulation could not achieve a first order stationary point, being these "infeas" occurrences connected with the lack of robustness of such a strategy. Among these failures, we highlight instance P1.1, thoroughly discussed as Example 1. For this particular instance, as can be seen in Table 3, the lack of preference of the KKT reformulation towards optimizers clearly leaves such a strategy in disadvantage in comparison with the four IR-based strategies.

Table 4 Numerical results of Problems 9 to 13.

	$\mathrm{IR_F}$		$\mathrm{IR_F^{\xi}}$		$\mathrm{IR_L}$		$\mathrm{IR_L^{\xi}}$		KKT	
Inst.	*F*	It	*F*	It	*F*	It	*F*	It	*F*	Status
P9.1	1.00e+00	2	***1.11e−21***	6	1.00e+00	2	***1.10e−21***	6	***2.16e−15***	✓
P9.2	***1.24e−21***	2	1.00e+00	5	***1.24e−21***	2	***3.46e−23***	5	***2.02e−10***	✓
P9.3	1.00e+00	2	1.00e+00	5	1.00e+00	2	1.00e+00	5	***7.61e−24***	✓
P9.4	1.00e+00	2	1.00e+00	5	1.00e+00	2	1.00e+00	5	***2.17e−12***	✓
P9.5	***6.35e−22***	2	1.00e+00	6	***6.35e−22***	2	1.00e+00	6	***6.53e−08***	✓
P9.6	1.00e+00	2	***5.10e−09***	4	1.00e+00	2	***5.10e−09***	4	1.00e+20	infeas
P9.7	1.00e+00	2	1.00e+00	5	1.00e+00	2	1.00e+00	5	***1.76e−09***	✓
P9.8	1.00e+00	2	1.00e+00	5	1.00e+00	2	1.00e+00	5	***1.79e−07***	✓
P9.9	1.00e+00	2	***1.10e−08***	4	1.00e+00	2	***1.10e−08***	4	1.00e+20	infeas
P9.10	1.00e+00	2	1.00e+00	5	1.00e+00	2	1.00e+00	5	***9.80e−08***	✓
P9.11	1.00e+00	2	1.00e+00	6	1.00e+00	2	1.00e+00	6	***2.03e−08***	✓
P10.1	1.00e+00	2	1.00e+00	5	1.00e+00	2	1.00e+00	5	**2.39e−03**	✓
P10.2	6.17e−02	2	1.00e+00	6	6.17e−02	2	1.00e+00	6	**2.29e−03**	✓
P10.3	1.00e+00	2	1.00e+00	4	1.00e+00	2	1.00e+00	4	**2.39e−03**	✓
P10.4	1.00e+00	2	2.56e−03	5	1.00e+00	2	2.56e−03	5	***2.02e−03***	✓
P10.5	6.17e−02	2	9.65e−01	5	6.17e−02	2	9.65e−01	5	**2.29e−03**	✓
P10.6	1.00e+00	2	**2.40e−03**	4	1.00e+00	2	**2.40e−03**	4	**2.39e−03**	✓
P10.7	1.00e+00	2	1.00e+00	5	1.00e+00	2	**9.65e−01**	5	1.00e+20	infeas
P10.8	9.65e−01	3	**2.29e−03**	5	9.65e−01	3	**2.29e−03**	5	2.09e−02	✓
P10.9	1.00e+00	2	1.00e+00	5	1.00e+00	2	9.65e−01	6	**6.81e−01**	✓
P10.10	1.00e+00	2	1.00e+00	4	1.00e+00	2	1.00e+00	5	***2.02e−03***	✓
P10.11	7.02e−01	3	9.65e−01	5	7.02e−01	3	9.65e−01	5	**2.29e−03**	✓
P11.1	1.99e+00	2	***5.82e−24***	5	1.99e+00	2	1.00e+00	5	1.00e+20	infeas
P11.2	**4.95e+00**	2	2.47e+01	5	**4.95e+00**	2	2.47e+01	5	1.00e+20	infeas
P11.3	**9.89e+00**	2	1.58e+01	5	**9.89e+00**	2	1.58e+01	5	1.00e+20	infeas
P11.4	5.11e+01	2	1.06e+01	8	5.11e+01	2	**8.90e+00**	6	1.14e+02	✓
P11.5	**5.11e+01**	2	1.24e+02	6	**5.11e+01**	2	1.24e+02	6	1.05e+02	✓
P11.6	5.11e+01	2	3.93e+01	6	5.11e+01	2	**2.05e+00**	5	1.00e+20	infeas
P11.7	4.28e+02	2	3.11e+02	6	4.28e+02	2	**1.95e+00**	6	3.66e+02	✓
P11.8	4.70e+02	2	4.60e+02	5	4.70e+02	2	**1.94e+02**	6	3.98e+02	✓
P11.9	2.50e+02	2	1.04e+02	6	2.50e+02	2	**2.14e+00**	6	1.00e+20	infeas
P11.10	7.17e+02	2	7.07e+02	7	7.17e+02	2	**4.31e+02**	6	7.10e+02	✓
P11.11	7.17e+02	2	1.00e+20	3	7.17e+02	2	**4.38e+02**	124	6.89e+02	✓
P12.1	***1.80e−19***	3	***4.80e−09***	5	***1.80e−19***	3	***4.80e−09***	5	***2.07e−11***	✓
P12.2	1.00e+00	2	***1.13e−08***	5	1.00e+00	2	***1.13e−08***	5	1.05e−01	✓
P12.3	1.00e+00	2	***7.77e−09***	5	1.00e+00	2	***7.77e−09***	5	***5.65e−11***	✓
P12.4	***3.24e−07***	6	1.00e+00	6	3.00e+00	4	1.00e+00	7	***3.07e−08***	✓
P13.1	***2.72e−08***	4	6.46e−03	5	***2.72e−08***	4	6.46e−03	5	5.71e−03	✓
P13.2	***2.71e−08***	4	***4.43e−08***	6	***2.71e−08***	4	***4.43e−08***	6	***3.66e−08***	✓
P13.3	6.76e−03	2	***8.81e−08***	7	6.76e−03	2	***2.95e−08***	8	1.00e+20	infeas
P13.4	***4.86e−07***	5	***1.12e−05***	16	4.00e+00	3	1.44e+00	6	3.68e−01	✓

5 Final remarks

We have applied inexact restoration strategies to address bilevel programming problems in which the lower level is a multiobjective constrained problem. Compared with the traditional strategy of reformulating the bilevel problem and replacing the lower level optimization by its first-order optimality conditions, our approaches have shown better computational performance, especially in terms of robustness. On one hand, the theoretical convergence analysis of the employed IR must encompass the optimality conditions of the lower level problem and its relationship with the upper level problem. On the other hand, the proposed approach allows a practical deep search into the lower level problem, exploiting its original and intrinsic structure. Moreover, nonlinear problems were solved, with general inequalities in the description of the feasible set of the multiobjective lower level problem, tackled by a perturbation scheme, with promising results. The weighted-sum scalarization reformulation allowed us to put our original and non standard problem as a classical bilevel one. Distinct possibilities for handling the lower level optimization problem, as well as a wider set of computational experiments are objects of future research.

A Test Problems

Problem 1. This is the problem (9)–(10) of Example 1 on page 7.

Problem 2. We define Problem 2 as the minimization of $F(x_1,x_2) = x_1^2 + x_2^2$ over the set $\mathscr{X}^*$ of efficient solutions of the multiobjective constrained problem [24]

$$\begin{array}{ll} \min & \{x_1, x_2\} \\ \text{s.t.} & 1 + 0.1\cos\left(16\arctan\frac{x_2}{x_1}\right) - x_1^2 - x_2^2 \leq 0 \\ & \left(x_1 - \frac{1}{2}\right)^2 + \left(x_2 - \frac{1}{2}\right)^2 \leq \frac{1}{2} \\ & 0 \leq x_1, x_2 \leq \pi. \end{array}$$

$\mathscr{X}^*$ is discontinuous and nonconvex. There are four optimal solutions with functional value $F^* \approx 0.95$ (see Figure 3).

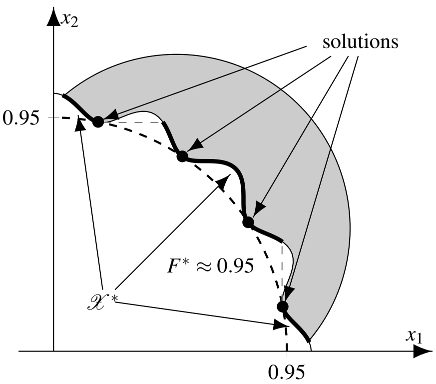

Fig. 3 Geometry of Problem 2. The feasible set $\mathscr{X}^*$ is discontinuous and nonconvex. There are four solutions, with optimal value $F^* \approx 0.95$.

Problem 3. Problem 3 consists in minimizing $F(x_1,x_2) = x_1^2 + x_2^2$ over the set of efficient solutions of

$$\begin{array}{ll} \min & \{-x_1, x_1 + x_2^2\} \\ \text{s.t.} & x_1^2 - x_2 \leq 0 \\ & x_1 + 2x_2 \leq 3 \\ & x_1, x_2 \geq 0. \end{array}$$

The feasible set is $\mathscr{X}^* = \{(x_1,x_2) \in \mathbb{R}^2; x_2 = x_1^2, 0 \leq x_1 \leq 2\}$ and $(0,0)^T$ is the optimal solution.

Problem 4. The lower level multiobjective problem is exactly the same of the Problem 3, and thus the feasible set $\mathscr{X}^*$ is the same. The upper level objective function of Problem 4 is $F(x_1,x_2) = (x_1-1)^2 + (x_2-1)^2$, and $(1,1)^T$ is the optimal solution.

Problem 5. This is a slight modification of Problem 4. It consists in minimizing $F(x_1,x_2) = (x_1-1)^2 + (x_2-1)^2$ over the set of efficient solutions of

$$\begin{aligned} \min\ & \{x_1,\ -x_1+x_2^2\} \\ \text{s.t.}\ & x_1^2 - x_2 \le 0 \\ & -x_1 + 2x_2 \le 3 \\ & x_1, x_2 \ge 0. \end{aligned}$$

The feasible set is $\mathscr{X}^* = \{(x_1,x_2) \in \mathbb{R}^2; x_2 = x_1^2, 0 \le x_1 \le 2^{-2/3}\}$ and the optimal solution is $(2^{-2/3}, 2^{-4/3})^T \approx (0.63, 0.40)^T$.

Problem 6. The lower level problem is

$$\begin{aligned} \min\ & \left\{x_1^2 + \tfrac{1}{4}\left(x_2 - \tfrac{9}{2}\right)^2, x_2^2 + \tfrac{1}{4}\left(x_1 - \tfrac{9}{2}\right)^2\right\} \\ \text{s.t.}\ & -x_1 - x_2 + 2 \le 0 \\ & x_1, x_2 \ge 0. \end{aligned}$$

Geometrically, the set $\mathscr{X}^*$ of efficient solutions of this multiobjective problem is composed by the simultaneously tangential points of the level sets of both objective functions, when it is contained in the semispace defined by $x_1 + x_2 \ge 2$, and by the line $x_1 + x_2 = 2$ otherwise. More specifically, $\mathscr{X}^*$ is the union of the curves

$$\begin{aligned} C_1 &: \left\{\left(\frac{9t}{2t-8}, \frac{9}{2-8t}\right); -\frac{1}{2} \le t \le 0\right\}, \\ C_2 &: \left\{(t, 2-t); \frac{1}{2} \le t \le \frac{3}{2}\right\}, \\ C_3 &: \left\{\left(\frac{9}{2-8t}, \frac{9t}{2t-8}\right); -\frac{1}{2} \le t \le 0\right\}. \end{aligned}$$

The upper level objetive function is $F(x_1,x_2) = (x_1 - 9/2)^2 + x_2^2$, for which $(9/2, 0)^T$ is the optimal solution. Figure 4 illustrates its geometry.

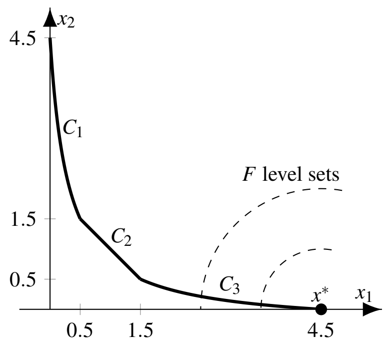

Fig. 4 Geometry of Problem 6. The feasible set is the union of curves C_1, C_2 and C_3. The optimal solution is $x^* = (9/2, 0)^T$.

Problem 7. This is the problem (11)-(12) of Example 2 on page 9.

Table 5 ZDT family instances.

Problem ID	f_1, g, h, box constraints	set of efficient solutions
Problem 8	$f_1 = x_1$ $g = 1 + 9\sum_{i=2}^{n} x_i/(n-1)$ $h = 1 - f_1/g$ $x \in [0,1]^n$	$x_2 = \cdots = x_n = 0$ $x_1 \in [0,1]$
Problem 9	$f_1 = x_1$ $g = 1 + 9\sum_{i=2}^{n} x_i/(n-1)$ $h = 1 - (f_1/g)^2$ $x \in [0,1]^n$	$x_2 = \cdots = x_n = 0$ $x_1 \in [0,1]$
Problem 10	$f_1 = x_1$ $g = 1 + 9\sum_{i=2}^{n} x_i/(n-1)$ $h = 1 - f_1/g(1 - \sin(10\pi f_1))$ $x \in [0,1]^n$	$x_2 = \cdots = x_n = 0$ x_1 in the intervals in which h decreases
Problem 11	$f_1 = x_1$ $g = 2n - 1 + \sum_{i=2}^{n}(x_i^2 - 2\cos(4\pi x_i))$ $h = 1 - (f_1/g)^2$ $x_1 \in [0,1], x_2, \ldots, x_n \in [-5,5]$	several efficient components

Problems 8 to 11 (ZDT family instances). Multiobjective problems of [25] (here named ZDT problems) are constructed with three functions f_1, g, h in the following way:

$$\min \{f_1(x_1),\, f_2(x_1,\ldots,x_n)\}$$

where

$$f_2(x_1,\ldots,x_n) = g(x_2,\ldots,x_n)h(f_1(x_1), g(x_2,\ldots,x_n))$$

(see also [7]). We were inspired in four combinations suggested in Definition 4 of [25]. For each combination, we minimize

$$F(x_1,x_2,\ldots,x_n) = (x_1-1)^2 + \sum_{i=2}^{n} x_i^2$$

over the set of efficient solutions defined by the minimization of functions f_1 and f_2 on certain box constraints. We summarized the properties of these test problems Table 5. For these problems, the set of efficient solutions is split into discontinuous components, one for each (local) minimum of g, formed with all possible values of f_1 (see Section 5.1 of [7]). Thus, Problems 8 to 10 have a unique set of efficient solutions component, for which $x_2 = \cdots = x_n = 0$ (i.e. $g(x_2,\ldots,x_n) = 1$), while Problem 11 has several components. For Problems 8, 9 and 11, the possible values of $x_1 = f_1(x_1)$ are the whole interval $[0,1]$. For Problem 10, the possible values of x_1 are the intervals for which h decreases (see Figure 5). Hence, $(1,0,\ldots,0)$ is the optimal solution of Problems 8, 9 and 11, whereas $(0.954, 0\ldots,0)^T$ is the (approximated) optimal solution of Problem 10. Problem 11 has 21^{n-1} sets of efficient solutions components, one for each combination of the local minimizers of $x_i^2 - 2\cos(4\pi x_i)$, $i = 2,\ldots,n$ (see Figure 6).

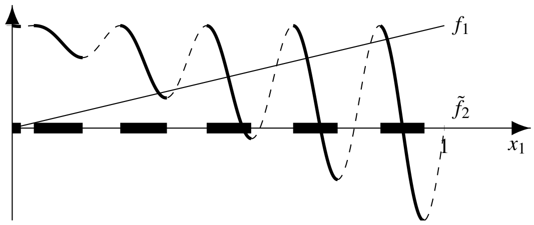

Fig. 5 Geometry of Problem 10. In the set of efficient solutions, $g(x_2^*,\ldots,x_n^*) = 1$ and x_1 lies on the intervals for which $\tilde{f}_2(x_1) = f_2(x_1,x_2^*,\ldots,x_n^*) = 1 - x_1(1-\sin(10\pi x_1))$ decreases (huge lines). These are the non-dominated values corresponding to the two objectives $f_1(x_1) = x_1$ and $\tilde{f}_2(x_1)$.

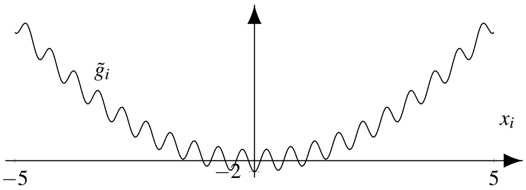

Fig. 6 Geometry of Problem 11. There is one set of efficient solutions component for each combination of local minimizers of $\tilde{g}_i(x_i) = x_i^2 - 2\cos(4\pi x_i)$, $i = 2,\ldots,n$.

Problems 12 and 13 (WFG family instances). Based on the Walking Fish Group (WFG) Toolkit ([16]), we consider two instances, for which particular choices of the parameters were made to come up with smooth functions:

– **Problem 12**, that consists in minimizing $F(x) = \sum_{i=1}^n x_i^2$, over the set of efficient solutions defined by the minimization of

$$f_1(x) = \frac{x_n}{2n} + 2\prod_{i=1}^{n-1}\left[1-\cos\left(\frac{\pi x_i}{4i}\right)\right],$$
$$f_j(x) = \frac{x_n}{2n} + 2j\left[1-\sin\left(\frac{\pi x_{n-j+1}}{4(n-j+1)}\right)\right]\prod_{i=1}^{n-j}\left[1-\cos\left(\frac{\pi x_i}{4i}\right)\right],\ j = 2,\ldots,n-1,$$
$$f_n(x) = \frac{x_n}{2n} + 2n\left[1-\frac{x_1}{2}-\frac{\cos(5\pi x_1+\pi/2)}{10\pi}\right]$$

on $[0,1]^n$. The optimal solution is the origin;

– **Problem 13**, that consists in minimizing $F(x) = \sum_{i=1}^n x_i^2$, over the set of efficient solutions defined by the minimization of

$$f_1(x) = \frac{x_n}{2n} + 2\prod_{i=1}^{n-1}\left[1-\cos\left(\frac{\pi x_i}{4i}\right)\right],$$
$$f_j(x) = \frac{x_n}{2n} + 2j\left[1-\sin\left(\frac{\pi x_{n-j+1}}{4(n-j+1)}\right)\right]\prod_{i=1}^{n-j}\left[1-\cos\left(\frac{\pi x_i}{4i}\right)\right],\ j = 2,\ldots,n-1,$$
$$f_n(x) = \frac{x_n}{2n} + 2n\left[1-\frac{x_1}{2}\cos^2\left(\frac{5\pi x_1}{2}\right)\right]$$

on $[0,1]^n$. Again, the optimal solution is the origin.

References

1. Andreani, R., Castro, S.L.C., Chela, J.L., Friedlander, A., Santos, S.A.: An inexact-restoration method for nonlinear bilevel programming problems. Computational Optimization and Applications **43**(3), 307–328 (2009)
2. Benson, H.P.: An improved definition of proper efficiency for vector maximization with respect to cones. Journal of Mathematical Analysis and Applications **71**(1), 232–241 (1979)
3. Benson, H.P.: Optimization over the efficient set. Journal of Mathematical Analysis and Applications **98**(2), 562–580 (1984)
4. Borwein, J.: Proper efficient points for maximizations with respect to cones. SIAM Journal on Control and Optimization **15**(1), 57–63 (1977)
5. Bskens, C., Wassel, D.: The ESA NLP Solver WORHP, *Springer Optimization and Its Applications*, vol. 73, chap. 8, pp. 85–110. Springer (2012)
6. Bueno, L.F., Haeser, G., Martínez, J.M.: An inexact restoration approach to optimization problems with multiobjective constraints under weighted-sum scalarization. Optimization Letters **10**(6), 1315–1325 (2016)
7. Deb, K.: Multi-objective genetic algorithms: Problem difficulties and construction of test problems. Evolutionary Computation **7**(3), 205–230 (1999)
8. El-Bakry, A.S., Tapia, R.A., Tsuchiya, T., Zhang, Y.: On the formulation and theory of the newton interior-point method for nonlinear programming. Journal of Optimization Theory and Applications **89**(3), 507–541 (1996)
9. Fletcher, R., Leyffer, S.: Numerical experience with solving MPECs as NLPs. Tech. rep., University of Dundee Report NA 210 (2002)
10. Fletcher, R., Leyffer, S.: Solving mathematical programs with complementarity constraints as nonlinear programs. Optimization Methods and Software **19**(1), 15–40 (2004)
11. Fletcher, R., Leyffer, S., Ralph, D., Scholtes, S.: Local convergence of SQP methods for mathematical programs with equilibrium constraints. SIAM Journal on Optimization **17**(1), 259–286 (2006)
12. Fulop, J.: On the equivalence between a linear bilevel programming problem and linear optimization over the efficient set. Technical report WP93-1, Laboratory of Operations Research and Decision Systems, Computer and Automation Institute, Hungarian Academy of Sciences (1993)
13. Geoffrion, A.M.: Proper efficiency and the theory of vector maximization. Journal of Mathematical Analysis and Applications **22**(3), 618–630 (1968)
14. Guo, X.L., Li, S.J.: Optimality conditions for vector optimization problems with difference of convex maps. Journal of Optimization Theory and Applications **162**(3), 821–844 (2014)
15. HSL: A collection of fortran codes for large scale scientific computation. Available at http://www.hsl.rl.ac.uk/
16. Huband, S., Barone, L., While, L., Hingston, P.: A Scalable Multi-objective Test Problem Toolkit, pp. 280–295. Springer, Berlin, Heidelberg (2005)
17. Kuhn, H.W., Tucker, A.W.: Nonlinear programming. In: J. Neyman (ed.) Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 481–492. University of California Press (1951)
18. Martínez, J.M.: Inexact-restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. Journal of Optimization Theory and Applications **3**(1), 39–58 (2001)
19. Martínez, J.M., Pilotta, E.A.: Inexact-restoration algorithm for constrained optimization. Journal of Optimization Theory and Applications **104**(1), 135–163 (2000)
20. Martínez, J.M., Svaiter, B.F.: A practical optimality condition without constraint qualifications for nonlinear programming. Journal of Optimization Theory and Applications **118**(1), 117–133 (2003)
21. Miettinen, K.M.: Nonlinear Multiobjective Optimization. Kluwer Academic Publishers, Boston/London/Dordrecht (1999)
22. Pilotta, E.A., Torres, G.A.: An inexact restoration package for bilevel programming problems. Applied Mathematics **15**(10A), 1252–1259 (2012)
23. Srinivas, N., Deb, K.: Muiltiobjective optimization using nondominated sorting in genetic algorithms. Evolutionary Computation **2**(3), 221–248 (1994)
24. Tanaka, M., Watanabe, H., Furukawa, Y., Tanino, T.: GA-based decision support system for multicriteria optimization. In: IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century, vol. 2, pp. 1556–1561 (1995)
25. Zitzler, E., Deb, K., Thiele, L.: Comparison of multiobjective evolutionary algorithms: Empirical results. Evolutionary Computation **8**(2), 173–195 (2000)