A first-order regularized approach to the order-value optimization problem*

G. Q. Álvarez[†] E. G. Birgin[‡]

August 16, 2023

Abstract

Minimization of the order-value function is part of a large family of problems involving functions whose value is calculated by sorting values from a set or subset of other functions. The order-value function has as particular cases the minimum and maximum functions of a set of functions and is well suited for applications involving robust estimation. In this paper, a first order method with quadratic regularization to solve the problem of minimizing the order-value function is proposed. An optimality condition for the problem and theoretical results of iteration complexity and evaluation complexity for the proposed method are presented. The applicability of the problem and method to a parameter estimation problem with outliers is illustrated.

Keywords: Order-value optimization, regularized models, complexity, algorithms, applications.

Mathematics Subject Classification: 90C30, 65K05, 49M37, 90C60, 68Q25.

1 Introduction

Generalized order-value functions, systematized in [15], are functions whose value $f(x)$, for a given x in the domain, depends on order relations on a set of the form $\{f_i(x)\}_{i\in I}$. One such function is the order-value function of order p defined in [3, 4]. Given m functions $f_1,\dots,f_m$, the value of the pth order-value function f at a point x in the domain corresponds to the value at the pth position when the values $f_1(x), f_2(x),\dots,f_m(x)$ are ordered from smallest to largest. For the particular choices $p=1$ and $p=m$, we have that $f(x)=\min\{f_1(x), f_2(x),\dots,f_m(x)\}$ and $f(x)=\max\{f_1(x), f_2(x),\dots,f_m(x)\}$, respectively. It is important to note that, even if all f_i are differentiable, it is almost certain that f will be non-differentiable.

*This work has been partially supported by the Brazilian agencies FAPESP (grants 2013/07375-0, 2018/24293-0, and 2022/05803-3) and CNPq (grant 302682/2019-8).

†Department of Applied Mathematics, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil. e-mail: gdavid@ime.usp.br

‡Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil. e-mail: egbirgin@ime.usp.br

If x is a vector of portfolio positions and $f_i(x)$ represents the expected loss for choosing x under scenario i, then the order-value function is the discrete value-at-risk (VaR) function, which is widely used in risk analysis; see [14]. If each function f_i represents the precision with which a certain model that depends on unknown parameters x fits the ith observation, minimizing the pth order-value function is equivalent to fitting the model's x parameters by discarding the poorest fitted $m - p$ observations. In general, the order-value function is well suited for applications involving robust estimation, i.e., estimation techniques that are not affected by slight deviations in the data or from the idealized premises.

In [3], it was introduced a steepest descent type method for the minimization of the pth order-value function restricted to a closed and convex set. Convergence to points that satisfy a weak optimality conditions was proven. In [4], stronger optimality conditions and a nonlinear programming reformulation with equilibrium constraints of the problem were given. In [5], it was introduced a quasi-Newton method that generalizes the method proposed in [3]. In [6], it was proposed a global optimization strategy that combines multistart and a tunneling approach. In [17], it was proved that the minimization of the order-value function is an NP-hard problem in the strong sense in the case that constraints are given by a polytope.

In 2006, [16] introduced the idea of computational complexity in continuous optimization. Since then, algorithms with complexity results have been developed for a wide variety of continuous optimization problems. See, for example, [8, 9]. In 2022, the first book [12] specifically dedicated to the subject was released. This paper contributes to this line of research by proposing a method that possesses complexity results for the problem of minimizing the order-value function with box constraints.

The rest of this paper is organized as follows. In Section 2, the problem of minimizing the order-value function and the proposed regularized method are defined. Section 3 is devoted to the definition of an adequate optimality condition and to prove the well-definiteness, convergence, and complexity results of the method. Illustrative numerical experiments are given in Section 4. Conclusions and final remarks are given in the last section.

Notation. The symbol $\|\cdot\|$ denotes the Euclidean norm. For $i = 1, \dots, n$, $e^i \in \mathbb{R}^n$ denotes the ith column of the identity matrix in $\mathbb{R}^{n \times n}$.

2 Quadratically regularized first-order method

Let $f_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \dots, m$ be given. For a given $p \in \{1, 2, \dots, m\}$, the pth-order-value function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$f(x) \equiv f_{i_p(x)}(x), \tag{1}$$

where the indices $\{i_1(x), i_2(x), \dots, i_m(x)\} = \{1, 2, \dots, m\}$ are such that

$$f_{i_1(x)}(x) \leq f_{i_2(x)}(x) \leq \cdots \leq f_{i_m(x)}(x), \tag{2}$$

that is, f is such that $f(x)$ corresponds to the value $f_i(x)$ which, when the values $f_1(x), f_2(x), \dots,$ $f_m(x)$ are ordered from smallest to largest, is ranked in the pth position. In the present work, we consider the order-value optimization (OVO) problem given by

$$\text{Minimize } f(x) \text{ subject to } x \in \Omega, \tag{3}$$

where $\Omega \equiv \{x \in \mathbb{R}^n \mid \ell \le x \le u\}$, $\ell, u \in \mathbb{R}^n$ and $\ell_i < u_i$ for $i = 1, \dots, n$.

We introduce hereafter a first-order method to tackle problem (3) which, at each step, minimizes a quadratically regularized linear model of $f(x)$. The specification of the algorithm requires the following definitions. Given $\delta > 0$, for all $x \in \Omega$, we define

$$I(x, \delta) \equiv \{i \in \{1, 2, \dots, m\} \mid f(x) - \delta \le f_i(x) \le f(x) + \delta\}. \tag{4}$$

For further reference, we define, for all $x \in \Omega$, $I(x) \equiv I(x, 0)$. In addition to $\delta > 0$, for given $\sigma > 0$ and $\bar{x} \in \Omega$, we also define

$$\Psi(x; \bar{x}, \delta, \sigma) \equiv \max_{i \in I(\bar{x}, \delta)} \left\{\nabla f_i(\bar{x})^T (x - \bar{x})\right\} + \frac{\sigma}{2} \|x - \bar{x}\|^2. \tag{5}$$

The proposed method follows below.

Algorithm 2.1: Let $\delta > 0$, $\sigma_{\min} > 0$, $\alpha \in (0, 1)$, $\gamma > 1$, and $x^0 \in \Omega$ be given. Initialize $k \leftarrow 0$.

Step 1. Initialize $j \leftarrow 0$ and choose $\sigma_{k,j} \ge \sigma_{\min}$.

Step 2. Compute $x_{\text{trial}}^{k,j}$ as a solution to

$$\text{Minimize } \Psi(x; x^k, \delta, \sigma_{k,j}) \text{ subject to } x \in \Omega. \tag{6}$$

Step 3. Consider condition

$$f(x) \le f(x^k) - \alpha \|x - x^k\|^2. \tag{7}$$

If (7) with $x \equiv x_{\text{trial}}^{k,j}$ does not hold, then set $\sigma_{k,j+1} = \gamma \sigma_{k,j}$, update $j \leftarrow j + 1$, and go to Step 2.

Step 4. Define $x^{k+1} = x_{\text{trial}}^{k,j}$, $\sigma_k = \sigma_{k,j}$, $j_k = j$, update $k \leftarrow k + 1$ and go to Step 1.

3 Convergence and complexity

In this section, we introduce an optimality condition $C(\delta, \epsilon)$ for problem (3) that depends on the parameter δ and an optimality tolerance ϵ. In the sequence, we show that Algorithm 2.1 is well defined and present complexity results for obtaining an iterate that satisfies the optimality condition $C(\delta, \epsilon)$ for prescribed values of $\delta > 0$ and $\epsilon > 0$.

The three theorems that follow (Theorems 3.1, 3.2, and 3.3) show that if x^* is a local minimizer of (3), then it is also a local minimizer of minimizing $\Psi(x; x^*, 0, 0)$, $\Psi(x; x^*, 0, \sigma)$, and $\Psi(x; x^*, \delta, \sigma)$ subject to $x \in \Omega$ for any $\delta > 0$ and $\sigma > 0$, respectively. These results will be used in the construction of the optimality condition $C(\delta, \epsilon)$ for problem (3).

Assumption A1. *Functions $f_1, f_2, \dots, f_m$ are continuously differentiable for all $x \in \Omega$.*

Theorem 3.1. *Suppose that Assumption A1 holds. Let x^* be a local minimizer of (3) and consider the problem minimize $\Psi(x; x^*, 0, 0)$ subject to $x \in \Omega$, i.e.*

$$\textit{Minimize } \max_{i \in I(x^*)} \left\{\nabla f_i(x^*)^T (x - x^*)\right\} \textit{ subject to } x \in \Omega. \tag{8}$$

Then, x^* is a solution to (8).

Proof. Assume that the thesis is not true. Then there exists $x \in \Omega$ such that

$$\max_{i \in I(x^*)} \left\{\nabla f_i(x^*)^T (x - x^*)\right\} < \max_{i \in I(x^*)} \left\{\nabla f_i(x^*)^T (x^* - x^*)\right\} = 0.$$

This means that $\nabla f_i(x^*)^T (x - x^*) < 0$ for all $i \in I(x^*)$. By Assumption A1, f_i is differentiable for all $i \in I(x^*)$, then we have that

$$\lim_{t \to 0} \frac{f_i(x^* + t(x - x^*)) - f_i(x^*)}{t} = \nabla f_i(x^*)^T (x - x^*) < 0,$$

for all $i \in I(x^*)$. Therefore, there exists $\bar{t}_i > 0$ such that $f_i(x^* + t(x - x^*)) < f_i(x^*)$ for all $t \in (0, \bar{t}_i]$. Taking $\bar{t} = \min_{i \in I(x^*)}\{\bar{t}_i\}$, we obtain that

$$f_i(x^* + t(x - x^*)) < f_i(x^*) = f(x^*) \text{ for all } i \in I(x^*) \text{ and } t \in (0, \bar{t}]. \tag{9}$$

Moreover, for all $j \in \{1, 2, \ldots, m\} \setminus I(x^*)$, if t is sufficiently small, by the continuity of the functions $f_1, \ldots, f_m$, one has that

$$f_i(x^* + t(x - x^*)) < f_j(x^* + t(x - x^*)) \text{ whenever } i \in I(x^*) \text{ and } f(x^*) < f_j(x^*)$$

and

$$f_i(x^* + t(x - x^*)) > f_j(x^* + t(x - x^*)) \text{ whenever } i \in I(x^*) \text{ and } f(x^*) > f_j(x^*).$$

This implies that, for all t small enough, there exists $i \in I(x^*)$ such that

$$f(x^* + t(x - x^*)) = f_i(x^* + t(x - x^*)).$$

Therefore, for all t small enough, by (9), $f(x^* + t(x - x^*)) < f(x^*)$. Since x and x^* belong to Ω, which is convex, $x^* + t(x - x^*) \in \Omega$ for all $t \in [0, 1]$ and, in particular, for all t sufficiently small. Hence, x^* can not be a local minimizer of (3). □

Theorem 3.2. *Suppose that Assumption A1 holds. Let x^* be a local minimizer of (3), $\sigma \geq 0$, and consider the problem minimize $\Psi(x; x^*, 0, \sigma)$ subject to $x \in \Omega$, i.e.*

$$\textit{Minimize } \max_{i \in I(x^*)} \left\{\nabla f_i(x^*)^T (x - x^*)\right\} + \frac{\sigma}{2}\|x - x^*\|^2 \textit{ subject to } x \in \Omega. \tag{10}$$

Then, x^* is a solution to (10).

Proof. Assume that the thesis is not true. Then there exists $x \in \Omega$ such that

$$\max_{i \in I(x^*)} \left\{\nabla f_i(x^*)^T (x - x^*)\right\} + \frac{\sigma}{2}\|x - x^*\|^2 < \max_{i \in I(x^*)} \left\{\nabla f_i(x^*)^T (x^* - x^*)\right\} + \frac{\sigma}{2}\|x^* - x^*\|^2 = 0,$$

that is,

$$\max_{i \in I(x^*)} \left\{\nabla f_i(x^*)^T (x - x^*)\right\} < -\frac{\sigma}{2}\|x - x^*\|^2 \leq 0.$$

In other words, there exists $x \in \Omega$ such that $\Psi(x; x^*, 0, 0) < 0$. But this is impossible because, by the Theorem 3.1, x^* is a solution to (8) and $\Psi(x^*; x^*, 0, 0) = 0$. □

Theorem 3.3. *Suppose that Assumption A1 holds. Let x^* be a local minimizer of* (3)*, $\sigma \geq 0$, $\delta \geq 0$, and consider the problem minimize $\Psi(x; x^*, \delta, \sigma)$ subject to $x \in \Omega$, i.e.*

$$\textit{Minimize } \max_{i \in I(x^*,\delta)} \left\{ \nabla f_i(x^*)^T (x - x^*) \right\} + \frac{\sigma}{2} \|x - x^*\|^2 \textit{ subject to } x \in \Omega. \tag{11}$$

Then, x^* is a solution to (11)*.*

Proof. Assume that the thesis is not true. Then there exists $x \in \Omega$ such that

$$\max_{i \in I(x^*,\delta)} \left\{ \nabla f_i(x^*)^T (x - x^*) \right\} + \frac{\sigma}{2} \|x - x^*\|^2 < 0.$$

Therefore, for all $i \in I(x^*, \delta)$,

$$\nabla f_i(x^*)^T (x - x^*) + \frac{\sigma}{2} \|x - x^*\|^2 < 0,$$

which, since $I(x^*) \subseteq I(x^*, \delta)$, implies that

$$\nabla f_i(x^*)^T (x - x^*) + \frac{\sigma}{2} \|x - x^*\|^2 < 0,$$

for all $i \in I(x^*)$. But, by Theorem 3.2 this is impossible. □

The following theorem (Theorem 3.4), which requires some technical lemmas and assumptions, is the theorem that motivates the definition of the optimality condition $C(\delta, \epsilon)$ for problem (3).

Definition 3.1. *Let $c : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable and consider the feasible set $\mathcal{C} = \{z \in \mathbb{R}^n \mid c(z) \leq 0\}$. We say that $z \in \mathcal{C}$ verifies the Mangasarian-Fromovitz Constraint Qualification (MFCQ) if there exists $d \in \mathbb{R}^n$ such that*

$$\nabla c_i(z)^T d < 0 \textit{ for all } i \in \{1, \ldots, m\} \textit{ such that } c_i(z) = 0.$$

Definition 3.2. *Given $q \in \mathbb{N}$, we define the unit simplex $\sum_q \subset \mathbb{R}^q$ by*

$$\textstyle\sum_q = \left\{ \lambda \in \mathbb{R}^q \mid \sum_{j=1}^q \lambda_j = 1 \textit{ and } \lambda_j \geq 0 \textit{ for } j = 1, \ldots, q \right\}.$$

Lemma 3.1. *Consider the problem*

$$\textit{Minimize } \max_{i \in \mathcal{I}} \varphi_i(x) \textit{ subject to } x \in \Omega, \tag{12}$$

where $\mathcal{I} \subset \mathbb{N}$ is a finite set of indices and $\varphi_i : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable for all $i \in \mathcal{I}$. Assume that $x^* \in \Omega$ is a local minimizer of (12)*. Then, there exist $\mu \in \sum_{|\mathcal{I}|}$ and $\nu^\ell, \nu^u \in \mathbb{R}^n_+$ such that $\sum_{i \in \mathcal{I}} \mu_i \nabla \varphi_i(x^*) + \sum_{i=1}^n (\nu_i^u - \nu_i^\ell) e^i = 0$.*

Proof. Let x^* be a local minimizer of (12). Then, $(x^*, \max_{i\in\mathcal{I}} \varphi_i(x^*)) \in \mathbb{R}^{n+1}$ is the solution to

$$\text{Minimize } y \text{ subject to } \varphi_i(x) \le y \text{ for all } i \in \mathcal{I} \text{ and } x \in \Omega.$$

Let $d \in \mathbb{R}^{n+1}$ be given by

$$d_j = \begin{cases} -1, & \text{if } x_j^* = \ell_j, \\ 1, & \text{if } x_j^* = u_j, \\ 0, & \text{otherwise,} \end{cases}$$

for $j = 1, \dots, n$ plus

$$d_{n+1} = \max_{i\in\mathcal{I}} \left\{ [\nabla \varphi_i(x^*)]^T (d_1, \dots, d_n) \right\} + 1.$$

This d shows that $(x^*, \max_{i\in\mathcal{I}} \varphi_i(x^*))$ satisfies MFCQ. Thus, the thesis follows using the KKT conditions for the problem above. □

The corollary below will be used later in the complexity results.

Corollary 3.1. *Suppose that Assumption A1 holds. Then, for every k and $j = 0, \dots, j_k$, there exist $\mu \in \sum_{|I(x^k,\delta)|}$ and $\nu^\ell, \nu^u \in \mathbb{R}^n_+$ such that*

$$x_{\text{trial}}^{k,j} - x^k = \frac{1}{\sigma_{k,j}} \left[\sum_{i=1}^{n} \left(\nu_i^\ell - \nu_i^u \right) e^i - \sum_{i\in I(x^k,\delta)} \mu_i \nabla f_i(x^k) \right]. \tag{13}$$

Proof. The thesis follows from Lemma 3.1 considering $\varphi_i(x) \equiv \nabla f_i(x^k)^T(x - x^k) + (\sigma_{k,j}/2)\|x - x^k\|^2$. Assumption A1 is used to guarantee the existence of $\nabla f_i(x^k)$ for $i = 1, \dots, m$. □

Theorem 3.4. *Suppose that Assumption A1 holds. Assume that x^* is a local minimizer of* (3)*. Given $\delta \ge 0$, there exist $\mu \in \sum_{|I(x^*,\delta)|}$ and $\nu^\ell, \nu^u \in \mathbb{R}^n_+$ such that*

$$\sum_{i\in I(x^*,\delta)} \mu_i \nabla f_i(x^*) + \sum_{i=1}^{n} \left(\nu_i^u - \nu_i^\ell \right) e^i = 0. \tag{14}$$

Proof. Let x^* a local minimizer of (3) and $\delta \ge 0$. By Theorem 3.3, x^* is a solution to

$$\text{Minimize } \Psi(x; x^*, \delta, \sigma) \text{ subject to } x \in \Omega,$$

for any $\sigma > 0$. Then, by Lemma 3.1, there exist $\mu \in \sum_{|I(x^*,\delta)|}$ and $\nu^\ell, \nu^u \in \mathbb{R}^n_+$ such that

$$\sum_{i\in I(x^*,\delta)} \mu_i \nabla \left[\nabla f_i(x^*)^T(x - x^*) + \frac{\sigma}{2}\|x - x^*\|^2 \right] \Big|_{x=x^*} + \sum_{i=1}^{n} \left(\nu_i^u - \nu_i^\ell \right) e^i = 0,$$

from which (14) follows. □

Theorem 3.4 leads to the definition of the following approximate necessary optimality condition.

Definition 3.3. *We say that x satisfies the approximate optimality condition $C(\delta, \epsilon)$ if there exist $\mu \in \sum_{|I(x,\delta)|}$ and $\nu^\ell, \nu^u \in \mathbb{R}^n_+$ such that*

$$\left\| \sum_{i \in I(x,\delta)} \mu_i \nabla f_i(x) + \sum_{i=1}^{n} \left(\nu_i^u - \nu_i^\ell \right) e^i \right\| \le \epsilon. \tag{15}$$

From here to the end of the section, we are devoted to show an upper bound for the cost of Algorithm 2.1, in terms of iterations and function evaluations, to, given $\epsilon > 0$, find an iterate x^k that satisfies the approximate optimality condition $C(\delta, \epsilon)$. We will also show that Algorithm 2.1 is well defined in the sense that the inner loop defined by Steps 2 and 3 terminates in a finite number of steps that does not depend on either k or ϵ. A few assumptions and technical lemmas precede the main results.

Assumption A2. *For all k and $j = 0, \ldots, j_k$, the associated Lagrange multipliers ν^ℓ and ν^u of Corollary 3.1 are bounded by a constant c_ν which depends neither on k nor on j.*

Note that MFCQ guarantees that, for every k and $j \in \{0, \ldots, j_k\}$, the associated Lagrange multipliers ν^ℓ and ν^u of Corollary 3.1 are bounded by a constant; while Assumption A2 says that there exists a constant for all k and $j \in \{0, \ldots, j_k\}$ which depends neither on k nor on j.

Assumption A3. $\|\nabla f_1(x)\|, \|\nabla f_2(x)\|, \ldots, \|\nabla f_m(x)\|$ *are bounded from above by a constant c_∇ for all $x \in \Omega$.*

Lemma 3.2. *Suppose that Assumptions A1, A2, and A3 hold. Then, for all k and $j = 0, \ldots, j_k$, there exist $c_x > 0$, which depends neither on k nor on j, such that*

$$\|x^{k,j}_{\rm trial} - x^k\| \le c_x / \sigma_{k,j}. \tag{16}$$

Proof. By Corollary 3.1 and Assumption A3 we have that

$$\|x^{k,j}_{\rm trial} - x^k\| \le \frac{1}{\sigma_{k,j}} \left[\sum_{i=1}^{n} \left| \nu_i^\ell - \nu_i^u \right| + c_\nabla \right].$$

By Assumption A2, (16) holds with $c_x = 2\, n\, c_\nu + c_\nabla$. □

Assumption A4. *All the gradients ∇f_i satisfy a Lipschitz condition, that is, there exists $L > 0$ such that, for $i = 1, \ldots, m$ and all $x, y \in \Omega$,*

$$\|\nabla f_i(y) - \nabla f_i(x)\| \le L \|y - x\|. \tag{17}$$

As a consequence of Assumption A4, for $i = 1, \ldots, m$ and all $x, y \in \Omega$,

$$\left| f_i(y) - \left[f_i(x) + \nabla f_i(x)^T (y - x) \right] \right| \le \frac{L}{2} \|y - x\|^2. \tag{18}$$

In particular, for $i = 1, \ldots, m$ and all $x, y \in \Omega$,

$$f_i(y) \le f_i(x) + \nabla f_i(x)^T (y - x) + \frac{L}{2} \|y - x\|^2. \tag{19}$$

See, for example, [9].

Lemma 3.3. *([3, Lemma 2.1]) Let* $a_1, \dots, a_r \in \mathbb{R}$, $b_1, \dots, b_r \in \mathbb{R}$, $\beta > 0$, *and* $\{i_1, \dots, i_r\} = \{1, \dots, r\}$ *be such that* $a_1 \le \cdots \le a_r$, $b_j \le a_j - \beta$ *for* $j = 1, \dots, r$, *and* $b_{i_1} \le \cdots \le b_{i_r}$. *Then,* $b_{i_q} \le a_q - \beta$ *for* $q = 1, \dots, r$.

Proof. By hypothesis, we have that, for any $q \in \{1, \dots, r\}$,

$$\begin{array}{rcl} b_{i_q} & \le & a_{i_q} - \beta \\ b_{i_q} & \le & b_{i_{q+1}} \le a_{i_{q+1}} - \beta \\ & \vdots & \\ b_{i_q} & \le & b_{i_{q+1}} \le \cdots \le b_{i_r} \le a_{i_r} - \beta. \end{array}$$

Therefore, $b_{i_q} \le \min\{a_{i_q}, a_{i_{q+1}}, \dots, a_{i_r}\} - \beta$. Since the set $\{a_{i_q}, a_{i_{q+1}}, \dots, a_{i_r}\}$ has $r - q + 1$ elements, then there exist $\tilde{q} \in \{1, \dots, q\}$ such that $a_{\tilde{q}} \in \{a_{i_q}, a_{i_{q+1}}, \dots, a_{i_r}\}$. Thus,

$$b_{i_q} \le a_{\tilde{q}} - \beta \le a_q - \beta,$$

as we wanted to prove. □

Lemma 3.4. *Suppose that Assumptions A1 and A4 hold. For every* k *and* $j = 0, \dots, j_k$, *if* $\sigma_{k,j} \ge L + 2\alpha$, *then*

$$f_i(x^{k,j}_{\text{trial}}) \le f_i(x^k) - \alpha \|x^{k,j}_{\text{trial}} - x^k\|^2, \tag{20}$$

for all $i \in I(x^k, \delta)$.

Proof. By (19), which is implied by Assumption A4, if $i \in I(x^k, \delta)$, then

$$\begin{array}{rcl} f_i(x^{k,j}_{\text{trial}}) & \le & f_i(x^k) + \nabla f_i(x^k)^T (x^{k,j}_{\text{trial}} - x^k) + \dfrac{L}{2} \|x^{k,j}_{\text{trial}} - x^k\|^2 \\ & = & f_i(x^k) + \nabla f_i(x^k)^T (x^{k,j}_{\text{trial}} - x^k) + \dfrac{\sigma_{k,j}}{2} \|x^{k,j}_{\text{trial}} - x^k\|^2 - \dfrac{\sigma_{k,j}}{2} \|x^{k,j}_{\text{trial}} - x^k\|^2 + \dfrac{L}{2} \|x^{k,j}_{\text{trial}} - x^k\|^2. \end{array}$$

But the objective function of (6), defined in (5), vanishes if $x = x^k$. Therefore,

$$\nabla f_i(x^k)^T (x^{k,j}_{\text{trial}} - x^k) + \frac{\sigma_{k,j}}{2} \|x^{k,j}_{\text{trial}} - x^k\|^2 \le 0 \text{ for all } i \in I(x^k, \delta).$$

Thus,

$$f_i(x^{k,j}_{\text{trial}}) \le f_i(x^k) - \frac{\sigma_{k,j}}{2} \|x^{k,j}_{\text{trial}} - x^k\|^2 + \frac{L}{2} \|x^{k,j}_{\text{trial}} - x^k\|^2 \text{ for all } i \in I(x^k, \delta).$$

So, if $\sigma_{k,j} \ge L + 2\alpha$, then we have that

$$f_i(x^{k,j}_{\text{trial}}) \le f_i(x^k) - \alpha \|x^{k,j}_{\text{trial}} - x^k\|^2 \text{ for all } i \in I(x^k, \delta).$$

□

The next theorem, together with the fact that, for all k, the initial value of the regularization parameter is equal to $\sigma_{\min} > 0$ and, whenever a new value is calculated, its value is multiplied by $\gamma > 1$, is what shows that the loop defined by Steps 2 and 3 is executed a finite number of times per iteration.

Theorem 3.5. *Suppose that Assumptions A1, A3 and A4 hold. Then, for all k and $j = 0, \ldots, j_k$, if*

$$\sigma_{k,j} \geq \max\left\{L + 2\alpha, 9c_x^2/(2\delta)\right\} \tag{21}$$

then $x_{\text{trial}}^{k,j}$ satisfies (7).

Proof. Assume that $\sigma_{k,j} \geq L + 2\,\alpha$. By the Cauchy-Schwarz inequality, Assumptions A3 and A4, and (16) in Lemma 3.2,

$$\left| f_i(x_{\text{trial}}^{k,j}) - f_i(x^k) \right| \leq \frac{c_x^2}{\sigma_{k,j}} + \frac{L\,c_x^2}{2\,\sigma_{k,j}^2} \text{ for } i = 1, \ldots, m.$$

Note that $L \leq \sigma_{k,j}$. Then,

$$\left| f_i(x_{\text{trial}}^{k,j}) - f_i(x^k) \right| \leq \frac{3\,c_x^2}{2\,\sigma_{k,j}} \text{ for } i = 1, \ldots, m.$$

Therefore, if $\sigma_{k,j} \geq \max\{L + 2\,\alpha, 9\,c_x^2/(2\,\delta)\}$, then we have that

$$\left| f_i(x_{\text{trial}}^{k,j}) - f_i(x^k) \right| \leq \frac{\delta}{3} \text{ for } i = 1, \ldots, m. \tag{22}$$

Thus, for $j = 1, \ldots, p$,

$$f_{i_j(x^k)}(x_{\text{trial}}^{k,j}) \leq f_{i_j(x^k)}(x^k) + \frac{\delta}{3} \leq f(x^k) + \frac{\delta}{3}$$

and, for $j = p, \ldots, m$,

$$f_{i_j(x^k)}(x_{\text{trial}}^{k,j}) \geq f_{i_j(x^k)}(x^k) - \frac{\delta}{3} \geq f(x^k) - \frac{\delta}{3}.$$

This means that p elements of the set $\{f_1(x_{\text{trial}}^{k,j}), f_2(x_{\text{trial}}^{k,j}), \ldots, f_m(x_{\text{trial}}^{k,j})\}$ are less than or equal to $f(x^k) + \delta/3$ and that $m - p + 1$ elements of that set are greater than or equal to $f(x^k) - \delta/3$. Then, at least one element satisfies both inequalities and, as a consequence, $f_{i_p(x_{\text{trial}}^{k,j})}(x_{\text{trial}}^{k,j}) = f(x_{\text{trial}}^{k,j})$ satisfies both inequalities, i.e.

$$f(x^k) - \frac{\delta}{3} \leq f(x_{\text{trial}}^{k,j}) \leq f(x^k) + \frac{\delta}{3}. \tag{23}$$

By (23) and the definition of $I(\cdot, \cdot)$ in (4), $i_p(x_{\text{trial}}^{k,j}) \in I(x^k, \delta)$. Let us write

$$I(x^k, \delta) = \{i_1, \ldots, i_r\} = \{i'_1, \ldots, i'_r\},$$

where

$$f_{i_1}(x^k) \leq \cdots \leq f_{i_r}(x^k) \text{ and } f_{i'_1}(x_{\text{trial}}^{k,j}) \leq \cdots \leq f_{i'_r}(x_{\text{trial}}^{k,j}).$$

Let j be such that $f_j(x^k) < f(x^k) - \delta$. Then, by (22), $f_j(x_{\text{trial}}^{k,j}) \leq f_j(x^k) + \delta/3 < f(x^k) - 2\delta/3 < f(x^k)$. This means that the indices $j \notin I(x^k, \delta)$ such that $f_j(x^k) < f(x^k)$ are the same as the indices $j \notin I(x^k, \delta)$ such that $f_j(x^{k,j}_{\text{trial}}) < f(x^k)$. Analogously, the indices $j \notin I(x^k, \delta)$ such that $f_j(x^k) > f(x^k)$ are the same as the indices $j \notin I(x^k, \delta)$ such that $f_j(x^{k,j}_{\text{trial}}) > f(x^k)$. Therefore, if $q \in \{1, \ldots, r\}$ is such that $i_p(x^k) = i_q$, then $i_p(x^{k,j}_{\text{trial}}) = i'_q$.

By Lemma 3.4,

$$f_{i_j}(x^{k,j}_{\text{trial}}) \leq f_{i_j}(x^k) - \alpha \|x^{k,j}_{\text{trial}} - x^k\|^2$$

for $j = 1, \ldots, r$. Therefore, by Lemma 3.3 taking $\beta = \alpha \|x^{k,j}_{\text{trial}} - x^k\|^2$, $a_j = f_{i_j}(x^k)$ and $b_j = f_{i_j}(x^{k,j}_{\text{trial}})$ for $j = 1, \ldots, r$, we have that

$$f_{i'_j}(x^{k,j}_{\text{trial}}) \leq f_{i_j}(x^k) - \alpha \|x^{k,j}_{\text{trial}} - x^k\|^2$$

for $j = 1, \ldots, r$. In particular, it holds for the index $q \in \{1, \ldots, r\}$ of the previous paragraph such that $i_p(x^k) = i_q$ and $i_p(x^{k,j}_{\text{trial}}) = i'_q$. Therefore,

$$f(x^{k,j}_{\text{trial}}) \leq f(x^k) - \alpha \|x^{k,j}_{\text{trial}} - x^k\|^2$$

as we wanted to prove. □

The theorem below shows that Algorithm 2.1 requires $O(\delta^{-2}\epsilon^{-2})$ iterations and $O(|\log(\delta)|)$ functional evaluations per iteration to find a point that satisfies the $C(\delta, \epsilon)$ optimality condition of problem (3).

Theorem 3.6. *Suppose that Assumptions A1, A3 and A4 hold and there exists $f_{\text{low}} \in \mathbb{R}$ such that $f(x) \geq f_{\text{low}}$ for all $x \in \Omega$. Then, σ_k is such that*

$$\sigma_k \leq \gamma \max\{L + 2\alpha, 9c_x^2/(2\delta)\}, \tag{24}$$

where c_x is a constant that depends on c_ν and c_∇, and at most

$$\left\lfloor 1 + \log_\gamma \left(\frac{\sigma_k}{\sigma_{\min}} \right) \right\rfloor \tag{25}$$

functional evaluations are done to get (7). Moreover, the number of iterations k at which $C(\delta, \epsilon)$ is not satisfied by x^k is bounded above by

$$\left\lfloor \left(\frac{\gamma^2 \max\{L + 2\alpha, 9c_x^2/(2\delta)\}^2}{\alpha} \right) \left(\frac{f(x^0) - f_{\text{low}}}{\epsilon^2} \right) \right\rfloor. \tag{26}$$

Proof. Applying Theorem 3.5, (24) and (25) follow from (21) and the fact that, at Step 3, Algorithm 2.1 updates the regularization parameter by multiplying its value by γ if (7) does not hold.

For the second part, let $K \subset \mathbb{N}$ be the set of indices k such that $C(\delta, \epsilon)$ is not satisfied by x^{k+1}. By the mechanism of Algorithm 2.1, $x^{k+1} = x^{k,j}_{\text{trial}}$, where $x^{k,j}_{\text{trial}}$ is a solution to (6) and satisfies (7). Then, on the one hand, by Corollary 3.1, for each $k \in K$, there exist $\mu \in \sum_{|I(x^k,\delta)|}$ and $\nu^\ell, \nu^u \in \mathbb{R}^n_+$ such that

$$x^{k+1} - x^k = \frac{1}{\sigma_k} \left[\sum_{i=1}^{n} \left(\nu_i^\ell - \nu_i^u \right) e^i - \sum_{i \in I(x^k,\delta)} \mu_i \nabla f_i(x^k) \right],$$

i.e.

$$\|x^{k+1} - x^k\| = \frac{1}{\sigma_k} \left\| \sum_{i \in I(x^k,\delta)} \mu_i \nabla f_i(x^k) + \sum_{i=1}^n \left(\nu_i^u - \nu_i^\ell\right) e^i \right\|,$$

and, by (24) and the fact that $C(\delta,\epsilon)$ does not hold at x^{k+1}, it holds

$$\|x^{k+1} - x^k\| \geq \frac{\epsilon}{\gamma \max\{L + 2\alpha, 9\, c_x^2/(2\delta)\}}.$$

On the other hand, since for each $k \in K$, x^{k+1} satisfies (7), we have that

$$f(x^{k+1}) \leq f(x^k) - \alpha \left(\frac{\epsilon}{\gamma \max\{L + 2\alpha, 9\, c_x^2/(2\delta)\}}\right)^2.$$

Summing for all $k \in K$,

$$\sum_{k \in K} \left(f(x^k) - f(x^{k+1})\right) \geq |K|\alpha \left(\frac{\epsilon}{\gamma \max\{L + 2\alpha, 9\, c_x^2/(2\delta)\}}\right)^2.$$

Since $f(x) \geq f_{\rm low}$ for all $x \in \mathbb{R}^n$,

$$f(x^0) - f_{\rm low} \geq |K|\alpha \left(\frac{\epsilon}{\gamma \max\{L + 2\alpha, 9\, c_x^2/(2\delta)\}}\right)^2,$$

from which (26) follows. □

Algorithm 2.1 defines at each iteration k a regularization parameter $\sigma_k \geq \sigma_{\min} > 0$, i.e., bounded away from zero. Specifically, the first trial $\sigma_{k,0}$ is an arbitrary value not smaller than $\sigma_{\min}$ that is then successively multiplied by γ. In practice, it may be adequate the first trial $\sigma_{k,0}$ at iteration $k > 1$ to be a fraction of σ_{k-1}. In this case, each $\sigma_k \geq \sigma_{\min}^k$, but it may be the case that $\sigma_{\min}^k \to 0$. For such a modified version of Algorithm 2.1, with a slightly different analysis than the one performed in Theorem 3.6, similar complexity bounds can also be obtained; see [7, §4].

4 Numerical illustration

In this section we intend to illustrate how the OVO problem, and in particular the method proposed to solve it, can be used to fit an epidemiological model in the case in which observations contain outliers.

Algorithm 2.1 was implemented in Fortran. As suggested in Lemma 3.1, subproblem (6) of Step 2 is reformulated as

$$\text{Minimize } y \text{ subject to } \nabla f_i(x^k)^T(x - x^k) + \frac{1}{\sigma_{k,j}}\|x - x^k\|^2 \leq y \text{ for all } i \in I(x^k,\delta) \text{ and } x \in \Omega. \quad (27)$$

Problem (27) is a smooth nonlinear programming problem and we chose to solve it with Algencan. Algencan [2, 10, 11] is a safeguarded augmented Langrangian method introduced in [1, 2].

Its convergence theory, properties and usage are described in detail in [10]. Complexity results and an extensive numerical comparison with another state-of-the-art method for nonlinear programming can be found in [11]. In this work we use Algencan with all its default parameters.

Codes were implemented in Fortran 90. Tests were conducted on a computer with a 3.9 GHz AMD Ryzen 5 5600G processor and 32GB 3200 MHz DDR3 RAM memory, running Windows 11 Pro and a Windows Subsystem for Linux with Debian GNU/Linux 11. Code was compiled by the GNU Fortran compiler (version 10.2.1) with the -O3 optimization directive enabled.

The considered epidemiological model was developed in [13] with the purpose of modeling a serological data set of 8870 people before the introduction of measles, mumps and rubella vaccine in United Kingdom. The model aims to describe the rate at which susceptible individuals acquire infection by the diseases mentioned above at different ages. The data in Table 1, taken from [13], show the estimated proportion of seropositive in the unvaccinated segment of the sample divided into 29 age groups.

Age group	Proportion seropositive			Age group	Proportion seropositive		
(years)	Measles	Mumps	Rubella	(years)	Measles	Mumps	Rubella
$[1, 2)$	0.207	0.115	0.126	$[17, 19)$	0.898	0.895	0.869
$[2, 3)$	0.301	0.147	0.171	$[19, 21)$	0.959	0.911	0.844
$[3, 4)$	0.409	0.389	0.184	$[21, 23)$	0.957	0.920	0.852
$[4, 5)$	0.589	0.516	0.286	$[23, 25)$	0.937	0.915	0.907
$[5, 6)$	0.757	0.669	0.400	$[25, 27)$	0.918	0.950	0.935
$[6, 7)$	0.669	0.768	0.503	$[27, 29)$	0.939	0.909	0.921
$[7, 8)$	0.797	0.786	0.524	$[29, 31)$	0.967	0.873	0.896
$[8, 9)$	0.818	0.798	0.634	$[31, 33)$	0.973	0.880	0.890
$[9, 10)$	0.866	0.878	0.742	$[33, 35)$	0.943	0.915	0.949
$[10, 11)$	0.859	0.861	0.664	$[35, 40)$	0.967	0.906	0.899
$[11, 12)$	0.908	0.844	0.735	$[40, 45)$	0.946	0.933	0.955
$[12, 13)$	0.923	0.881	0.815	$[45, 55)$	0.961	0.917	0.937
$[13, 14)$	0.889	0.895	0.768	$[55, 65)$	0.968	0.898	0.933
$[14, 15)$	0.936	0.882	0.842	$[65, +\infty)$	0.968	0.839	0.917
$[15, 17)$	0.889	0.869	0.760				

Table 1: Proportion of seropositive for measles, mumps and rubella by age group.

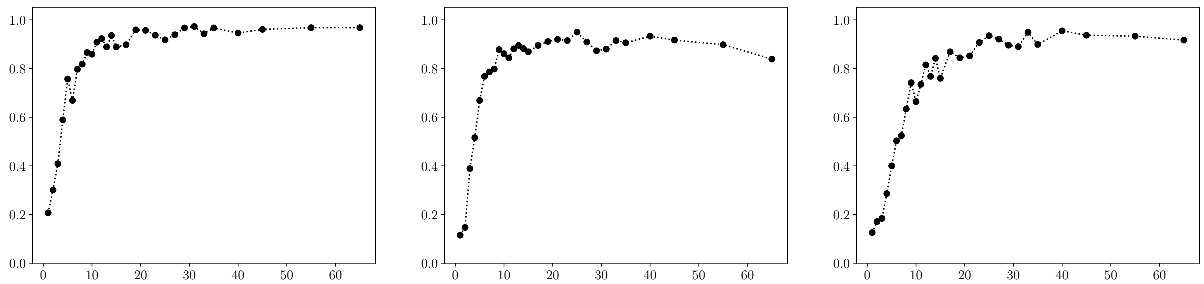
The model we wish to fit to the data in Table 1 is given by

$$y(t, x) = 1 - \exp\left\{ \frac{x_1}{x_2} t e^{-x_2 t} + \frac{x_1}{x_2} \left(\frac{x_1}{x_2} - x_3 \right) \left(e^{-x_2 t} - 1 \right) - x_3 t \right\}, \tag{28}$$

where x_1, x_2, x_3 are non-negative unknown parameters. The amount of data is $m = 29$, and we wish to estimate the parameters x_1, x_2, x_3 of model (28) for each of the three diseases separately. That is, we consider three independent problems. To transform the model parameter fitting problem into an OVO type problem, we define

$$f_i(x) = \frac{1}{2} \left(y(t_i, x) - y_i \right)^2,$$

for $i = 1, \ldots, m$, where t_i represents the left limit of an age range $[t_{\min}, t_{\max})$ and y_i represents the corresponding observation. (Considering $t_i = (t_{\min} + t_{\max})/2$ would also be another valid alternative.) Figure 1 shows a graphical representation of the data in Table 1, with the definition of t_i mentioned above. Since x_1, x_2, x_3 are non-negative, we define $\Omega = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid (x_1, x_2, x_3) \geq 0\}$.



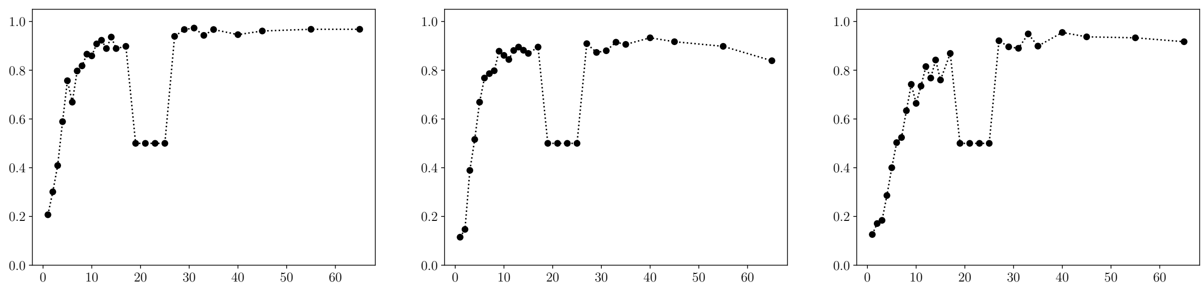
(a) measles (b) mumps (c) rubella

Figure 1: Observed proportion of seropositive for the three considered diseases.

In Algorithm 2.1, we considered $\delta = 5 \times 10^{-4}$, $\sigma_{\min} = 0.1$, $\alpha = 10^{-8}$, and $\gamma = 5$. As stopping criterion, we checked the satisfaction of the optimality condition $C(\delta, \epsilon)$ in-between Steps 3 and 4. It is worth noticing that, when solving the reformulation of subproblem (6) with Algencan as mentioned above, Algencan returns, besides the solution $x_{\text{trial}}^{k,j}$ sought, estimations of the associated Lagrange multipliers μ, ν^ℓ, and ν^u required to check $C(\delta, \epsilon)$. For the stopping criterion, we considered $\epsilon = 10^{-4}$. The ϵ tolerance value is standard when using first order methods. The value of the parameters $\sigma_{\min}$, α, and γ is quite standard in the literature of methods using regularized models and the method is not very sensitive to variations in these parameters. The choice of δ is more difficult. It is dimensional, problem dependent, and was chosen by trial and error. As initial guess $x^0 \in \Omega$, we considered the solution reported in [13] obtained by applying linear least squares, namely, $x^0 = (0.197, 0.287, 0.021)$ for measles, $x^0 = (0.156, 0.250, 0.000)$ for mumps, and $x^0 = (0.063, 0.178, 0.020)$ for rubella.

To illustrate the result of tackling a parameter fitting problem in the presence of outliers using the OVO approach, we contaminated the observations of the age groups $[19, 21)$, $[21, 23)$, $[23, 25)$, and $[25, 27)$, replacing the corresponding observation with 0.5. The modified observations are shown in Figure 2. Assuming that the number of outliers is unknown, we solved the OVO problem (3) with $p = m - o$ and $o \in \{1, 2, \ldots, 10\}$, where o represents the presumed number of outliers in the data. Table 2 and Figure 3 show the results. The table shows, for each value of o, the optimal value of the OVO function (column $f(x^*)$) and, as a measure of Algorithm 2.1 performance, the number of iterations (column "#it"), the number of functional evaluations (column "#fcnt"), and the CPU time in seconds (column "Time") that were necessary to meet the stopping criterion. The figures in the table show that the optimal value of the objective function of the OVO problem is on the order of 10^{-3} when $o \in \{1, 2, 3\}$ and drops by an order of magnitude when $o \geq 4$. This shows that this approach might be used to automatically detect the number of outliers contained in the data. The numbers in the table also show that problems where the number of outliers is underestimated are much more difficult to solve. In the figure, some of the curves appear overlapped, but as expected the models whose parameters were fitted considering $1 \leq o \leq 3$ fail to reproduce the observed data.

Figure 4 shows, on the left, the models adjusted when considering $o \in \{4, 5, 6\}$. It is not entirely clear that the model found by considering $o = 4$ is "the best"; and comparing the optimal values $f(x^*)$ obtained in the three cases does not help to decide, since it is natural that the more observations are left out, the better (smaller) is the optimal value found. That suggests that, assuming model (28) is "correct", there are already outliers in the observed data available in [13]. Figure 4 shows on the right side the fitted models considering $o = 10$. In these plots, the observations that the optimal solution of the OVO problem points out as outliers are highlighted in red. It is clear that choosing these observations manually would be practically impossible.

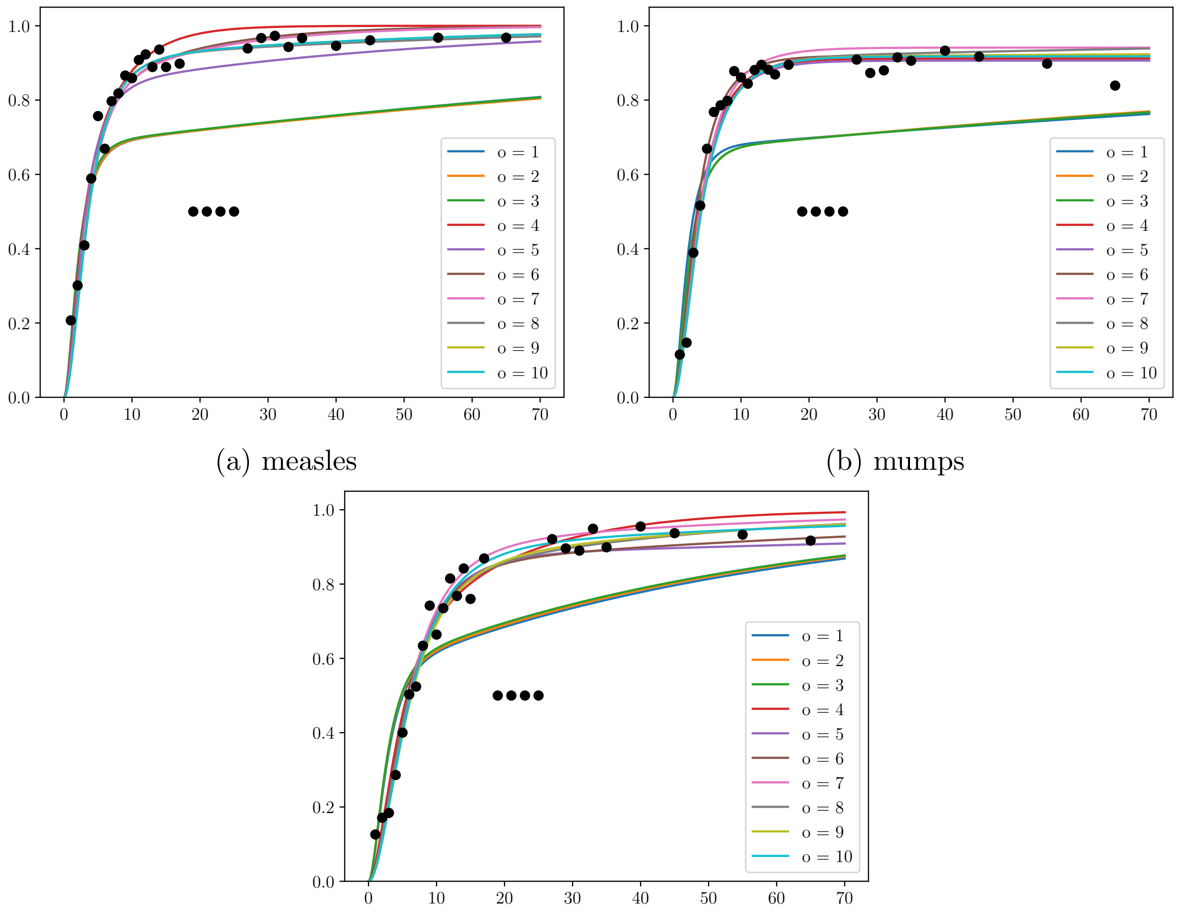


(a) measles (b) mumps (c) rubella

Figure 2: Observed proportion of seropositive for the three considered diseases after the inclusion of outliers.

o	measles				mumps				rubella			
	$f(x^*)$	#it	#fcnt	Time	$f(x^*)$	#it	#fcnt	Time	$f(x^*)$	#it	#fcnt	Time
1	2.664e−02	259	459	0.13	2.149e−02	110	197	0.03	2.034e−02	413	885	0.34
2	2.650e−02	256	457	0.05	2.100e−02	104	174	0.03	1.933e−02	359	754	0.23
3	2.605e−02	219	354	0.06	2.088e−02	107	160	0.03	1.817e−02	372	746	0.21
4	3.193e−03	27	55	0.00	2.982e−03	5	19	0.00	3.165e−03	26	80	0.01
5	2.755e−03	12	31	0.00	2.481e−03	3	9	0.00	1.914e−03	6	22	0.00
6	1.199e−03	6	8	0.00	1.060e−03	6	14	0.00	1.782e−03	4	25	0.00
7	9.910e−04	7	18	0.00	1.981e−03	4	22	0.00	1.725e−03	2	6	0.00
8	5.496e−04	4	16	0.00	1.660e−03	5	6	0.00	1.300e−03	6	18	0.00
9	4.549e−04	1	10	0.00	1.491e−03	2	10	0.00	1.061e−03	3	6	0.00
10	3.620e−04	1	10	0.00	1.300e−03	1	2	0.00	6.666e−04	2	5	0.00

Table 2: Details of applying Algorithm 2.1 for solving the OVO problem with $p = m - o$ and $o \in \{1, 2, \ldots, 10\}$.



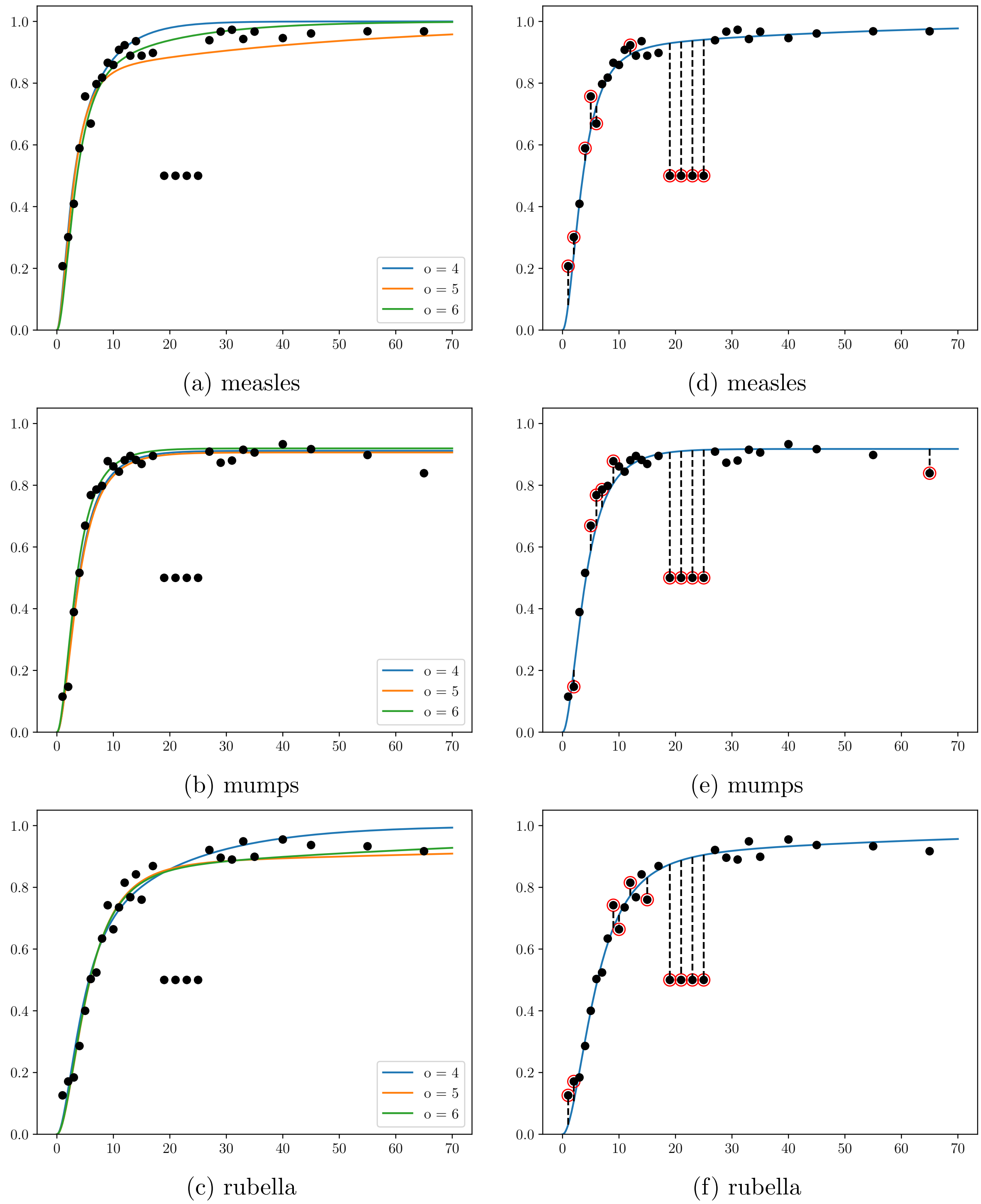
(a) measles (b) mumps (c) rubella

Figure 3: Models adjusted by solving the OVO problem with $p = m - o$ and $o \in \{1, 2, \ldots, 10\}$.

5 Final remarks

In this paper we introduced a method for the problem of minimizing the order-value function with box constraints. The method is of first order and uses quadratic regularization. As lines of future work we can mention the development of methods for problems with more general constraints and methods using higher order models. More generally, proposing methods with complexity results for other problems of the GOVO family is also a possible line of future work.

Disclosure statement: The authors report there are no competing interests to declare.



(a) measles (d) measles

(b) mumps (e) mumps

(c) rubella (f) rubella

Figure 4: On the left side, the models fitted with $o \in \{4, 5, 6\}$. On the right side, the models fitted with $o = 10$, highlighting the observations that the optimal solution to the OVO problem points to as outliers.

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification, *Mathematical Programming* 111, pp. 5–32, 2008.

[2] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, On Augmented Lagrangian methods with general lower-level constraints, *SIAM Journal on Optimization* 18, pp. 1286–1309, 2008.

[3] R. Andreani, C. Dunder and J. M. Martínez, Order-Value Optimization: Formulation and solution by means of a primal Cauchy method, *Mathematical Methods of Operations Research* 58, pp. 387–399, 2003.

[4] R. Andreani, C. Dunder and J. M. Martínez, Nonlinear-programming reformulation of the order-value optimization problem, *Mathematical Methods of Operations Research* 61, pp. 365–384, 2005.

[5] R. Andreani, J. M. Martínez, M. Salvatierra, and F. S. Yano, Quasi-Newton methods for Order-Value Optimization and Value-at-Risk calculations, *Pacific Journal of Optimization* 2, pp. 11–33, 2006.

[6] R. Andreani, J. M. Martínez, M. Salvatierra, and F. S. Yano, Global order-value optimization by means of a multistart harmonic oscillator tunneling strategy, in *Global Optimization: From Theory to Implementation*, L. Liberti and N. Maculan (eds.), Springer, Boston, MA, 2006, pp. 379–404.

[7] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, and S. A. Santos, On the use of third-order models with fourth-order regularization for unconstrained optimization, *Optimization Letters* 14, pp. 815–838, 2020.

[8] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos, and Ph. L. Toint, Evaluation complexity for nonlinear constrained optimization using unscaled KKT conditions and high-order models, *SIAM Journal on Optimization* 26, pp. 951–967, 2016.

[9] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos, and Ph. L. Toint, Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models, *Mathematical Programming* 163, pp. 359–368, 2017.

[10] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, Society for Industrial and Applied Mathematics, Philadelphia, 2014.

[11] E. G. Birgin and J. M. Martínez, Complexity and performance of an Augmented Lagrangian algorithm, *Optimization Methods and Software* 35, pp. 885–920, 2020.

[12] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *Evaluation Complexity of Algorithms for Nonconvex Optimization: Theory, Computation, and Perspectives*, SIAM, Philadelphia, PA, 2022.

[13] C. P. Farrington, Modelling forces of infection for measles, mumps and rubella, *Statistics in Medicine* 9, pp. 953–967, 1990.

[14] P. Jorion, *Value at Risk: The new benchmark for managing financial risk*, 3rd. ed., McGraw-Hill, 2009.

[15] J. M. Martínez, Generalized Order-Value Optimization, *TOP* 20, pp. 75–98, 2012.

[16] Y. Nesterov and B. T. Polyak, Cubic regularization of Newton method and its global performance, *Mathematical Programming* 108, 177–205, 2006.

[17] Z. Jiang, Q. Hu, and X. Zheng, Optimality condition and complexity of order-value optimization problems and low order-value optimization problems, *Journal of Global Optimization* 69, pp. 511–523, 2017.