Journal of Convex Analysis
Volume 32 (2025), No. 1, 199–226
https://doi.org/10.68381/jca32013

Splitting Algorithms for Distributionally Robust Optimization

Luis Briceño-Arias
Departamento de Matemática, Universidad Técnica Federico Santa María, Santiago, Chile
luis.briceno@usm.cl

Sergio López-Rivera
Departamento de Ingeniería Matemática, Universidad de Chile, Santiago, Chile
sergio.lopez@dim.uchile.cl

Emilio Vilches
Instituto de Ciencias de la Ingeniería, Universidad de O'Higgins, Rancagua, Chile.
emilio.vilches@uoh.cl

Dedicated to Ralph Tyrrell Rockafellar on the occasion of his 90th birthday.

Received: October 16, 2024
Accepted: October 21, 2024

We propose different splitting methods for solving distributionally robust optimization problems in cases where the uncertainties are described by discrete distributions. The first method involves computing the proximity operator of the supremum function that appears in the optimization problem. The second method solves an equivalent monotone inclusion formulation derived from the first-order optimality conditions, where the resolvents of the monotone operators involved in the inclusion are computable. The proposed methods are applied to solve the Couette inverse problem with uncertainty and the denoising problem with uncertainty. We present numerical results to compare the efficiency of the algorithms.

Keywords: Distributionally robust optimization, supremum function, proximal mapping, splitting algorithms.

2020 Mathematics Subject Classification: 47H05, 65K05, 90C15, 90C17, 90C25, 90C47.

1. Introduction

Stochastic optimization models decision problems that involve uncertainty or randomness. The goal is to find solutions that work well under different conditions by optimizing an objective function that accounts for the uncertainty, often focusing on expected results or reducing risks (see, e.g., [14]). However, this approach usually assumes full knowledge of the probability distributions, which may not always be practical in real situations. To address this situation, the theoretical framework of Distributionally Robust Optimization (DRO) has been introduced, which assumes that the true probability distribution belongs to a set of distributions known as the *ambiguity set*. Optimization is then performed based on the worst-case scenario within this ambiguity set. By considering different ambiguity sets, DRO encompasses Robust Optimization when the ambiguity set includes all possible distributions, and Stochastic Optimization when the ambiguity set consists of a single

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

distribution. Hence, the DRO framework is well-suited for addressing problems with partial information, promoting distributed robustness in decision-making. We refer to [14, 15, 16] for more details.

Let $(\Omega, \mathcal{A})$ be a measurable space, and let $\mathcal{P}$ be a nonempty, closed, and convex subset of probability measures defined on $(\Omega, \mathcal{A})$, supported on $\Xi \subset \mathcal{G}$, where $\mathcal{G}$ is a real Hilbert space. In this paper, we aim to study splitting methods for solving the following distributionally robust optimization problem:

$$\min_{x\in\mathcal{Q}} \Big\{ h(x) + \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[F(x,\xi)\right] \Big\}, \tag{1}$$

where $\mathcal{Q}$ is a nonempty closed convex subset of a real Hilbert space $\mathcal{H}$ and $h\colon \mathcal{H} \to \mathbb{R}$ is a convex and differentiable function with β^{-1}-Lipschitz gradient for some $\beta > 0$, which represents a cost associated to the decision variable x. Besides, $\xi\colon \Omega \to \Xi$ is a random vector, and $F\colon \mathcal{H} \times \Xi \to \mathbb{R}$ is the random cost function, with $F(\cdot, \xi)$ being proper, lower semicontinuous, and convex. The set $\mathcal{P}$ is called the *ambiguity set* and accounts for the level of knowledge about the probability model of the problem. This set can be constructed based on empirical statistical information or beliefs about the moments of the distribution itself (see [15]). Moreover, when the ambiguity set $\mathcal{P}$ is a singleton, the DRO model (1) reduces to a stochastic optimization problem. Conversely, if $\mathcal{P}$ represents the set of all probability measures defined on $(\Omega, \mathcal{A})$ and supported on Ξ, then problem (1) becomes a robust optimization problem.

It is well-known that both robust and stochastic optimization have certain limitations. First, robust optimization can be very conservative, as it ignores valuable probabilistic information, while stochastic optimization may require too much information about the probability distribution, which may not be available to the modeler. Second, the solutions provided by robust or stochastic optimization models may perform poorly in out-of-sample tests or may have intrinsic bias that cannot be eliminated simply by increasing the size of the sampled data. Third, robust models can be computationally difficult to solve. Additionally, stochastic programs may involve high-dimensional integration, which is also generally intractable. In this context, the DRO problem provides a potent modeling framework and addresses some of the disadvantages of both stochastic and robust optimization. We refer to [14] for a thorough discussion.

One approach to tackle the problem (1) is to use the techniques proposed in [15, Section 4.2] or in [14, Section 7.4], that is, formulate the dual program of the inner maximization problem:

$$\max_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[\overline{F}(x,\xi)] := \mathbb{E}_{\mathbb{P}}[h(x) + F(x,\xi)]. \tag{2}$$

Let us illustrate this technique considering $\mathcal{G} = \mathbb{R}^m$ and the following ambiguity set [15, eq. 4.2.1]

$$\mathcal{P}(\Xi, \boldsymbol{\mu}^+, \boldsymbol{\mu}^-) = \{\mathbb{P} \in \mathcal{M}_+(\mathcal{A}) \,:\, \mathbb{P}(\xi \in \Xi) = 1,\ \boldsymbol{\mu}^- \leq \mathbb{E}_{\mathbb{P}}[\xi] \leq \boldsymbol{\mu}^+\},$$

where $\mathcal{M}_+(\mathcal{A})$ is the set of all the measures defined on $\mathcal{A}$, $\boldsymbol{\mu}^+ \in \mathbb{R}^m$, and $\boldsymbol{\mu}^- \in \mathbb{R}^m$.

Denote by $\mathcal{M}_+(\mathbb{R}^m)$ the set of all the measures induced by the random vector ξ. That is, $\mu \in \mathcal{M}_+(\mathbb{R}^m)$ if $\mu(A) = \mathbb{P}(\xi \in A)$, for all $A \subset \mathbb{R}^m$ Borel measurable set, where $\mathbb{P} \in \mathcal{M}_+(\mathcal{A})$. Then, the problem (2) is equivalent to the following maximization problem

$$\max_{\mu\in\mathcal{M}_+(\mathbb{R}^m)} \int_\Xi \overline{F}(x,\xi)d\mu(\xi) \tag{3}$$
$$\text{s.t.} \int_\Xi d\mu(\xi) = 1, \quad \int_\Xi \xi d\mu(\xi) \leq \boldsymbol{\mu}^+, \quad \int_\Xi -\xi d\mu(\xi) \leq -\boldsymbol{\mu}^-,$$

whose dual problem (3) is given by

$$\min_{\lambda\in\mathbb{R},\ \boldsymbol{\beta}\geq 0,\ \boldsymbol{\gamma}\geq 0} \lambda + \boldsymbol{\beta}^\top\boldsymbol{\mu}^+ - \boldsymbol{\gamma}^\top\boldsymbol{\mu}^- \tag{4}$$
$$\text{s.t. } \overline{F}(x,\xi) - \lambda + (\boldsymbol{\gamma}-\boldsymbol{\beta})^\top\xi \leq 0, \quad \forall \xi \in \Xi.$$

If the strong duality hold between the primal problem (3) and its dual (4) (for example, if the Slater conditions holds for (3)), we can replace the inner maximization problem with its dual in the original DRO problem (1). Thus, the problem (1) is equivalent to the following static robust optimization problem

$$\min_{x\in\mathcal{H},\ \lambda\in\mathbb{R},\ \boldsymbol{\beta}\geq 0,\ \boldsymbol{\gamma}\geq 0} \lambda + \boldsymbol{\beta}^\top\boldsymbol{\mu}^+ - \boldsymbol{\gamma}^\top\boldsymbol{\mu}^- \tag{5}$$
$$\text{s.t. } \overline{F}(x,\xi) - \lambda + (\boldsymbol{\gamma}-\boldsymbol{\beta})^\top\xi \leq 0, \quad \forall \xi \in \Xi,$$
$$x \in \mathcal{Q}.$$

The difficulty of the program (5) depends on the structure of the function $\overline{F}(x,\xi)$ as well as of the set $\mathcal{Q}$. We focus on the discrete case, that is, when the uncertainty set Ξ associated to ξ is finite: $\Xi = \{\xi_1,\dots,\xi_N\} \subset \mathcal{G}$ with $N \geq 1$. For every $\mathbb{P} \in \mathcal{P}$ and $i \in \{1,\dots,N\}$, we denote $p_i := \mathbb{P}(\{\omega \in \Omega : \xi(\omega) = \xi_i\})$. Then, the set of all the probability measures is given by the following set (the probability simplex)

$$\Delta_N := \Big\{p \in \mathbb{R}^N_+ : \sum_{i=1}^N p_i = 1\Big\}$$

and $\mathbb{E}_\mathbb{P}[F(x,\xi)] = \sum_{i=1}^N p_i F(x,\xi_i)$. Hence, the problem (1) becomes

$$\min_{x\in\mathcal{Q}} \Big\{h(x) + \sup_{p\in\mathcal{P}} \sum_{i=1}^N p_i f_i(x)\Big\}, \tag{6}$$

where $f_i := F(\cdot,\xi_i)$ for all $i \in \{1,\dots,N\}$. As far as we know, for the above problem, there are no available splitting algorithms in the distributionally robust literature. Hence, the aim of this paper is twofold. First, to propose a new algorithm for solving the following generalization of problem (1) in the discrete case:

$$\min_{\mathbf{x}=(x_1,\dots,x_N)\in\mathcal{H}^N} \Big\{H(\mathbf{x}) + \sup_{p\in\mathcal{P}} \sum_{i=1}^N p_i f_i(x_i)\Big\} \quad \text{s.t.} \quad \mathbf{x} \in V \cap \bigtimes_{i=1}^N \mathcal{Q}_i, \tag{7}$$

where $H\colon \mathcal{H}^N \to \mathbb{R}$ is a convex differentiable function with β^{-1}-Lipschitz gradient, $\mathcal{Q}_i \subset \mathcal{H}$ is a nonempty closed convex subset, $V \subset \mathcal{H}^N$ is a closed vector subspace,

and $\mathcal{P} \subset \Delta_N$ is a nonempty closed convex set. Second, to propose different methods to solve the problem (6).

On the one hand, we observe that whenever $V = \mathcal{D} := \{\mathbf{x} \in \mathcal{H}^N \,:\, x_1 = \cdots = x_N\}$, $H(\mathbf{x}) = h(x_1)$, and $\mathcal{Q}_i = \mathcal{Q}$ for all $i \in \{1, \ldots, N\}$, we have that $(x, \ldots, x) \in \mathcal{D}$ is solution to problem (7) if and only if $x \in \mathcal{H}$ is a solution to (6). On the other hand, when $\mathcal{H} = \mathbb{R}^n$, V is the nonanticipativity set, and $H = 0$, problem (7) reduces to the problem considered in [10, Example 4]. In order to solve problem (7), we consider the following functions $(f_i)_{i=1}^N$ and ambiguity sets $\mathcal{P}$. For every $i \in \{1, \ldots, N\}$,

(i) $f_i(x) = \langle a_i, x\rangle + \xi_i$, where $a_i \in \mathcal{H}\backslash\{0\}$ and $\xi_i \in \mathbb{R}$.
  (a) $\mathcal{P} = \Delta_N$.
  (b) $\mathcal{P} = \{p \in \Delta_N : p \le q\}$, where $q \in \mathbb{R}^N$ is such that $\operatorname{int}(\mathbb{R}^N_+) \cap \operatorname{int}(\mathcal{P}_2) \cap \mathcal{P}_1 \neq \emptyset$ with $\mathcal{P}_1 = \{p \in \mathbb{R}^N \,:\, \sum_{i=1}^N p_i = 1\}$ and $\mathcal{P}_2 = \{p \in \mathbb{R}^N \,:\, p \le q\}$.
  (c) $\mathcal{P} = \{p \in \Delta_N \,:\, \mu_- \le \sum_{i=1}^N p_i\xi_i \le \mu_+\}$, where $\mu_- \in \mathbb{R}$, and $\mu_+ \in \mathbb{R}$ are such that $\operatorname{int}(\mathbb{R}^N_+) \cap \operatorname{int}(\mathcal{P}_3) \cap \mathcal{P}_1 \neq \emptyset$ with $\mathcal{P}_3 = \{p \in \mathbb{R}^N \,:\, \mu_- \le \sum_{i=1}^N p_i\xi_i \le \mu_+\}$.

(ii) $f_i(x) = \|x - \xi_i\|^2$, where $\xi_i \in \mathcal{H}$, and $\mathcal{P} = \Delta_N$.

Remark 1.1. Note that in the context of problem (6), the case when we have $f_i(x) = \langle x, Qx\rangle + \langle b_i, x\rangle + c_i$, where $Q\colon \mathcal{H} \to \mathcal{H}$ is a symmetric positive semidefinite operator, $b_i \in \mathcal{H}$, and $c_i \in \mathbb{R}$, reduces to the case (i). Indeed, since $\mathcal{P} \subset \Delta_N$, we have in this case that

$$h(x) + \sup_{p\in\mathcal{P}} \sum_{i=1}^N p_i f_i(x) = \underbrace{h(x) + \langle x, Qx\rangle}_{\widetilde{h}(x)} + \sup_{p\in\mathcal{P}} \sum_{i=1}^N p_i(\langle b_i, x\rangle + c_i),$$

where $\widetilde{h}$ is convex and differentiable with Lipschitz gradient. □

The case (i)(b) covers the case when the ambiguity set is associated to the Conditional Value-at-Risk, which is

$$\mathcal{P}_{\mathrm{CVaR}_\alpha} := \Big\{p \in \Delta_N \,:\, p \le \frac{\overline{p}}{1-\alpha}\Big\}, \tag{8}$$

where $\overline{p} \in \mathbb{R}^N$ is a probability vector and $\alpha \in]0, 1[$. The ambiguity set in (8) appears in [10] and is the domain of the conjugate of the well-known Conditional Value-at-Risk, which is a coherent-risk measure. On the other hand, the case (i)(c) is motivated by the ambiguity set defined in [15, eq. 4.2.1].

In the case (i), problem (7) is

$$\min_{\mathbf{x}\in\mathcal{H}^N} \Big\{H(\mathbf{x}) + \sup_{p\in\mathcal{P}} \sum_{i=1}^N p_i(\langle a_i, x_i\rangle + \xi_i)\Big\} \quad \text{s.t.} \quad \mathbf{x} \in V \cap \bigtimes_{i=1}^N \mathcal{Q}_i. \tag{9}$$

In the case (ii), problem (7) reduces to

$$\min_{\mathbf{x}\in\mathcal{H}^N} \Big\{H(\mathbf{x}) + \max_{1\le i\le N} \|x_i - \xi_i\|^2\Big\} \quad \text{s.t.} \quad \mathbf{x} \in V \cap \bigtimes_{i=1}^N \mathcal{Q}_i. \tag{10}$$

On the other hand, in the case (i), problem (6) is equivalent to

$$\min_{x\in\mathcal{Q}} \Big\{ h(x) + \sup_{p\in\mathcal{P}} \sum_{i=1}^{N} p_i(\langle a_i, x\rangle + \xi_i) \Big\}. \tag{11}$$

Notation: Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|$. For every $\lambda > 0$, the *Moreau envelope* of $f\colon \mathcal{H} \to \mathbb{R}\cup\{+\infty\}$ of parameter λ is denoted by

$$x \in \mathcal{H} \mapsto e_\lambda f(x) := \inf_{u\in\mathcal{H}} \Big\{ f(u) + \frac{1}{2\lambda}\|x-u\|^2 \Big\}$$

and the *proximity operator* is denoted by

$$x \in \mathcal{H} \mapsto P_\lambda f(x) = \arg\min_{y\in\mathcal{H}} \Big\{ f(y) + \frac{1}{2\lambda}\|x-y\|^2 \Big\}.$$

Moreover, the *projection* on a set $\mathcal{Q} \subset \mathcal{H}$ is denoted by

$$x \in \mathcal{H} \mapsto \mathrm{proj}_{\mathcal{Q}}(x) = \arg\min_{y\in\mathcal{Q}} \|x-y\|.$$

The *indicator function* of a set $\mathcal{Q}$ is denoted by $\iota_{\mathcal{Q}}$, which is equal to zero in $\mathcal{Q}$ and $+\infty$ otherwise. On the other hand, the *normal cone* to a closed convex set $\mathcal{Q}$ is denoted by $N_{\mathcal{Q}}$ and the *interior* of $\mathcal{Q}$ is denoted by $\mathrm{int}(\mathcal{Q})$. Moreover, the set of the functions $f\colon \mathcal{H} \to \mathbb{R}\cup\{+\infty\}$ that are proper, lower semicontinuous, and convex is denoted by $\Gamma_0(\mathcal{H})$. The *convex subdifferential* of a function $f \in \Gamma_0(\mathcal{H})$ is denoted by ∂f. In addition, the *domain* of f is denoted by $\mathrm{dom}\, f = \{x \in \mathcal{H} : f(x) < +\infty\}$. We have that if $f \in \Gamma_0(\mathcal{H})$, then $P_\lambda f(x)$ is unique. In particular, if $\mathcal{Q}$ is a nonempty closed convex set, then $\mathrm{proj}_{\mathcal{Q}}(x)$ is unique.

The *conjugate* of a function $f \in \Gamma_0(\mathcal{H})$ is denoted by f^* and the *resolvent* of a maximally monotone operator A is denoted by J_A. Now, the class of bounded linear operators from $\mathcal{H}$ to a real Hilbert space $\mathcal{G}$ is denoted by $\mathcal{L}(\mathcal{H},\mathcal{G})$ and if $\mathcal{H} = \mathcal{G}$ this class is denoted by $\mathcal{L}(\mathcal{H})$. Given $L \in \mathcal{L}(\mathcal{H},\mathcal{G})$, its adjoint operator is denoted by $L^* \in \mathcal{L}(\mathcal{G},\mathcal{H})$. Finally, the expected value of a random variable $X\colon \Omega \to \Xi$ with respect to a probability measure $\mathbb{P}$ is denoted by $\mathbb{E}_{\mathbb{P}}[X] = \int_\Omega X(\omega)d\mathbb{P}(\omega)$.

The remainder of this paper is organized as follows. In Section 2, we compute the proximity operator of certain supremum functions. In some cases, we provide an algorithm that converges to the proximity operator, while in a particular case we provide a closed-form for the proximity operator. Next, in Section 3, we will show the different algorithms for solving the problems described in the introduction. In Section 4, we provide some applications that can be written as a particular case of the main problem. In Section 5, we illustrate the numerical experiments of the proposed algorithms. Finally, conclusions are detailed in Section 6.

2. Proximity operator of a supremum function

Consider the following function

$$\mathbf{x} \in \mathcal{H}^N \mapsto f(\mathbf{x}) := \sup_{p\in\mathcal{P}} \sum_{i=1}^{N} p_i f_i(x_i). \tag{12}$$

In the following result, we compute the proximity operator of f in the cases (i) and (ii). We will see that in the case (ii), the proximity of f has a closed-form.

Proposition 2.1. *Let $\{f_i\}_{i=1}^N \subset \Gamma_0(\mathcal{H})$ and let $\mathcal{P} \subset \Delta_N$ be a nonempty closed convex set. Let $\mathbf{x} \in \mathcal{H}^N$ and let $\lambda > 0$. Then the following hold.*

(i) *If for every $i \in \{1,\dots,N\}$, $f_i(x) = \langle a_i, x\rangle + \xi_i$ with $a_i \in \mathcal{H}\backslash\{0\}$ and $\xi_i \in \mathbb{R}$, then*

$$P_\lambda f(\mathbf{x}) = (x_i - \lambda \overline{p}_i a_i)_{i=1}^N,$$

where $\overline{p} \in \mathcal{P}$ is the unique solution to

$$\min_{p\in\mathcal{P}} \frac{1}{2} p^\top D p - p^\top d, \tag{13}$$

with $D := \operatorname{diag}(\lambda\|a_1\|^2,\dots,\lambda\|a_N\|^2)$ and $d := (\langle a_i, x_i\rangle + \xi_i)_{i=1}^N$.

(ii) *If $\mathcal{P} = \Delta_N$ and for every $i \in \{1,\dots,N\}$, $f_i(x) = \|x - \xi_i\|^2$ with $\xi_i \in \mathcal{H}$, then*

$$P_\lambda f(\mathbf{x}) = \left(\frac{x_i + 2\lambda\overline{p}_i\xi_i}{2\lambda\overline{p}_i + 1}\right)_{i=1}^N,$$

where $\overline{p} \in \Delta_N$ is a solution to

$$\max_{p\in\Delta_N} \ell(p) := \sum_{i=1}^N \left(\frac{p_i}{1 + 2\lambda p_i}\right)\alpha_i, \tag{14}$$

and, for every $i \in \{1,\dots,N\}$, $\alpha_i = \|x_i - \xi_i\|^2$.

Proof. Let us define the function $\mathbf{x} \in \mathcal{H}^N \mapsto g_p(\mathbf{x}) = \sum_{i=1}^N p_i f_i(x_i)$ for all $p \in \mathcal{P}$. Note that, since $f_i \in \Gamma_0(\mathcal{H})$ for all $i \in \{1,\dots,N\}$, then $\{g_p\}_{p\in\mathcal{P}} \subset \Gamma_0(\mathcal{H}^N)$. Moreover, the function $p \mapsto g_p(\mathbf{x})$ is concave and upper semicontinuous for all $\mathbf{x} \in \mathcal{H}^N$. In addition, since f_i is proper for all $i \in \{1,\dots,N\}$ and $\mathcal{P} \subset \Delta_N$, then $f = \sup_{p\in\mathcal{P}} g_p$ is proper. Furthermore, $\mathcal{P}$ is a nonempty compact convex set. Then, by [13, Theorem 3.5], we have that

$$P_\lambda f(\mathbf{x}) = P_\lambda g_{\overline{p}}(\mathbf{x}), \quad \text{with } \overline{p} \in \operatorname*{arg\,max}_{p\in\mathcal{P}} e_\lambda g_p(\mathbf{x}) \tag{15}$$

Case (i): By [2, Proposition 24.8(i) & Proposition 24.11], we deduce that

$$P_\lambda g_{\overline{p}}(\mathbf{x}) = (x_i - \lambda\overline{p}_i a_i)_{i=1}^N. \tag{16}$$

Let us compute $e_\lambda g_p(\mathbf{x})$. By (16), we obtain that

$$\begin{aligned}
e_\lambda g_p(\mathbf{x}) &= g_p(P_\lambda g_p(\mathbf{x})) + \frac{1}{2\lambda}\|\mathbf{x} - P_\lambda g_p(\mathbf{x})\|^2 \\
&= \sum_{i=1}^N p_i(\langle a_i, x_i - \lambda p_i a_i\rangle + \xi_i) + \frac{1}{2\lambda}\sum_{i=1}^N \|x_i - x_i + \lambda p_i a_i\|^2 \\
&= \sum_{i=1}^N p_i(\langle a_i, x_i\rangle + \xi_i) - \lambda\sum_{i=1}^N p_i^2\|a_i\|^2 + \frac{1}{2\lambda}\sum_{i=1}^N \lambda^2 p_i^2\|a_i\|^2 \\
&= \sum_{i=1}^N p_i(\langle a_i, x_i\rangle + \xi_i) - \frac{\lambda}{2}\sum_{i=1}^N p_i^2\|a_i\|^2.
\end{aligned}$$

Set $\alpha_i = \lambda\|a_i\|^2$ and $d_i = \langle a_i, x_i\rangle + \xi_i$ for all $i \in \{1,\dots,N\}$. Then, in order to find $\overline{p}$ in (15), we need to solve the following problem

$$\max_{p\in\mathcal{P}} \sum_{i=1}^{N} p_i d_i - \frac{1}{2}\sum_{i=1}^{N} p_i^2\alpha_i,$$

which is equivalent to

$$\min_{p\in\mathcal{P}} \psi(p) := \frac{1}{2}p^\top Dp - p^\top d, \tag{17}$$

where $D = \operatorname{diag}(\alpha_1,\dots,\alpha_N) \in \mathbb{R}^{N\times N}_{++}$ and $d = (d_i)_{i=1}^N \in \mathbb{R}^N$. Note that, since $\nabla^2\psi(p) = D$ is positive definite, ψ is strictly convex and coercive. Thus, from [2, Proposition 11.15(i)], we have that the problem (17) has a unique solution.

Case (ii): By [2, Proposition 24.11 & Proposition 24.8(i)], we obtain that

$$P_\lambda g_{\overline{p}}(\mathbf{x}) = \Big(\frac{1}{2\lambda\overline{p}_i+1}(x_i + 2\lambda\overline{p}_i\xi_i)\Big)_{i=1}^N. \tag{18}$$

Let us calculate $e_\lambda g_p(\mathbf{x})$. By (18), we have that

$$\begin{aligned}
e_\lambda g_p(\mathbf{x}) &= g_p(P_\lambda g_p(\mathbf{x})) + \frac{1}{2\lambda}\|\mathbf{x} - P_\lambda g_p(\mathbf{x})\|^2\\
&= \sum_{i=1}^{N} p_i\left\|\frac{x_i+2\lambda p_i\xi_i}{1+2\lambda p_i} - \xi_i\right\|^2 + \frac{1}{2\lambda}\sum_{i=1}^{N}\left\|\frac{x_i+2\lambda p_i\xi_i}{1+2\lambda p_i} - x_i\right\|^2\\
&= \sum_{i=1}^{N} p_i\left\|\frac{x_i-\xi_i}{1+2\lambda p_i}\right\|^2 + \frac{1}{2\lambda}\sum_{i=1}^{N}\left\|\frac{2\lambda p_i\xi_i - 2\lambda p_i x_i}{1+2\lambda p_i}\right\|^2\\
&= \sum_{i=1}^{N}\frac{p_i\|x_i-\xi_i\|^2}{(1+2\lambda p_i)^2} + \frac{(2\lambda p_i)^2\|\xi_i-x_i\|^2}{2\lambda(1+2\lambda p_i)^2}\\
&= \sum_{i=1}^{N}\frac{p_i(1+2\lambda p_i)}{(1+2\lambda p_i)^2}\|x_i-\xi_i\|^2 = \sum_{i=1}^{N}\Big(\frac{p_i}{1+2\lambda p_i}\Big)\|x_i-\xi_i\|^2.
\end{aligned}$$

Set $\alpha_i = \|x_i - \xi_i\|^2$ for all $i \in \{1,\dots,N\}$. Then, in order to find $\overline{p}$ in (15), we need to solve in this case the following problem

$$\max_{p\in\Delta_N} \ell(p) := \sum_{i=1}^{N}\Big(\frac{p_i}{1+2\lambda p_i}\Big)\alpha_i. \tag{19}$$

Note that ℓ is continuous and Δ_N is a nonempty compact set. Therefore, the problem (19) has solutions. □

Remark 2.2. In the case when $\mathcal{P} = \Delta_N$ (case (i)(a)) or $\mathcal{P} = \{p \in \Delta_N : p \le q\}$ (case (i)(b)), the problem in (13) can be solved by the method proposed in [8].

The following result provides a method for solving (13) when

$$\mathcal{P} = \{p \in \Delta_N : \mu_- \le \textstyle\sum_{i=1}^N p_i\xi_i \le \mu_+\}$$

(case (i)(c)). Recall that $\mathcal{P}_3 = \{p \in \mathbb{R}^N : \mu_- \le \sum_{i=1}^N p_i\xi_i \le \mu_+\}$.

Proposition 2.3. *In the context of problem* (13)*, let* $R := \sqrt{D}$*, let* $x_0 = R^{-1}d$*, and let* $p_0 = q_0 = 0$*. For every* $k \in \mathbb{N}$*, we consider the following routine*

$$\left\lfloor \begin{aligned} y_k &= \operatorname{proj}_{R\mathcal{P}_3}(x_k + p_k) \\ p_{k+1} &= x_k + p_k - y_k \\ x_{k+1} &= \operatorname{proj}_{R\Delta_N}(y_k + q_k) \\ q_{k+1} &= y_k + q_k - x_{k+1}. \end{aligned} \right. \tag{20}$$

Then the sequence (x_k) *converges to a point* q *and* $p = R^{-1}q$ *is the solution to* (13) *with* $\mathcal{P} = \{p \in \Delta_N \,:\, \mu_- \leq \sum_{i=1}^N p_i\xi_i \leq \mu_+\}$.

Proof. Note that the problem (13) with $\mathcal{P} = \{p \in \Delta_N \,:\, \mu_- \leq \sum_{i=1}^N p_i\xi_i \leq \mu_+\}$ is equivalent to

$$\min_{p\in\Delta_N\cap\mathcal{P}_3} \psi(p) := \frac{1}{2}p^\top Dp - p^\top d. \tag{21}$$

Let us consider the variable change $q = Rp$. Then, since $R = \sqrt{D}$ and $D \in \mathbb{R}^{N\times N}_{++}$ is a diagonal matrix, we have

$$\begin{aligned} \psi(p) &= \frac{1}{2}(R^{-1}q)^\top DR^{-1}q - (R^{-1}q)^\top d = \frac{1}{2}q^\top R^{-1}DR^{-1}q - q^\top R^{-1}d \\ &= \frac{1}{2}q^\top q - q^\top R^{-1}d = \frac{1}{2}\|q - R^{-1}d\|^2 - \frac{1}{2}\|R^{-1}d\|^2. \end{aligned}$$

Thus, since R is invertible, problem (21) is equivalent to find

$$\overline{q} = \operatorname{proj}_{R\Delta_N\cap R\mathcal{P}_3}(R^{-1}d). \tag{22}$$

From [7, Proposition 5.3], the sequence (x_k) generated by the algorithm in (20) (Dykstra's projection algorithm) converges to the projection $\overline{q}$ in (22). Therefore, $p = R^{-1}\overline{q}$ is the solution to (21). □

Remark 2.4. In the context of problem (13), denote $D = \operatorname{diag}(\alpha_1, \ldots, \alpha_N)$. Note that $R\Delta_N = \{q \in \mathbb{R}^N_+ \,:\, \sum_{i=1}^N \frac{q_i}{\sqrt{\alpha_i}} = 1\}$ and

$$R\mathcal{P}_3 = \{q \in \mathbb{R}^N \,:\, \mu_- \leq \sum_{i=1}^N q_i\frac{\xi_i}{\sqrt{\alpha_i}} \leq \mu_+\}.$$

Then, the projection on $R\Delta_N$ can be calculated using the algorithm proposed in [8] and the projection on $R\mathcal{P}_3$ can be calculated through of [2, Example 29.21]. □

Let us now consider a method for solving the problem (14), where $\alpha_i \geq 0$ for all $i \in \{1, \ldots, N\}$. Note that if $\alpha_i = 0$ for all $i \in \{1, \ldots, N\}$, then every $p \in \Delta_N$ is a solution to (14). Thus, we can assume that $\max\limits_{1\leq i\leq N} \alpha_i > 0$. On the other hand, note that for every $p \in \Delta_N$, we have

$$(\nabla^2\ell(p))_{ij} = \begin{cases} -4\lambda\alpha_i(1 + 2\lambda p_i)^{-3} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad \text{for all } i, j \in \{1, \ldots, N\}.$$

Then, $\nabla^2(-\ell)(p)$ is a positive semidefinite matrix for all $p \in \Delta_N$ and thus $-\ell$ is convex on Δ_N. The following result provides an explicit solution to problem (14).

Proposition 2.5. *In the context of problem* (14)*, let* $\{\ell_i\}_{i=1}^N$ *be such that we have* $\alpha_{\ell_1} \leq \cdots \leq \alpha_{\ell_N}$ *with* $\alpha_{\ell_N} > 0$ *and let* $I_i := \{\ell_1, \ldots, \ell_i\}$ *for all* $i \in \{1, \ldots, N\}$.

$$\text{Define} \qquad k := \min\Big\{i \in \{0, \ldots, N-1\} \,:\, (N-i+2\lambda)\sqrt{\alpha_{\ell_{i+1}}} > \sum_{j \notin I_i} \sqrt{\alpha_j}\Big\}, \tag{23}$$

where $I_0 := \emptyset$*. Then* $\overline{p} \in \mathbb{R}^N$ *defined by*

$$(\forall i \in \{1, \ldots, N\}) \quad \overline{p}_i = \begin{cases} 0, & \text{if } i \in I_k; \\ \dfrac{1}{2\lambda}\left[\dfrac{(N-k+2\lambda)\sqrt{\alpha_i}}{\sum_{j \notin I_k}\sqrt{\alpha_j}} - 1\right], & \text{if } i \notin I_k \end{cases} \tag{24}$$

is a solution to problem (14)*.*

Proof. First of all, the set in (23) is nonempty since $N-1$ is in that set. Indeed, $(1+2\lambda)\sqrt{\alpha_{\ell_N}} > \sqrt{\alpha_{\ell_N}}$. Hence k is well defined. Note that

$$\nabla \ell(p) = \Big(\frac{\alpha_i}{(1+2\lambda p_i)^2}\Big)_{i=1}^N \text{ for all } p \in \mathbb{R}^N.$$

Then, since $-\ell$ is convex on Δ_N, by the definition of Δ_N and the KKT's conditions, it follows that it is enough to prove that there exists $\tau \in \mathbb{R}$ and $(\mu_i)_{i=1}^N \in \mathbb{R}_+^N$ such that

$$\frac{-\alpha_i}{(1+2\lambda p_i)^2} + \tau - \mu_i = 0, \quad \mu_i p_i = 0, \quad p_i \geq 0 \quad \text{for all } i \in \{1, \ldots, N\}, \tag{25}$$

and

$$\sum_{i=1}^N p_i = 1, \tag{26}$$

where $p \in \mathbb{R}^N$ is defined by (24). Consider

$$\tau = \frac{1}{(N-k+2\lambda)^2}\Big(\sum_{j \notin I_k} \sqrt{\alpha_j}\Big)^2 \in \mathbb{R}$$

and $(\mu_i)_{i=1}^N \in \mathbb{R}^N$ defined by $\mu_i = \tau - \alpha_i$ if $i \in I_k$ and $\mu_i = 0$ if $i \notin I_k$. Thus, we have the second condition in (25). Let us now prove the first condition in (25). Let $i \in \{1, \ldots, N\}$. If $i \in I_k$, then $p_i = 0$ and

$$\frac{-\alpha_i}{(1+2\lambda p_i)^2} + \tau - \mu_i = -\alpha_i + \tau - \mu_i = 0.$$

If $i \notin I_k$, then by (24), we have that

$$(1+2\lambda p_i)^2 = \frac{(N-k+2\lambda)^2 \alpha_i}{\Big(\sum_{j \notin I_k}\sqrt{\alpha_j}\Big)^2},$$

hence

$$\frac{-\alpha_i}{(1+2\lambda p_i)^2} + \tau - \mu_i = -\frac{\Big(\sum_{j \notin I_k}\sqrt{\alpha_j}\Big)^2}{(N-k+2\lambda)^2} + \tau - \mu_i = -\mu_i = 0,$$

proving the first condition in (25). We claim that $\mu_i \geq 0$ for all $i \in I_k$ (note that if $k = 0$, the latter is direct since $I_0 = \emptyset$, so in order to prove this claim we assume that $k > 0$). Let $i \in I_k$. Then $\alpha_i \leq \alpha_{\ell_k}$.

Now, by definition of k, we have that $k-1$ is not in the set in (23), that is,

$$(N-k+1+2\lambda)\sqrt{\alpha_{\ell_k}} \leq \sum_{j\notin I_{k-1}} \sqrt{\alpha_j} = \sqrt{\alpha_{\ell_k}} + \cdots + \sqrt{\alpha_{\ell_N}},$$

which yields that

$$(N-k+2\lambda)\sqrt{\alpha_i} \leq (N-k+2\lambda)\sqrt{\alpha_{\ell_k}} \leq \sqrt{\alpha_{\ell_{k+1}}} + \cdots + \sqrt{\alpha_{\ell_N}} = \sum_{j\notin I_k} \sqrt{\alpha_j},$$

which implies that $\mu_i = \tau - \alpha_i = \frac{1}{(N-k+2\lambda)^2}\Big(\sum_{j\notin I_k}\sqrt{\alpha_j}\Big)^2 - \alpha_i \geq 0$. We claim now that $p_i \geq 0$ for all $i \notin I_k$. Let $i \notin I_k$. Then $\alpha_i \geq \alpha_{\ell_{k+1}}$ and therefore

$$\frac{(N-k+2\lambda)\sqrt{\alpha_i}}{\sum_{j\notin I_k}\sqrt{\alpha_j}} \geq \frac{(N-k+2\lambda)\sqrt{\alpha_{\ell_{k+1}}}}{\sum_{j\notin I_k}\sqrt{\alpha_j}} > 1,$$

where the last inequality is by the definition of k. Thus, by definition in (24), we obtain that $p_i \geq 0$. Finally, since $|I_k^c| = N-k$, we deduce from definition in (24), that $\sum_{i=1}^N p_i = 1$. In summary, we have proved (25)-(26). □

Consider now the following function

$$x \in \mathcal{H} \mapsto \widetilde{f}(x) = \sup_{p\in\mathcal{P}} \sum_{i=1}^N p_i(\langle x, Qx\rangle + \langle b_i, x\rangle + c_i), \tag{27}$$

where $\mathcal{P} \subset \Delta_N$ is a nonempty closed convex set, $Q\colon \mathcal{H} \to \mathcal{H}$ is a bounded linear operator which is symmetric positive semidefinite, $b_i \in \mathcal{H}$, and $c_i \in \mathbb{R}$. The following result provides an algorithm for compute the proximity operator of the function $\widetilde{f}$ in (27).

Proposition 2.6. *Let $x \in \mathcal{H}$, $\lambda > 0$, $C = (\mathrm{Id} + 2\lambda Q)^{-1}$ and let $B \in \mathcal{L}(\mathbb{R}^N, \mathcal{H})$ be defined by $Bp = \sum_{j=1}^N p_j b_j$. Let $M = \lambda B^*(\mathrm{Id} - \lambda C^*Q - \frac{1}{2}C^*)CB \in \mathcal{L}(\mathbb{R}^N)$ and let $\gamma \in \mathbb{R}^N$ be defined by*

$$\gamma_i = \langle b_i, C(2\lambda QC - 2\,\mathrm{Id} + C)x\rangle - c_i.$$

Then the following hold.

(i) *M is a symmetric positive semidefinite operator.*

(ii) *Let $p^0 \in \mathbb{R}^N$, let $\overline{p}^0 \in \mathbb{R}^N$, and let $q^0 \in \mathbb{R}^N$ such that $p^0 = \overline{p}^0$. Let $L = M^{1/2}$ and let $\tau, \sigma > 0$ such that $\tau\sigma\|L\|^2 < 1$. For every $k \in \mathbb{N}$, we consider the following routine*

$$\left\lfloor \begin{aligned} x^k &= C(x - \lambda B p^k)\\ q^{k+1} &= \tfrac{2}{\sigma+2}(q^k + \sigma L\overline{p}^k)\\ u^{k+1} &= p^k - \tau(L^*q^{k+1} + \gamma)\\ p^{k+1} &= \mathrm{proj}_{\mathcal{P}}(u^{k+1})\\ \overline{p}^{k+1} &= p^{k+1} + u^{k+1} - p^k. \end{aligned} \right. \tag{28}$$

Then $x^k \to P_\lambda \widetilde{f}(x)$.

Proof. For every $p \in \mathcal{P}$, let us define the function

$$x \in \mathcal{H} \mapsto \widetilde{g}_p(x) = \sum_{i=1}^{N} p_i(\langle x, Qx\rangle + \langle b_i, x\rangle + c_i).$$

Note that $\{\widetilde{g}_p\}_{p\in\mathcal{P}} \subset \Gamma_0(\mathcal{H})$. Moreover, the function $p \mapsto \widetilde{g}_p(x)$ is concave and upper semicontinuous for all $x \in \mathcal{H}$. In addition, since $\mathcal{P} \subset \Delta_N$, then $\widetilde{f} = \sup_{p\in\mathcal{P}} \widetilde{g}_p$ is proper. Furthermore, $\mathcal{P}$ is a nonempty compact convex set. Then, by [13, Theorem 3.5], we have that

$$P_\lambda \widetilde{f}(x) = P_\lambda \widetilde{g}_{\overline{p}}(x), \quad \text{with } \ \overline{p} \in \operatorname*{arg\,max}_{p\in\mathcal{P}} e_\lambda \widetilde{g}_p(x). \tag{29}$$

Note that $\widetilde{g}_p$ is differentiable. Then, given $\overline{p} \in \mathcal{P} \subset \Delta_N$, we have

$$\begin{aligned} y = P_\lambda \widetilde{g}_{\overline{p}}(x) &\Leftrightarrow\ x = y + \lambda \nabla \widetilde{g}_{\overline{p}}(y) \ \Leftrightarrow\ x = y + \lambda \sum_{i=1}^{N} \overline{p}_i(2Qy + b_i) \\ &\Leftrightarrow\ x = (\mathrm{Id} + 2\lambda Q)y + \lambda \sum_{i=1}^{N} \overline{p}_i b_i \\ &\Leftrightarrow\ y = (\mathrm{Id} + 2\lambda Q)^{-1}\Big(x - \lambda \sum_{i=1}^{N} \overline{p}_i b_i\Big) \ \Leftrightarrow\ y = C(x - \lambda B\overline{p}), \end{aligned} \tag{30}$$

where we defined $C := (\mathrm{Id} + 2\lambda Q)^{-1}$, and B is here the linear operator defined by $p \in \mathbb{R}^N \mapsto Bp = \sum_{i=1}^{N} p_i b_i$. Let us compute $e_\lambda \widetilde{g}_p(x)$ with $p \in \mathcal{P} \subset \Delta_N$. By (30), we obtain that

$$\begin{aligned} e_\lambda \widetilde{g}_p(x) &= \widetilde{g}_p(P_\lambda \widetilde{g}_p(x)) + \frac{1}{2\lambda}\|x - P_\lambda \widetilde{g}_p(x)\|^2 \\ &= \sum_{i=1}^{N} p_i(\langle C(x - \lambda Bp), QC(x - \lambda Bp)\rangle + \langle b_i, C(x - \lambda Bp)\rangle + c_i) \\ &\quad + \frac{1}{2\lambda}\|x - C(x - \lambda Bp)\|^2 \\ &=: \ell_1(p, x) + \ell_2(p, x). \end{aligned} \tag{31}$$

Let $c = (c_i)_{i=1}^N$. Then the first term in (31) is

$$\begin{aligned} \ell_1(p, x) &= \langle C(x - \lambda Bp), QC(x - \lambda Bp)\rangle + \langle Bp, C(x - \lambda Bp)\rangle + p^\top c \\ &= \langle Cx, QCx\rangle - \lambda\langle Cx, QCBp\rangle - \lambda\langle CBp, QCx\rangle + \lambda^2\langle CBp, QCBp\rangle \\ &\quad + \langle Bp, Cx\rangle - \lambda\langle Bp, CBp\rangle + p^\top c \\ &= \langle Cx, QCx\rangle - 2\lambda\langle Cx, QCBp\rangle + \lambda^2\langle p, (CB)^* QCBp\rangle \\ &\quad + \langle Bp, Cx\rangle - \lambda\langle p, B^* CBp\rangle + p^\top c \\ &= \langle Cx, QCx\rangle + \lambda^2\langle p, (CB)^* QCBp\rangle - \lambda\langle p, B^* CBp\rangle \\ &\quad + \langle Bp, Cx - 2\lambda CQCx\rangle + p^\top c. \end{aligned}$$

On the other hand, the second term in (31) is

$$\begin{aligned}\ell_2(p,x) &= \frac{1}{2\lambda}\|(\mathrm{Id}-C)x+\lambda CBp\|^2\\ &= \frac{1}{2\lambda}\|(\mathrm{Id}-C)x\|^2+\langle(\mathrm{Id}-C)x, CBp\rangle+\frac{\lambda}{2}\|CBp\|^2.\\ &= \frac{1}{2\lambda}\|(\mathrm{Id}-C)x\|^2+\langle Bp, C(\mathrm{Id}-C)x\rangle+\frac{\lambda}{2}\langle p,(CB)^*CBp\rangle.\end{aligned}$$

Let $\gamma\in\mathbb{R}^N$ defined by

$$\gamma_i=\langle b_i, 2\lambda CQCx-Cx-C(\mathrm{Id}-C)x\rangle-c_i=\langle b_i, C(2\lambda QC-2\,\mathrm{Id}+C)x\rangle-c_i.$$

Then $$e_\lambda\widetilde{g}_p(x)=\frac{1}{2\lambda}\|(\mathrm{Id}-C)x\|^2+\langle Cx,QCx\rangle-p^\top\gamma-\langle p,Mp\rangle,$$

where $M=\lambda B^*CB-\lambda^2(CB)^*QCB-\frac{\lambda}{2}(CB)^*CB=\lambda B^*(\mathrm{Id}-\lambda C^*Q-\frac{1}{2}C^*)CB$. Therefore the problem $\max_{p\in\mathcal{P}} e_\lambda\widetilde{g}_p(x)$ reduces to

$$\max_{p\in\mathcal{P}} -p^\top\gamma-\langle p,Mp\rangle \Leftrightarrow \min_{p\in\mathcal{P}} p^\top\gamma+\langle p,Mp\rangle. \tag{32}$$

Observe that, since C and Q are symmetric, M also is symmetric. Now, note that $p\mapsto e_\lambda\widetilde{g}_p(x)$ is concave since is the infimum of affine linear functions. Then, $p\mapsto\langle p,Mp\rangle+p^\top\gamma$ is convex. Thus, M is positive semidefinite. Hence there exists $M^{1/2}$ symmetric positive semidefinite such that $M=M^{1/2}M^{1/2}$. Therefore we have $\langle p,Mp\rangle=\|M^{1/2}p\|^2$. Thus, the problem (32) is equivalent to

$$\min_{p\in\mathcal{P}}\overline{g}(Lp)+\overline{h}(p), \tag{33}$$

where $\overline{g}=\|\cdot\|^2$, $L=M^{1/2}$, and $\overline{h}$ is defined by $\overline{h}(p)=p^\top\gamma$. Now, we have that $P_\sigma\overline{g}(x)=x/(1+2\sigma)$, which implies that

$$P_\sigma\overline{g}^*(x)=x-\sigma P_{1/\sigma}\overline{g}(x/\sigma)=x-\frac{\sigma}{1+2/\sigma}\cdot\frac{x}{\sigma}=\frac{2}{\sigma+2}x.$$

In addition, $\nabla\overline{h}(p)=\gamma$ is Lipschitz. Thus, the algorithm (28) is equivalent to

$$\left\lfloor\begin{aligned}x^k &= C(x-\lambda Bp^k)\\ q^{k+1} &= P_\sigma\overline{g}^*(q^k+\sigma L\overline{p}^k)\\ u^{k+1} &= p^k-\tau(L^*q^{k+1}+\nabla\overline{h}(p^k))\\ p^{k+1} &= \mathrm{proj}_{\mathcal{P}}(u^{k+1})\\ \overline{p}^{k+1} &= p^{k+1}+u^{k+1}-p^k.\end{aligned}\right.$$

From [6, Theorem 3.1], there exists $\overline{p}\in\mathcal{P}$ solution to (33) such that $p^k\to\overline{p}$. That is $\overline{p}\in\arg\max_{p\in\mathcal{P}} e_\lambda\widetilde{g}_p(x)$. Now, by continuity, $x^k\to C(x-\lambda B\overline{p})$. Finally, by (29) and (30), $C(x-\lambda B\overline{p})=P_\lambda\widetilde{f}(x)$. □

Remark 2.7. In the case when $\mathcal{P}=\Delta_N$, the projection onto $\mathcal{P}$ can be obtained efficiently [11]. In the case when $\mathcal{P}$ is the ambiguity set given in (i)(b), we have that $\mathcal{P}=\Delta_N\cap\mathcal{P}_2$ and the projection onto $\mathcal{P}$ can be calculated using the Dykstra's algorithm [7, Proposition 5.3], which requires the computation of the projection onto Δ_N and $\mathcal{P}_2$.

Note that by using [2, Proposition 29.3] the projection onto $\mathcal{P}_2$ has an explicit form. Similarly, if $\mathcal{P}$ is the set in (i)(c), then $\mathcal{P}=\Delta_N\cap\mathcal{P}_3$ and we can use the Dykstra's algorithm for computing the projection onto $\mathcal{P}$. Note that the projection onto $\mathcal{P}_3$ has an explicit form given in [2, Example 29.21].

3. Algorithms

In this section we propose splitting algorithms for solving problems (9) and (10). This method is based on the results of Section 2, in which we compute the proximity operator of the supremum function appearing in problem (7). Note that the difficulty of calculating this proximity operator depends on the functions f_i and the set $\mathcal{P}$.

On the other hand, to solve problem (11), we propose a second method, which focuses in the resolution of an equivalent monotone inclusion problem, coming from the optimality conditions via the computation of resolvents of maximally monotone operators involved given in [3, Proposition 4.1]. In addition, we propose a third and fourth method to solve the problem (11), which correspond to the application of the algorithms proposed in [4] and [9], respectively, to solve a product space formulation involving the normal cone to a closed vector subspace.

3.1. A proximal algorithm

Note that the problem (7) is equivalent to

$$\min_{\mathbf{x}\in V} f(\mathbf{x}) + \iota_{\overline{\mathcal{Q}}}(\mathbf{x}) + H(\mathbf{x}), \tag{34}$$

where $\overline{\mathcal{Q}} = \bigtimes_{i=1}^{N} \mathcal{Q}_i$ and f is the function defined in (12). The problem (34) can be solved using the algorithm proposed in [5], which requires calculating the proximity of f, the projection on $\overline{\mathcal{Q}}$, and ∇H.

Let us define the following functions

$$(\alpha, d) \in \mathbb{R}^N_{++} \times \mathbb{R}^N \mapsto \rho(\alpha, d) = \operatorname*{arg\,min}_{p\in\mathcal{P}} \Big\{ \frac{1}{2}\sum_{i=1}^{N} p_i^2\alpha_i - \sum_{i=1}^{N} p_i d_i \Big\},$$

$$(\alpha, \lambda) \in \mathbb{R}^N_{+} \times \mathbb{R}_{++} \mapsto \varphi(\alpha, \lambda) = \operatorname*{arg\,max}_{p\in\Delta_N} \sum_{i=1}^{N} \Big(\frac{p_i}{1+2\lambda p_i}\Big)\alpha_i.$$

In order to solve the problem in the definition of ρ, we use the method proposed in [8] for the cases (i)(a) and (i)(b), and the algorithm (20) for the case (i)(c). On the other hand, the explicit form of φ is given by (24). From Proposition 2.1(i), [2, Proposition 24.8(ix)], and [5, Theorem 3.1], we obtain the following

Proposition 3.1. *Let $\mathbf{u}^0 \in \mathcal{H}^N$, let $\mathbf{x}^0 \in V$, let $\overline{\mathbf{x}}^0 = \mathbf{x}^0$, and let $\mathbf{y}^0 \in V^\perp$. Let $\lambda \in]0, 2\beta[$ and let $\gamma > 0$ such that $\gamma < \frac{1}{\lambda} - \frac{1}{2\beta}$. For every $k \in \mathbb{N}$, we consider the following routine.*

$$\left\lfloor \begin{aligned}
&u_i^{k+1} = u_i^k + \gamma\overline{x}_i^k - \gamma\,\mathrm{proj}_{\mathcal{Q}_i}\Big(\tfrac{u_i^k}{\gamma} + \overline{x}_i^k\Big) \quad \text{for all } i \in \{1,\dots,N\}\\
&\overline{\mathbf{z}}^k = \mathbf{x}^k + \lambda\mathbf{y}^k - \lambda\,\mathrm{proj}_V(\mathbf{u}^{k+1} + \nabla H(\mathbf{x}^k))\\
&p^k = \rho((\lambda\|a_i\|^2)_{i=1}^N, (\langle a_i, \overline{z}_i^k\rangle + \xi_i)_{i=1}^N)\\
&w_i^{k+1} = \overline{z}_i^k - \lambda p_i^k a_i \quad \text{for all } i \in \{1,\dots,N\}\\
&\mathbf{x}^{k+1} = \mathrm{proj}_V\,\mathbf{w}^{k+1}\\
&\mathbf{y}^{k+1} = \mathbf{y}^k + (\mathbf{x}^{k+1} - \mathbf{w}^{k+1})/\lambda\\
&\overline{\mathbf{x}}^{k+1} = 2\mathbf{x}^{k+1} - \mathbf{x}^k.
\end{aligned} \right. \tag{35}$$

Then there exists a solution $\mathbf{x} \in \mathcal{H}^N$ to problem (9) such that $\mathbf{x}^k \rightharpoonup \mathbf{x}$.

Similarly, from Proposition 2.1(ii), [2, Proposition 24.8(ix)], and [5, Theorem 3.1], we obtain the following

Proposition 3.2. *Let $\mathbf{u}^0 \in \mathcal{H}^N$, let $\mathbf{x}^0 \in V$, let $\overline{\mathbf{x}}^0 = \mathbf{x}^0$, and let $\mathbf{y}^0 \in V^\perp$. Let $\lambda \in]0, 2\beta[$ and let $\gamma > 0$ such that $\gamma < \frac{1}{\lambda} - \frac{1}{2\beta}$. For every $k \in \mathbb{N}$, we consider the following routine.*

$$\left\lfloor \begin{aligned} u_i^{k+1} &= u_i^k + \gamma\overline{x}_i^k - \gamma \operatorname{proj}_{\mathcal{Q}_i}\left(\frac{u_i^k}{\gamma} + \overline{x}_i^k\right) \quad \text{for all } i \in \{1,\dots,N\}\\ \overline{\mathbf{z}}^k &= \mathbf{x}^k + \lambda\mathbf{y}^k - \lambda \operatorname{proj}_V(\mathbf{u}^{k+1} + \nabla H(\mathbf{x}^k))\\ p^k &= \varphi((\|\overline{z}_i^k - \xi_i\|^2)_{i=1}^N, \lambda)\\ w_i^{k+1} &= \frac{\overline{z}_i^k + 2\lambda p_i^k \xi_i}{1 + 2\lambda p_i^k} \quad \text{for all } i \in \{1,\dots,N\}\\ \mathbf{x}^{k+1} &= \operatorname{proj}_V \mathbf{w}^{k+1}\\ \mathbf{y}^{k+1} &= \mathbf{y}^k + (\mathbf{x}^{k+1} - \mathbf{w}^{k+1})/\lambda\\ \overline{\mathbf{x}}^{k+1} &= 2\mathbf{x}^{k+1} - \mathbf{x}^k. \end{aligned} \right. \tag{36}$$

Then there exists a solution $\mathbf{x} \in \mathcal{H}^N$ to problem (10) such that $\mathbf{x}^k \rightharpoonup \mathbf{x}$.

3.2. Distributed forward-backward method

In this section, we propose an alternative method for solving problem (6). Note that in the cases (i) and (ii), we have that $\operatorname{dom} f_i = \mathcal{H}$ for all $i \in \{1,\dots,N\}$. In addition, $\operatorname{dom} h = \mathcal{H}$. Thus, by the Fermat's rule and [2, Corollary 16.50(iv)], the problem (6) is equivalent to solve the following inclusion (optimality condition)

$$\text{Find} \quad (x,p) \in \mathcal{H} \times \mathbb{R}^N \text{ such that } \begin{cases} 0 \in \nabla h(x) + N_{\mathcal{Q}}(x) + \sum_{i=1}^N p_i \partial f_i(x)\\ 0 \in N_{\mathcal{P}}(p) - G(x), \end{cases} \tag{37}$$

where $G(x) = (f_1(x),\dots,f_N(x))^\top \in \mathbb{R}^N$. If we have $\mathcal{P} = \Delta_N$, then $\mathcal{P} = \mathbb{R}^N_+ \cap \mathcal{P}_1$, where we recall that $\mathcal{P}_1 = \{p \in \mathbb{R}^N : \sum_{i=1}^N p_i = 1\}$. Since $\operatorname{int}(\mathbb{R}^N_+) \cap \mathcal{P}_1 \neq \emptyset$, then by [2, Corollary 16.48(ii)], $N_{\mathcal{P}} = N_{\mathbb{R}^N_+} + N_{\mathcal{P}_1}$. Now, by defining $\phi = \iota_{\mathbb{R}_-}$, we have that $N_{\mathbb{R}_+} = \partial\phi^*$. Define the following operators

$$(x,p) \in \mathcal{H} \times \mathbb{R}^N \mapsto B_i(x,p) := \begin{pmatrix} p_i \partial f_i(x)\\ \partial\phi^*(p_i)e_i - f_i(x)e_i \end{pmatrix} \text{ for all } i \in \{1,...,N\} \tag{38}$$

$$(x,p) \in \mathcal{H} \times \mathbb{R}^N \mapsto A_1(x,p) = \begin{pmatrix} N_{\mathcal{Q}}(x)\\ N_{\mathcal{P}_1}(p) \end{pmatrix}$$

$$(x,p) \in \mathcal{H} \times \mathbb{R}^N \mapsto C(x,p) = \begin{pmatrix} \nabla h(x)\\ 0 \end{pmatrix},$$

where e_i is the canonical vector in $\mathbb{R}^N$. Then the inclusion (37) is equivalent to

$$\begin{pmatrix} 0\\ 0 \end{pmatrix} \in A_1(x,p) + \sum_{i=1}^N B_i(x,p) + \sum_{i=1}^N C_i(x,p), \tag{39}$$

where $C_1 = C$ and $C_i = 0$ for all $i \in \{2,\dots,N\}$.

If $\mathcal{P} = \{p \in \Delta_N : p \leq q\}$, then $\mathcal{P} = \mathbb{R}^N_+ \cap \mathcal{P}_1 \cap \mathcal{P}_2$, where $\mathcal{P}_2 = \{p \in \mathbb{R}^N : p \leq q\}$. By assumption in (i)(b), $\mathcal{P}_1 \cap \operatorname{int}(\mathbb{R}^N_+) \cap \operatorname{int}(\mathcal{P}_2) \neq \emptyset$. Hence, by [2, Corollary 16.50(iv)], $N_{\mathcal{P}} = N_{\mathbb{R}^N_+} + N_{\mathcal{P}_1} + N_{\mathcal{P}_2}$. Thus, in this case the inclusion (37) is equivalent to

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in A_1(x,p) + A_2(x,p) + \sum_{i=1}^{N} B_i(x,p) + \sum_{i=1}^{N+1} C_i(x,p), \tag{40}$$

where $(x,p) \in \mathcal{H} \times \mathbb{R}^N \mapsto A_2(x,p) = \begin{pmatrix} 0 \\ N_{\mathcal{P}_2}(p) \end{pmatrix}$ and $C_{N+1} = 0$. Similarly, if $\mathcal{P} = \{p \in \Delta_N : \mu_- \leq \sum_{i=1}^N p_i\xi_i \leq \mu_+\}$, we deduce that the inclusion (37) is equivalent to

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in A_1(x,p) + A_3(x,p) + \sum_{i=1}^{N} B_i(x,p) + \sum_{i=1}^{N+1} C_i(x,p), \tag{41}$$

where $$(x,p) \in \mathcal{H} \times \mathbb{R}^N \mapsto A_3(x,p) = \begin{pmatrix} 0 \\ N_{\mathcal{P}_3}(p) \end{pmatrix}$$

and we recall that $$\mathcal{P}_3 = \{p \in \mathbb{R}^N : \mu_- \leq \sum_{i=1}^{N} p_i\xi_i \leq \mu_+\}.$$

Note that $(A_i)_{i=1}^3$ are maximally monotone and $(C_i)_{i=1}^{N+1}$ are cocoercive. Thus, to apply the splitting methods in order to solve the previous inclusions, we need to prove that B_i is maximally monotone and compute $J_{\gamma B_i}$ with $\gamma > 0$.

Proposition 3.3. *Let $i \in \{1, \ldots, N\}$ and consider the operator B_i defined in (38), where $f_i \in \Gamma_0(\mathcal{H})$ satisfies $\operatorname{dom} f_i = \mathcal{H}$. Then the following hold.*

(i) *B_i is maximally monotone.*

(ii) *Let $(x,p) \in \mathcal{H} \times \mathbb{R}^N$, let $\gamma > 0$, and let $\overline{\omega}$ the unique real number in $[0, +\infty[$ such that*

$$\overline{\omega} = \begin{cases} 0 & \text{if } p_i + \gamma f_i(x) \leq 0 \\ p_i + \gamma f_i(P_{\gamma\overline{\omega}} f_i(x)) & \text{if } p_i + \gamma f_i(x) > 0. \end{cases} \tag{42}$$

Then $$J_{\gamma B_i}(x,p) = \begin{cases} (x,\omega) & \text{if } p_i + \gamma f_i(x) \leq 0 \\ (P_{\gamma\omega_i} f_i(x), \omega) & \text{if } p_i + \gamma f_i(x) > 0, \end{cases}$$

where $\omega_i = \overline{\omega}$ and $\omega_j = p_j$ for all $j \neq i$.

Proof. (i): First, we prove that B_i is monotone. Let therefore $(u,v) \in B_i(x,p)$ and $(u',v') \in B_i(x',p')$. Then $u \in p_i \partial f_i(x)$, $u' \in p'_i \partial f_i(x')$, $v_i \in \partial\phi^*(p_i) - f_i(x)$, and $v'_i \in \partial\phi^*(p'_i) - f_i(x')$. In addition, $v_j = v'_j = 0$ for all $j \neq i$. By [3, Proposition 3.1(v)], we have that the following operator

$$(x,\xi) \in \mathcal{H} \times \mathbb{R} \mapsto \xi \partial f_i(x) \times (\partial\phi^*(\xi) - f_i(x)) \tag{43}$$

is maximally monotone. Thus,

$$\langle (u,v) - (u',v'), (x,p) - (x',p') \rangle = \langle u - u', x - x' \rangle + \langle v_i - v'_i, p_i - p'_i \rangle \geq 0,$$

which proves that B_i is monotone. By [2, Theorem 21.1], it is enough to prove that $\operatorname{ran}(\operatorname{Id} + B_i) = \mathcal{H} \times \mathbb{R}^N$. Let $(y,\omega) \in \mathcal{H} \times \mathbb{R}^N$. Then $(y,\omega_i) \in \mathcal{H} \times \mathbb{R}$.

Since the operator (43) is maximally monotone, then from [2, Theorem 21.1], there exists $(x,\xi)\in\mathcal{H}\times\mathbb{R}$ such that

$$(y,\omega_i)\in(x,\xi)+\begin{pmatrix}\xi\partial f_i(x)\\ \partial\phi^*(\xi)-f_i(x)\end{pmatrix}.$$

Let $p\in\mathbb{R}^N$ defined by $p_i=\xi$ and $p_j=\omega_j$ for all $j\neq i$. Then

$$(y,\omega)\in(x,p)+\begin{pmatrix}p_i\partial f_i(x)\\ \partial\phi^*(p_i)e_i-f_i(x)e_i\end{pmatrix}=(x,p)+B_i(x,p).$$

That is, $\operatorname{ran}(\mathrm{Id}+B_i)=\mathcal{H}\times\mathbb{R}^N$.

(ii): Let $(y,\omega)\in\mathcal{H}\times\mathbb{R}^N$. Then

$$\begin{aligned}(y,\omega)=J_{\gamma B_i}(x,p)\quad&\Leftrightarrow\quad(x,p)\in(y,\omega)+\gamma B_i(y,\omega)\\ &\Leftrightarrow\quad\begin{cases}x\in y+\gamma\omega_i\partial f_i(y)\\ p_i\in\omega_i+\gamma\partial\phi^*(\omega_i)-\gamma f_i(y)\\ p_j=\omega_j\text{ for all }j\neq i\end{cases}\\ &\Leftrightarrow\quad\begin{cases}y=P_{\gamma\omega_i}f_i(x)\\ \omega_i=P_\gamma\phi^*(p_i+\gamma f_i(y))\\ \omega_j=p_j\text{ for all }j\neq i\end{cases}\end{aligned}$$

Using [3, Remark 4.1 & Example 4.1] and that $\operatorname{dom}f_i=\mathcal{H}$, we deduce that

$$\omega_i=\begin{cases}0 & \text{if } p_i+\gamma f_i(x)\leq 0\\ p_i+\gamma f_i(P_{\gamma\omega_i}f_i(x)) & \text{if } p_i+\gamma f_i(x)>0.\end{cases}\tag{44}$$

Therefore, $$J_{\gamma B_i}(x,p)=\begin{cases}(x,\omega) & \text{if } p_i+\gamma f_i(x)\leq 0\\ (P_{\gamma\omega_i}f_i(x),\omega) & \text{if } p_i+\gamma f_i(x)>0,\end{cases}$$

where $\omega_i\in[0,+\infty[$ is the unique solution to (44) and $\omega_j=p_j$ for all $j\neq i$. □

In the case (i) we have $f_i(x)=\langle a_i,x\rangle+\xi_i$. Therefore we get $P_{\gamma\omega}f_i(x)=x-\gamma\omega a_i$ and thus $f_i(P_{\gamma\omega}f_i(x))=\langle a_i,x-\gamma\omega a_i\rangle+\xi_i$. Hence

$$\begin{aligned}\omega=p_i+\gamma f_i(P_{\gamma\omega}f_i(x))\quad&\Leftrightarrow\omega=p_i+\gamma\langle a_i,x-\gamma\omega a_i\rangle+\gamma\xi_i\\ &\Leftrightarrow\omega=\frac{p_i+\gamma(\langle a_i,x\rangle+\xi_i)}{1+\gamma^2\|a_i\|^2}.\end{aligned}\tag{45}$$

It follows from (45) that, in the case (i), the solution to (42) reduces to

$$\overline{\omega}=\begin{cases}0 & \text{if } p_i+\gamma(\langle a_i,x\rangle+\xi_i)\leq 0\\ \dfrac{p_i+\gamma(\langle a_i,x\rangle+\xi_i)}{1+\gamma^2\|a_i\|^2} & \text{if } p_i+\gamma(\langle a_i,x\rangle+\xi_i)>0.\end{cases}$$

On the other hand, $J_{\gamma A_1}(x,p) = (\mathrm{proj}_{\mathcal{Q}}(x), \mathrm{proj}_{\mathcal{P}_1}(p))$, $J_{\gamma A_2}(x,p) = (x, \mathrm{proj}_{\mathcal{P}_2}(p))$, and $J_{\gamma A_3}(x,p) = (x, \mathrm{proj}_{\mathcal{P}_3}(p))$. Once calculated $J_{\gamma A_1}$, $J_{\gamma A_2}$, $J_{\gamma A_3}$, and $J_{\gamma B_i}$ for all $i \in \{1,\dots,N\}$, we can apply the method proposed in [1] for solving problems (39), (40), and (41). Define $\mathbf{H} = \mathcal{H} \times \mathbb{R}^N$. From [1, Theorem 3], we deduce the following results.

Proposition 3.4. *Let $\lambda \in \left]0, 2\beta\right[$ and let $\gamma \in \left]0, 1 - \frac{\lambda}{2\beta}\right[$.*

Let $\overline{\mathbf{z}}^0 = ((\overline{x}_1^0, \overline{p}_1^0), ..., (\overline{x}_N^0, \overline{p}_N^0)) \in \mathbf{H}^N$. For every $k \in \mathbb{N}$, we consider the following routine.

$$\left\lfloor \begin{array}{l} x_1^k := \mathrm{proj}_{\mathcal{Q}}(\overline{x}_1^k) \text{ and } p_1^k := \mathrm{proj}_{\mathcal{P}_1}(\overline{p}_1^k) \\ (x_2^k, p_2^k) := J_{\lambda B_1}(\overline{x}_2^k + x_1^k - \overline{x}_1^k - \lambda \nabla h(x_1^k), \overline{p}_2^k + p_1^k - \overline{p}_1^k) \\ (x_i^k, p_i^k) := J_{\lambda B_{i-1}}(\overline{x}_i^k + x_{i-1}^k - \overline{x}_{i-1}^k, \overline{p}_i^k + p_{i-1}^k - \overline{p}_{i-1}^k) \text{ for all } i \in \{3, ..., N\} \\ (x_{N+1}^k, p_{N+1}^k) := J_{\lambda B_N}(x_1^k + x_N^k - \overline{x}_N^k, p_1^k + p_N^k - \overline{p}_N^k) \\ \left.\begin{array}{l} \overline{x}_j^{k+1} = \overline{x}_j^k + \gamma(x_{j+1}^k - x_j^k) \\ \overline{p}_j^{k+1} = \overline{p}_j^k + \gamma(p_{j+1}^k - p_j^k) \end{array}\right\} \text{ for all } j \in \{1,\dots,N\} \end{array}\right. \tag{46}$$

Then there exists $(x,p) \in \mathbf{H}$ solution to problem (39) such that $(x_1^k, p_1^k) \rightharpoonup (x,p)$.

Proposition 3.5. *Let $\lambda \in \left]0, 2\beta\right[$ and let $\gamma \in \left]0, 1 - \frac{\lambda}{2\beta}\right[$.*
Let $\overline{\mathbf{z}}^0 = ((\overline{x}_1^0, \overline{p}_1^0), \dots, (\overline{x}_{N+1}^0, \overline{p}_{N+1}^0)) \in \mathbf{H}^{N+1}$. For every $k \in \mathbb{N}$, we consider the following routine.

$$\left\lfloor \begin{array}{l} x_1^k := \mathrm{proj}_{\mathcal{Q}}(\overline{x}_1^k) \text{ and } p_1^k := \mathrm{proj}_{\mathcal{P}_1}(\overline{p}_1^k) \\ (x_2^k, p_2^k) := (\overline{x}_2^k + x_1^k - \overline{x}_1^k - \lambda \nabla h(x_1^k), \mathrm{proj}_{\mathcal{P}}(\overline{p}_2^k + p_1^k - \overline{p}_1^k)) \\ (x_i^k, p_i^k) := J_{\lambda B_{i-2}}(\overline{x}_i^k + x_{i-1}^k - \overline{x}_{i-1}^k, \overline{p}_i^k + p_{i-1}^k - \overline{p}_{i-1}^k) \text{ for all } i \in \{3, ..., N+1\} \\ (x_{N+2}^k, p_{N+2}^k) := J_{\lambda B_N}(x_1^k + x_{N+1}^k - \overline{x}_{N+1}^k, p_1^k + p_{N+1}^k - \overline{p}_{N+1}^k) \\ \left.\begin{array}{l} \overline{x}_j^{k+1} = \overline{x}_j^k + \gamma(x_{j+1}^k - x_j^k) \\ \overline{p}_j^{k+1} = \overline{p}_j^k + \gamma(p_{j+1}^k - p_j^k) \end{array}\right\} \text{ for all } j \in \{1,\dots,N+1\} \end{array}\right. \tag{47}$$

Then the following hold.

(i) *If $\mathcal{P} = \mathcal{P}_2$, then there exists $(x,p) \in \mathbf{H}$ solution to problem (40) such that $(x_1^k, p_1^k) \rightharpoonup (x,p)$.*

(ii) *If $\mathcal{P} = \mathcal{P}_3$, then there exists $(x,p) \in \mathbf{H}$ solution to problem (41) such that $(x_1^k, p_1^k) \rightharpoonup (x,p)$.*

Remark 3.6. By [2, Example 29.18], the projection onto $\mathcal{P}_1$ is given by

$$\mathrm{proj}_{\mathcal{P}_1}(x) = x + \Big(\frac{1 - \sum_{i=1}^N x_i}{N}\Big)\mathbf{1} \quad \text{for all } x \in \mathbb{R}^N,$$

where $\mathbf{1} = (1,\dots,1) \in \mathbb{R}^N$.

On the other hand, note that $\mathcal{P}_2 = \bigtimes_{i=1}^N \{p \in \mathbb{R} : p \le q_i\}$.

Then by [2, Proposition 29.3],

$$(\mathrm{proj}_{\mathcal{P}_2}(x))_i = \begin{cases} x_i & \text{if } x_i \leq q_i \\ q_i & \text{if } x_i > q_i \end{cases} \quad \text{for all } i \in \{1,\dots,N\} \text{ and } x \in \mathbb{R}^N.$$

Finally, the projection onto $\mathcal{P}_3$ can be calculated using [2, Example 29.21]. □

3.3. Forward-backward with subspaces

Let us see another method to solve the inclusions (39), (40), and (41). Consider the spaces $\mathsf{H}_1 = (\mathcal{H}\times\mathbb{R}^N)^{N+1}$ and $\mathsf{H}_2 = (\mathcal{H}\times\mathbb{R}^N)^{N+2}$. Let $\widetilde{A}_1\colon \mathsf{H}_1 \to 2^{\mathsf{H}_1}$, $\widetilde{A}_2\colon \mathsf{H}_2 \to 2^{\mathsf{H}_2}$, $\widetilde{A}_3\colon \mathsf{H}_2 \to 2^{\mathsf{H}_2}$, $\widetilde{C}_1\colon \mathsf{H}_1 \to \mathsf{H}_1$, and $\widetilde{C}_2\colon \mathsf{H}_2 \to \mathsf{H}_2$ the operators defined by

$$\begin{aligned}
\widetilde{A}_1 &= B_1 \times \cdots \times B_N \times A_1 \\
\widetilde{C}_1 &= (C_1, \underbrace{0,\dots,0}_{N-\text{times}}) \\
\widetilde{A}_2 &= B_1 \times \cdots \times B_N \times A_1 \times A_2 \\
\widetilde{A}_3 &= B_1 \times \cdots \times B_N \times A_1 \times A_3 \\
\widetilde{C}_2 &= (C_1, \underbrace{0,\dots,0}_{(N+1)-\text{times}}).
\end{aligned}$$

Then the inclusion (39) is equivalent to

$$\text{Find} \quad z \in \mathsf{H}_1 \ \text{ such that } \ 0 \in \widetilde{A}_1(z) + \widetilde{C}_1(z) + N_{V_1}(z), \tag{48}$$

where $V_1 = \{z \in \mathsf{H}_1 \,:\, z_1 = \ldots = z_{N+1}\}$. Indeed, we have that

$$V_1^{\perp} = \{z \in \mathsf{H}_1 \,:\, \textstyle\sum_{i=1}^{N+1} z_i = 0\} = N_{V_1}(z) \ \text{ for all } z \in V_1.$$

Now, define $D_i = B_i$ for all $i \in \{1,\dots,N\}$, $D_{N+1} = A_1$, and $C_i = 0$ for all $i \in \{2,\dots,N+1\}$. Thus, denoting $\mathbf{x} = (x,p)$, (39) is equivalent to

$$\begin{aligned}
0 \in \sum_{i=1}^{N+1} D_i(\mathbf{x}) + C_i(\mathbf{x}) &\Leftrightarrow \Big(\exists (y_i)_{i=1}^{N+1} \in \bigtimes_{i=1}^{N+1} D_i(\mathbf{x})\Big) \quad 0 = \sum_{i=1}^{N+1}(-y_i - C_i(\mathbf{x})) \\
&\Leftrightarrow \Big(\exists (y_i)_{i=1}^{N+1} \in \bigtimes_{i=1}^{N+1} D_i(\mathbf{x})\Big) \quad -(y_1,\dots,y_{N+1}) - \widetilde{C}_1(\mathbf{x},\dots,\mathbf{x}) \in V_1^{\perp} \\
&\Leftrightarrow 0 \in \widetilde{A}_1(\mathbf{x},\dots,\mathbf{x}) + \widetilde{C}_1(\mathbf{x},\dots,\mathbf{x}) + N_{V_1}(\mathbf{x},\dots,\mathbf{x}).
\end{aligned}$$

Similarly, the inclusion (40) is equivalent to

$$\text{Find} \quad z \in \mathsf{H}_2 \ \text{ such that } \ 0 \in \widetilde{A}_2(z) + \widetilde{C}_2(z) + N_{V_2}(z), \tag{49}$$

where $V_2 = \{z \in \mathsf{H}_2 \,:\, z_1 = \ldots = z_{N+2}\}$. Finally, the inclusion (41) is equivalent to

$$\text{Find} \quad z \in \mathsf{H}_2 \ \text{ such that } \ 0 \in \widetilde{A}_3(z) + \widetilde{C}_2(z) + N_{V_2}(z). \tag{50}$$

Remark 3.7. Note that the projection on V_1 is given by

$$\operatorname{proj}_{V_1}(z) = \Big(\frac{1}{N+1}\sum_{j=1}^{N+1} z_j\Big)_{i=1}^{N+1} \quad \text{for all } z \in \mathsf{H}_1, \tag{51}$$

and the projection onto V_2 is similar. □

In order to solve (48), (49), and (50), we use the algorithm proposed in [4, Corollary 5.3] when $\lambda_n \equiv 1$ (which coincides with the algorithm proposed in [5] in this case). From [4, Corollary 5.3], [2, Proposition 23.18], and (51), we deduce the following results.

Proposition 3.8. *Let $\gamma \in]0, 2\beta[$, let $z^0 = ((x_1^0, p_1^0), \dots, (x_{N+1}^0, p_{N+1}^0)) \in V_1$, and let $\overline{z}^0 = ((\overline{x}_1^0, \overline{p}_1^0), \dots, (\overline{x}_{N+1}^0, \overline{p}_{N+1}^0)) \in V_1^\perp$. For every $k \in \mathbb{N}$, we consider the following routine.*

$$\left\lfloor \begin{aligned}
&(\widetilde{x}_i^k, \widetilde{p}_i^k) := J_{\gamma B_i}\Big(x_i^k + \gamma\overline{x}_i^k - \tfrac{\gamma}{N+1}\nabla h(x_1^k), p_i^k + \gamma\overline{p}_i^k\Big) \text{ for all } i \in \{1, \dots, N\}\\
&(\widetilde{x}_{N+1}^k, \widetilde{p}_{N+1}^k) := \big(\operatorname{proj}_{\mathcal{Q}}\big(x_{N+1}^k + \gamma\overline{x}_{N+1}^k - \tfrac{\gamma}{N+1}\nabla h(x_1^k)\big), \operatorname{proj}_{\mathcal{P}_1}(p_{N+1}^k + \gamma\overline{p}_{N+1}^k)\big)\\
&z_i^{k+1} := (x_i^{k+1}, p_i^{k+1}) = \tfrac{1}{N+1}\textstyle\sum_{j=1}^{N+1}(\widetilde{x}_j^k, \widetilde{p}_j^k) \text{ for all } i \in \{1, \dots, N+1\}\\
&\overline{z}_i^{k+1} := (\overline{x}_i^{k+1}, \overline{p}_i^{k+1}) = (\overline{x}_i^k, \overline{p}_i^k) + \tfrac{1}{\gamma}(x_i^{k+1} - \widetilde{x}_i^k, p_i^{k+1} - \widetilde{p}_i^k) \text{ for all } i \in \{1, ..., N+1\}.
\end{aligned} \right. \tag{52}$$

Then there exists $z \in V_1$ solution to problem (48) such that $z^k \rightharpoonup z$.

Proposition 3.9. *Let $\gamma \in]0, 2\beta[$, let $z^0 = ((x_1^0, p_1^0), \dots, (x_{N+2}^0, p_{N+2}^0)) \in V_2$, and let $\overline{z}^0 = ((\overline{x}_1^0, \overline{p}_1^0), \dots, (\overline{x}_{N+2}^0, \overline{p}_{N+2}^0)) \in V_2^\perp$. For every $k \in \mathbb{N}$, we consider the following routine.*

$$\left\lfloor \begin{aligned}
&(\widetilde{x}_i^k, \widetilde{p}_i^k) := J_{\gamma B_i}\Big(x_i^k + \gamma\overline{x}_i^k - \tfrac{\gamma}{N+2}\nabla h(x_1^k), p_i^k + \gamma\overline{p}_i^k\Big) \text{ for all } i \in \{1, \dots, N\}\\
&(\widetilde{x}_{N+1}^k, \widetilde{p}_{N+1}^k) := \big(\operatorname{proj}_{\mathcal{Q}}\big(x_{N+1}^k + \gamma\overline{x}_{N+1}^k - \tfrac{\gamma}{N+2}\nabla h(x_1^k)\big), \operatorname{proj}_{\mathcal{P}_1}(p_{N+1}^k + \gamma\overline{p}_{N+1}^k)\big)\\
&(\widetilde{x}_{N+2}^k, \widetilde{p}_{N+2}^k) := \big(x_{N+2}^k + \gamma\overline{x}_{N+2}^k - \tfrac{\gamma}{N+2}\nabla h(x_1^k), \operatorname{proj}_{\mathcal{P}}(p_{N+2}^k + \gamma\overline{p}_{N+2}^k)\big)\\
&z_i^{k+1} := (x_i^{k+1}, p_i^{k+1}) = \tfrac{1}{N+2}\textstyle\sum_{j=1}^{N+2}(\widetilde{x}_j^k, \widetilde{p}_j^k) \text{ for all } i \in \{1, \dots, N+2\}\\
&\overline{z}_i^{k+1} := (\overline{x}_i^{k+1}, \overline{p}_i^{k+1}) = (\overline{x}_i^k, \overline{p}_i^k) + \tfrac{1}{\gamma}(x_i^{k+1} - \widetilde{x}_i^k, p_i^{k+1} - \widetilde{p}_i^k) \text{ for all } i \in \{1, ..., N+2\}.
\end{aligned} \right. \tag{53}$$

Then the following hold.

(i) *If $\mathcal{P} = \mathcal{P}_2$, then there exists $z \in V_2$ solution to problem (49) such that $z^k \rightharpoonup z$.*

(ii) *If $\mathcal{P} = \mathcal{P}_3$, then there exists $z \in V_2$ solution to problem (50) such that $z^k \rightharpoonup z$.*

3.4. Davis-Yin's formulation

Another method for solving the problems (48), (49), and (50) (and therefore the inclusions (39), (40), and (41)) is the algorithm proposed in [9] since the normal cone to a closed vector subspace is a maximally monotone operator. From [9, Theorem 2.1], [2, Proposition 23.18], and (51), we obtain the following results.

Proposition 3.10. *Let $\gamma \in]0, 2\beta[$ and let $z^0 = ((x_1^0, p_1^0), \ldots, (x_{N+1}^0, p_{N+1}^0)) \in \mathsf{H}_1$. For every $k \in \mathbb{N}$, we consider the following routine.*

$$\left\lfloor \begin{array}{l} \overline{z}_i^k := (\overline{x}_i^k, \overline{p}_i^k) = \frac{1}{N+1}\sum_{j=1}^{N+1}(x_j^k, p_j^k) \text{ for all } i \in \{1, \ldots, N+1\} \\ (\widetilde{x}_1^k, \widetilde{p}_1^k) = J_{\gamma B_1}(2\overline{x}_1^k - x_1^k - \gamma\nabla h(\overline{x}_1^k), 2\overline{p}_1^k - p_1^k) \\ (\widetilde{x}_i^k, \widetilde{p}_i^k) = J_{\gamma B_i}(2\overline{x}_i^k - x_i^k, 2\overline{p}_i^k - p_i^k) \quad \text{for all } i \in \{2, \ldots, N\} \\ (\widetilde{x}_{N+1}^k, \widetilde{p}_{N+1}^k) = (\mathrm{proj}_{\mathcal{Q}}(2\overline{x}_{N+1}^k - x_{N+1}^k), \mathrm{proj}_{\mathcal{P}_1}(2\overline{p}_{N+1}^k - p_{N+1}^k)) \\ (x_i^{k+1}, p_i^{k+1}) = (x_i^k + \widetilde{x}_i^k - \overline{x}_i^k, p_i^k + \widetilde{p}_i^k - \overline{p}_i^k) \quad \text{for all } i \in \{1, \ldots, N+1\}. \end{array} \right. \tag{54}$$

Then there exists $z \in \mathsf{H}_1$ solution to problem (48) such that $\overline{z}^k \rightharpoonup z$.

Proposition 3.11. *Let $\gamma \in]0, 2\beta[$ and let $z^0 = ((x_1^0, p_1^0), \ldots, (x_{N+2}^0, p_{N+2}^0)) \in \mathsf{H}_2$. For every $k \in \mathbb{N}$, we consider the following routine.*

$$\left\lfloor \begin{array}{l} \overline{z}_i^k := (\overline{x}_i^k, \overline{p}_i^k) = \frac{1}{N+2}\sum_{j=1}^{N+2}(x_j^k, p_j^k) \text{ for all } i \in \{1, \ldots, N+2\} \\ (\widetilde{x}_1^k, \widetilde{p}_1^k) = J_{\gamma B_1}(2\overline{x}_1^k - x_1^k - \gamma\nabla h(\overline{x}_1^k), 2\overline{p}_1^k - p_1^k) \\ (\widetilde{x}_i^k, \widetilde{p}_i^k) = J_{\gamma B_i}(2\overline{x}_i^k - x_i^k, 2\overline{p}_i^k - p_i^k) \quad \text{for all } i \in \{2, \ldots, N\} \\ (\widetilde{x}_{N+1}^k, \widetilde{p}_{N+1}^k) = (\mathrm{proj}_{\mathcal{Q}}(2\overline{x}_{N+1}^k - x_{N+1}^k), \mathrm{proj}_{\mathcal{P}_1}(2\overline{p}_{N+1}^k - p_{N+1}^k)) \\ (\widetilde{x}_{N+2}^k, \widetilde{p}_{N+2}^k) = (2\overline{x}_{N+2}^k - x_{N+2}^k, \mathrm{proj}_{\mathcal{P}}(2\overline{p}_{N+2}^k - p_{N+2}^k)) \\ (x_i^{k+1}, p_i^{k+1}) = (x_i^k + \widetilde{x}_i^k - \overline{x}_i^k, p_i^k + \widetilde{p}_i^k - \overline{p}_i^k) \quad \text{for all } i \in \{1, \ldots, N+2\}. \end{array} \right. \tag{55}$$

Then the following hold.

(i) *If $\mathcal{P} = \mathcal{P}_2$, then there exists $z \in \mathsf{H}_2$ solution to problem (49) such that $\overline{z}^k \rightharpoonup z$.*

(ii) *If $\mathcal{P} = \mathcal{P}_3$, then there exists $z \in \mathsf{H}_2$ solution to problem (50) such that $\overline{z}^k \rightharpoonup z$.*

3.5. Dual method

In this section, we use the techniques proposed in [15, Section 4.2] for solving problem (6) for the ambiguity sets considered in the cases (i)(a), (i)(b), and (i)(c). In each case, we reformulate the dual of the following inner maximization problem

$$\max_{p \in \mathcal{P}} h(x) + \sum_{i=1}^{N} p_i f_i(x) = \sum_{i=1}^{N} p_i (f_i(x) + h(x)). \tag{56}$$

Observe that by assumption of each case, we have that the problem (56) is a convex problem that satisfy the Slater conditions. Therefore, the strong duality hold.

Case (i)(a): In this case $\mathcal{P} = \Delta_N$ and the Lagrangian function is

$$\begin{aligned} L(p, \lambda) &= \sum_{i=1}^{N} p_i (f_i(x) + h(x)) + \lambda\Big(1 - \sum_{i=1}^{N} p_i\Big) \\ &= \lambda + \sum_{i=1}^{N} p_i (f_i(x) + h(x) - \lambda). \end{aligned}$$

Then, (56) is equivalent to the following dual problem

$$\min_{\lambda\in\mathbb{R}}\max_{p\geq 0}\Big\{\lambda+\sum_{i=1}^{N}p_i(f_i(x)+h(x)-\lambda)\Big\}\;\Leftrightarrow\min_{\lambda\in\mathbb{R}}\Big\{\lambda+\max_{p\geq 0}\sum_{i=1}^{N}p_i(f_i(x)+h(x)-\lambda)\Big\}$$
$$\Leftrightarrow\min_{\lambda\in\mathbb{R}}\lambda\quad\text{s.t. } f_i(x)+h(x)-\lambda\leq 0,\quad\forall i\in\{1,\dots,N\}.$$

Thus, problem (6) is equivalent to

$$\begin{aligned}&\min_{x\in\mathcal{H},\,\lambda\in\mathbb{R}}\quad\lambda\\&\text{s.t. } f_i(x)+h(x)-\lambda\leq 0,\quad\forall i\in\{1,\dots,N\}\\&x\in\mathcal{Q}\end{aligned}\tag{57}$$

Case (i)(b): In this case $\mathcal{P}=\{p\in\Delta_N : p\leq q\}$ and hence the Lagrangian function is

$$\begin{aligned}L(p,\lambda,\mu)&=\sum_{i=1}^{N}p_i(f_i(x)+h(x))+\lambda(1-\sum_{i=1}^{N}p_i)+\sum_{i=1}^{N}\mu_i(q_i-p_i)\\&=\lambda+\mu^\top q+\sum_{i=1}^{N}p_i(f_i(x)+h(x)-\mu_i-\lambda).\end{aligned}$$

Then, from strong duality, (56) is equivalent to the following dual problem

$$\begin{aligned}&\min_{\substack{\lambda\in\mathbb{R}\\ \mu\in\mathbb{R}^N_+}}\max_{p\geq 0}\Big\{\lambda+\mu^\top q+\sum_{i=1}^{N}p_i(f_i(x)+h(x)-\mu_i-\lambda)\Big\}\\&\Leftrightarrow\min_{\substack{\lambda\in\mathbb{R}\\ \mu\in\mathbb{R}^N_+}}\lambda+\mu^\top q\\&\qquad\text{s.t. } f_i(x)+h(x)-\mu_i-\lambda\leq 0,\quad\forall i\in\{1,\dots,N\}.\end{aligned}$$

Thus, problem (6) is equivalent to

$$\begin{aligned}&\min_{\substack{x\in\mathcal{H},\,\lambda\in\mathbb{R}\\ \mu\in\mathbb{R}^N_+}}\lambda+\mu^\top q\\&\text{s.t. } f_i(x)+h(x)-\mu_i-\lambda\leq 0,\quad\forall i\in\{1,\dots,N\}\\&x\in\mathcal{Q}\end{aligned}\tag{58}$$

Case (i)(c): Here $\mathcal{P}=\{p\in\Delta_N\,:\,\mu_-\leq\sum_{i=1}^{N}p_i\xi_i\leq\mu_+\}$ and hence the Lagrangian function is

$$\begin{aligned}L(p,\lambda,\beta,\gamma)&=\sum_{i=1}^{N}p_i(f_i(x)+h(x))+\lambda\big(1-\sum_{i=1}^{N}p_i\big)+\beta\big(\mu_+-\sum_{i=1}^{N}p_i\xi_i\big)\\&\quad+\gamma\big(-\mu_-+\sum_{i=1}^{N}p_i\xi_i\big)\\&=\lambda+\beta\mu_+-\gamma\mu_-+\sum_{i=1}^{N}p_i(f_i(x)+h(x)-\lambda+(\gamma-\beta)\xi_i).\end{aligned}$$

Therefore, from strong duality, (56) is equivalent to the following dual problem

$$\begin{aligned}&\min_{\substack{\lambda\in\mathbb{R}\\ \beta\geq 0,\ \gamma\geq 0}} \max_{p\geq 0}\Big\{\lambda+\beta\mu_+-\gamma\mu_-+\sum_{i=1}^{N}p_i(f_i(x)+h(x)-\lambda+(\gamma-\beta)\xi_i)\Big\}\\ &\quad\Leftrightarrow \min_{\substack{\lambda\in\mathbb{R}\\ \beta\geq 0,\gamma\geq 0}} \lambda+\beta\mu_+-\gamma\mu_-\\ &\qquad \text{s.t. } f_i(x)+h(x)-\lambda+\gamma\xi_i-\beta\xi_i\leq 0,\quad \forall i\in\{1,\dots,N\}.\end{aligned}$$

Thus, problem (6) is equivalent to

$$\begin{aligned}&\min_{\substack{x\in\mathcal{H},\,\lambda\in\mathbb{R}\\ \beta\geq 0,\ \gamma\geq 0}} \lambda+\beta\mu_+-\gamma\mu_-\\ &\quad \text{s.t. } f_i(x)+h(x)-\lambda+\gamma\xi_i-\beta\xi_i\leq 0,\quad \forall i\in\{1,\dots,N\}\\ &\quad x\in\mathcal{Q}\end{aligned}\tag{59}$$

4. Applications

4.1. Couette inverse problem

Consider a rheometer of coaxial cylinders with fluid. Let $\Omega\colon\mathbb{R}\to\mathbb{R}$ representing the angular velocity of the inner cylinder. The function Ω has the following form [12]:

$$\Omega(\tau)=\frac{1}{2}\int_{\beta\tau}^{\tau}\frac{f(t)}{t}dt,$$

where τ is the tangential shear stress on the cylindrical surface, $f\colon\mathbb{R}\to\mathbb{R}$ is the rheological curve of the fluid, and $\beta<1$. Given $\Omega\colon\mathbb{R}\to\mathbb{R}$, the Couette inverse problem consists in to find the function f. Suppose that we have a sample of size r of values $(\Omega_i,\tau_i)_{i=1}^r$ representing measurements obtained from the rheometer. Assume that the function f has the form $f(t)=\sum_{j=1}^{\ell}a_je_j(t)$, where $a_j\in\mathbb{R}$ and e_j is a function such that $t\mapsto e_j(t)/t$ is integrable for all $j\in\{1,\dots,\ell\}$. Given $(\Omega_i,\tau_i)_{i=1}^r$, the problem is to find the constants $(a_j)_{j=1}^{\ell}$ such that

$$\Omega_i=\sum_{j=1}^{\ell}a_j\cdot\frac{1}{2}\int_{\beta\tau_i}^{\tau_i}\frac{e_j(t)}{t}dt,\quad \text{for all } i\in\{1,\dots,r\},$$

which is equivalent to $\Omega=Ax$, where $\Omega=(\Omega_i)_{i=1}^r\in\mathbb{R}^r$, $A\in\mathbb{R}^{r\times\ell}$ is the matrix defined by $A_{ij}=\frac{1}{2}\int_{\beta\tau_i}^{\tau_i}\frac{e_j(t)}{t}dt$, and $x=(a_j)_{j=1}^{\ell}$. It is possible that the system $Ax=\Omega$ has no solutions or that it has several solutions. In these cases, a common approach to finding an approximate solution is to solve the following regularized least squares problem

$$\min_{x\in\mathbb{R}^{\ell}}\|Ax-\Omega\|^2+\lambda R(x),\tag{60}$$

where R is the regularization function which represents a prior information on x and $\lambda>0$ is the regularization parameter. In the case when $R(x)=\|Dx\|^2$, where $D\in\mathbb{R}^{p\times\ell}$ and assuming that $A^\top A+\lambda D^\top D$ is positive definite (for example, if $D=\mathrm{Id}$), the Fermat's rule implies that the solution to (60) is

$$\overline{x}=(A^\top A+\lambda D^\top D)^{-1}A^\top\Omega.$$

Suppose now that we have N samples of size r of values $(\Omega_i^1)_{i=1}^r, \ldots, (\Omega_i^N)_{i=1}^r$. Let $\xi\colon \overline{\Omega} \to \Xi$ the random vector defined on a measurable space $(\overline{\Omega}, \mathcal{A})$ whose values are in the finite set $\Xi = \{\Omega^1, \ldots, \Omega^N\}$ and let $\mathcal{P} \subset \Delta_N$ a nonempty closed convex subset of probability measures on the space $(\overline{\Omega}, \mathcal{A})$ supported on Ξ. For every $\mathbb{P} \in \mathcal{P}$ and $k \in \{1, \ldots, N\}$, we denote $p_k = \mathbb{P}(\{\omega \in \overline{\Omega} : \xi(\omega) = \Omega^k\})$. Thus, we consider the following DRO problem

$$\min_{x\in\mathbb{R}^\ell} \sup_{p\in\mathcal{P}} \sum_{k=1}^N p_k \|Ax - \Omega^k\|^2 + \lambda\|Dx\|^2. \tag{61}$$

Note that, since $\mathcal{P} \subset \Delta_N$, problem (61) is equivalent to

$$\min_{x\in\mathbb{R}^\ell} \|Ax\|^2 + \lambda\|Dx\|^2 + \sup_{p\in\mathcal{P}} \sum_{k=1}^N p_k(-2x^\top A^\top \Omega^k + \|\Omega^k\|^2),$$

which is equivalent to problem (11) when $\mathcal{Q} = \mathbb{R}^\ell$, $h(x) = \|Ax\|^2 + \lambda\|Dx\|^2$ (which has Lipschitz gradient), $a_k = -2A^\top\Omega^k$, and $\xi_k = \|\Omega^k\|^2$.

Note that if $D = \mathrm{Id}$, then the problem (61) reduces to

$$\min_{x\in\mathbb{R}^\ell} \frac{1}{2\mu}\|x\|^2 + \sup_{p\in\mathcal{P}} \sum_{k=1}^N p_k(\langle x, Qx\rangle + \langle b_k, x\rangle + c_k), \tag{62}$$

where $Q = A^\top A$, $b_k = -2A^\top\Omega^k$, $c_k = \|\Omega^k\|^2$, and $\mu = \frac{1}{2\lambda}$. The problem (62) is equivalent to find $P_\mu \widetilde{f}(0)$, where $\widetilde{f}$ is defined in (27). Thus, (62) can be solved using the algorithm proposed in Proposition 2.6.

4.2. Denoising

Suppose that we have a noisy measurement of a signal $x \in \mathbb{R}^n$:

$$b = x + w,$$

where x is an unknown signal, w is an unknown noise vector, and b is the known measurement vector. The denoising problem is the following: Given b, find a good estimate of x. The regularized least squares problem associated with this problem is

$$\min_{x\in\mathbb{R}^n} \|x - b\|^2 + \lambda R(x), \tag{63}$$

where $R(x)$ is a regularization term which represents some a priori information on the signal and $\lambda > 0$ is a given regularization parameter. Typically, it is considered a regularization function of the form

$$R(x) = \sum_{i=1}^{n-1} (x_i - x_{i+1})^2.$$

Note that $R(x) = \|Lx\|^2$, where $L \in \mathbb{R}^{(n-1)\times n}$ is given by

$$L = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}.$$

In this case, by the Fermat's rule, we have that the solution to (63) is given by

$$\overline{x} = (\mathrm{Id} + \lambda L^\top L)^{-1} b.$$

Suppose that we have N noisy measurement $b^1, \dots, b^N$:

$$b^i = x^i + w^i,$$

where, for all $i \in \{1, \dots, N\}$, $x^i \in \mathbb{R}^n$ is an unknown signal and w^i is an unknown noise vector. The objective is to find a good estimate of x^i for all $i \in \{1, \dots, N\}$ in order to minimize the largest norm of the noise vectors. Thus, we consider the following robust optimization problem

$$\min_{\mathbf{x}=(x^1,\dots,x^N)\in\mathbb{R}^{nN}} \Big\{ \lambda \sum_{j=1}^{N} \|L_j x^j\|^2 + \max_{1\le i\le N} \|x^i - b^i\|^2 \Big\}, \tag{64}$$

where $\|L_j x^j\|^2$ represents the regularization term of the variable x^j. The problem (64) is a particular case of problem (10) when we have $V = \mathcal{Q}_i = \mathcal{H} = \mathbb{R}^n$, $H(\mathbf{x}) = \lambda \sum_{j=1}^N \|L_j x^j\|^2$, and $\xi_i = b^i$. Note that $\nabla H(\mathbf{x}) = 2\lambda(L_1^\top L_1 x^1, ..., L_N^\top L_N x^N)$. Then, for every $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^{nN}$, we have

$$\begin{aligned} \|\nabla H(\mathbf{x}) - \nabla H(\mathbf{y})\|^2 &= \sum_{j=1}^{N} \|2\lambda L_j^\top L_j (x^j - y^j)\|^2 \\ &\le (2\lambda)^2 \max_{1\le j\le N} \|L_j^\top L_j\|^2 \|\mathbf{x} - \mathbf{y}\|^2. \end{aligned}$$

Hence ∇H is β^{-1}-Lipschitz with $\beta^{-1} = 2\lambda \max_{1\le j\le N} \|L_j^\top L_j\|$.

5. Numerical experiments

Consider the problem (6) when $\mathcal{H} = \mathbb{R}^n$, $f_i(x) = \langle a_i, x\rangle + \xi_i$ for all $i \in \{1, \dots, N\}$, and $\mathcal{Q} = \{x \in \mathbb{R}^n : Ax = b\}$, where $a_i \in \mathbb{R}^n \backslash \{0\}$, $\xi_i \in \mathbb{R}$, $A \in \mathbb{R}^{m\times n}$ satisfies $\ker A^\top = \{0\}$, and $b \in \mathrm{ran}(A)$. We consider two cases for the function h. First, we consider $h(x) = \frac{1}{2} x^\top M x$, where $M \in \mathbb{R}^{n\times n}$ is symmetric positive definite. In this case, the problem (6) is

$$\min_{x\in\mathcal{Q}} \Big\{ \frac{1}{2} x^\top M x + \sup_{p\in\mathcal{P}} \sum_{i=1}^{N} p_i(\langle a_i, x\rangle + \xi_i) \Big\}. \tag{65}$$

Second, we consider $h(x) = c^\top x$, where $c \in \mathbb{R}^n$. In this case, problem (6) is

$$\min_{x\in\mathcal{Q}} \Big\{ c^\top x + \sup_{p\in\mathcal{P}} \sum_{i=1}^{N} p_i(\langle a_i, x\rangle + \xi_i) \Big\}. \tag{66}$$

We consider two ambiguity sets $\mathcal{P}$. First, we consider $\mathcal{P} = \Delta_N$. Second, we consider $\mathcal{P} = \widetilde{\mathcal{P}} := \{p \in \Delta_N : \mu_- \le \sum_{i=1}^N p_i \xi_i \le \mu_+\}$, where $\xi_i \in \mathbb{R}$, $\mu_- \in \mathbb{R}$, and $\mu_+ \in \mathbb{R}$ satisfies the conditions in (i)(c).

Note that, since M is positive definite, then $x \mapsto \frac{1}{2} x^\top M x$ is coercive and strictly convex. Thus, by [2, Proposition 11.15(i)], the problem (65) has a unique solution. On the other hand, we assume that problem (66) has solutions.

For solving problems (65) and (66) with $\mathcal{P} = \Delta_N$, we use the algorithms (35) (with $V = \mathcal{D}$, $\mathcal{Q}_i = \mathcal{Q}$, and $H(\mathbf{x}) = h(x_1)$), (46), (52), and (54), while that for solving problems (65) and (66) with $\mathcal{P} = \widetilde{\mathcal{P}}$, we use the algorithms (35), (47), (53), and (55). Note that in the context of problem (66), the equivalent formulations (57) and (59) reduce to a linear program, which can be solved using the function `linprog` of MATLAB. Thus, also we use the equivalent formulation (57) for solving (66) when $\mathcal{P} = \Delta_N$ and we use the formulation (59) for solving (66) when $\mathcal{P} = \widetilde{\mathcal{P}}$.

Table 1: Average execution time (number of iterations) with relative error tolerance $e = 10^{-5}$ for solving problem (65) with $\mathcal{P} = \Delta_N$.

(n, m, N)	prox max	distributed FB	FB with subspaces	Davis-Yin
$(100, 100, 10)$	2.224 (1137)	4.723 (3806)	7.892 (5880)	7.497 (5530)
$(100, 100, 50)$	3.163 (580)	17.317 (12849)	6.260 (3493)	5.377 (3314)
$(100, 100, 100)$	11.728 (1160)	67.524 (23416)	6.056 (2018)	6.828 (2086)
$(200, 200, 50)$	9.962 (580)	19.387 (12835)	2.231 (498)	1.539 (937)
$(200, 200, 100)$	41.647 (1160)	100.361 (23404)	7.065 (725)	8.887 (1722)

For each method, we obtain the average execution time (in seconds) and the average number of iterations from 20 random instances for the matrices A and M, the vectors $\{a_i\}$, b, and c, and the scalars $\{\xi_i\}$, μ_-, and μ_+ (we consider $\mu_- \in [0, 1/2]$ and $\mu_+ \in [1/2, 1]$). We measure the efficiency for different values of N, n, and m. We choose $n = m$. We label the algorithm in (35) as *prox max*, algorithms in (46) and (47) as *distributed FB*, algorithms in (52) and (53) as *FB with subspaces*, and algorithms in (54) and (55) as *Davis-Yin*. In addition, we label the method proposed in Section 3.5 for solving (66) as *dual method*. We choose to stop every algorithm when the norm of the difference between two consecutive iterations is less than 10^{-5} or the number of iterations exceeds 30000.

Table 2: Average execution time (number of iterations) with relative error tolerance $e = 10^{-5}$ for solving problem (65) with $\mathcal{P} = \widetilde{\mathcal{P}}$.

(n, m, N)	prox max	distributed FB	FB with subspaces	Davis-Yin
$(100, 100, 10)$	2.277 (1160)	5.417 (4057)	7.043 (4799)	6.242 (4304)
$(100, 100, 50)$	4.663 (580)	18.313 (13065)	6.312 (3055)	4.666 (2899)
$(100, 100, 100)$	24.465 (1160)	67.241 (23624)	5.823 (1915)	6.311 (2012)
$(200, 200, 50)$	10.312 (580)	20.348 (13050)	2.505 (542)	1.591 (948)
$(200, 200, 100)$	46.704 (1160)	86.954 (23613)	8.723 (868)	8.092 (1737)

With respect to problem (65), we observe that in the cases $(n, m, N) = (100, 100, 10)$ and $(n, m, N) = (100, 100, 50)$, the most efficient algorithm is the proximal algorithm, where in the case of problem (65) with $\mathcal{P} = \widetilde{\mathcal{P}}$ and $(n, m, N) = (100, 100, 50)$ the Davis-Yin's method has a similar performance with the proximal method.

Now, for the problem (65) with $\mathcal{P} = \Delta_N$, the forward-backward method with subspaces is the most efficient in the case $(n, m, N) = (100, 100, 100)$ and in the case $(n, m, N) = (200, 200, 100)$, while that in the case $(n, m, N) = (200, 200, 50)$ the fastest method is the Davis-Yin's formulation. On the other hand, for the problem (65) with $\mathcal{P} = \widetilde{\mathcal{P}}$, in the case $(n, m, N) = (100, 100, 100)$ the forward-backward algorithm with subspaces is the most efficient, while that in the case $(n, m) = (200, 200)$ the Davis-Yin's method has the best performance.

Table 3: Average execution time (number of iterations) with relative error tolerance $e = 10^{-5}$ for solving problem (66) with $\mathcal{P} = \Delta_N$.

(n, m, N)	prox max	distributed FB	FB with subspaces	Davis-Yin	Dual method
$(100, 100, 10)$	0.028 (11)	1.034 (813)	0.403 (277)	2.661 (1945)	0.057
$(100, 100, 50)$	0.065 (11)	13.310 (8586)	1.923 (1051)	13.107 (7515)	0.064
$(100, 100, 100)$	0.113 (10)	7.387 (2317)	5.189 (1943)	22.904 (7108)	0.059
$(200, 200, 50)$	0.164 (10)	1.920 (1256)	1.676 (1112)	8.792 (5632)	0.093
$(200, 200, 100)$	0.345 (11)	9.393 (2664)	7.844 (2022)	27.166 (6459)	0.105

Table 4: Average execution time (number of iterations) with relative error tolerance $e = 10^{-5}$ for solving problem (66) with $\mathcal{P} = \widetilde{\mathcal{P}}$.

(n, m, N)	prox max	distributed FB	FB with subspaces	Davis-Yin	Dual method
$(100, 100, 10)$	0.040 (11)	1.128 (864)	0.463 (301)	2.907 (2060)	0.058
$(100, 100, 50)$	0.204 (11)	10.441 (7558)	1.914 (1072)	11.944 (6647)	0.059
$(100, 100, 100)$	1.166 (10)	8.750 (2993)	5.130 (1965)	17.486 (5607)	0.061
$(200, 200, 50)$	0.256 (10)	3.659 (2328)	1.702 (1140)	6.885 (4257)	0.092
$(200, 200, 100)$	0.850 (10)	13.278 (3726)	7.992 (2041)	26.853 (5857)	0.107

With respect to the problem (66), we note that in the case $(n, m, N) = (100, 100, 10)$ the proximal algorithm is the most efficient for both ambiguity sets, while that in the other cases the dual method is the most efficient. Note that in the case $(n, m, N) = (100, 100, 50)$ with $\mathcal{P} = \Delta_N$, the proximal algorithm and the dual method have similar performance. In addition, the forward backward algorithm with subspaces also is an efficient alternative for solving problem (66).

6. Conclusions

In this paper, we provide different splitting algorithms for solving the discrete version of the distributionally robust optimization problem. This problem includes a supremum function in the objective function. The first method is based on calculating the proximity operator of said supremum function. In some cases, we propose an algorithm that converges to the proximity operator, while in a particular case, we provide a closed form for the proximity operator. On the other hand, under qualifications conditions, we prove that the problem is equivalent to solving a monotone

inclusion that involves the sum of finitely many monotone operators, and we compute the resolvent of the monotone operators involved in the inclusion. The second method uses an algorithm specialized for that type of inclusion, which is proposed in [1]. Additionally, we reformulate the inclusion as one that involves two monotone operators and the normal cone to a vector subspace. For solving this reformulation, we use the algorithm proposed in [4] (third method) and the algorithm proposed in [9] (fourth method).

The proposed algorithms can be applied to solve the Couette inverse problem with uncertainty and the denoising problem with uncertainty. In addition, we prove the efficiency of the algorithms in two particular problems. The first numerical experiment shows that the algorithm which computes the proximity of the supremum function (proximal algorithm) is the most efficient when the dimension of the problem and the size of the uncertainty set are small, whereas the third and fourth methods are more efficient when the dimension of the problem is larger. On the other hand, the second numerical experiment shows that the proximal algorithm is more efficient when the uncertainty set is small, while the method proposed in the literature performs better in other cases.

Acknowledgements. The work of Luis M. Briceño-Arias is supported by Centro de Modelamiento Matemático (CMM), FB210005, BASAL fund for centers of excellence, and grants FONDECYT 1230257 and MATHAmSud AMSUD230036 from ANID-Chile. The work of Sergio López-Rivera is supported ANID-Chile through Beca de Doctorado Nacional 21210951. The work of Emilio Vilches is supported by Centro de Modelamiento Matemático (CMM) and ANID-Chile through BASAL funds for Center of Excellence FB210005, Fondecyt Regular projects 1200283 and 1240120, Fondecyt Exploración 13220097, the ECOS-Anid Project ECOS230027 and the MathAmsud Project AMSUD230036.

References

[1] F. J. Aragón-Artacho, Y. Malitsky, M. K. Tam, D. Torregrosa-Belén: *Distributed forward-backward methods for ring networks*, Comput. Optim. Appl. 86/3 (2023) 845–870.

[2] H. H. Bauschke, P. L. Combettes: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., CMS Books in Mathematics, Springer, New York (2017).

[3] L. Briceño-Arias, P. L. Combettes: *A perturbation framework for convex minimization and monotone inclusion problems with nonlinear compositions*, Math. Oper. Res. 49/3 (2023) 1890–1914.

[4] L. Briceño-Arias: *Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions*, Optimization 64/5 (2015) 1239–1261.

[5] L. Briceño-Arias, J. Deride, S. López-Rivera, F. J. Silva: *A primal-dual partial inverse algorithm for constrained monotone inclusions: Applications to stochastic programming and mean field games*, Appl. Math. Optim. 87/2 (2023), art.no. 21, 36p.

[6] L. Briceño-Arias, S. López Rivera: *A projected primal-dual method for solving constrained monotone inclusions*, J. Optim. Theory Appl. 180/3 (2019) 907–924.

[7] P. L. Combettes, J. C. Pesquet: *Proximal splitting methods in signal processing*, in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*,

H. H. Bauschke et al. (eds.), Banff 2009, Springer Optimization and its Applications Vol. 49, Springer, New York (2011) 185–212.

[8] R. Cominetti, W. F. Mascarenhas, P. J. S. Silva: *A Newton's method for the continuous quadratic knapsack problem*, Math. Program. Comput. 6/2 (2014) 151–169.

[9] D. Davis, W. Yin: *A three-operator splitting scheme and its optimization applications*, Set-Valued Var. Analysis 25 (2017) 829–858.

[10] W. de Oliveira: *Risk-averse stochastic programming and distributionally robust optimization via operator splitting*, Set-Valued Var. Analysis 29/4 (2021) 861–891.

[11] J. Duchi, S. Shalev-Shwartz, Y. Singer, T. Chandra: *Efficient projections onto the l1-ball for learning in high dimensions*, in: Proc. 25th Int. Conf. on Machine learning (2008) 272–279.

[12] I. M. Krieger, S. H. Maron: *Direct determination of the flow curves of non-newtonian fluids*, J. Appl. Physics 23/1 (1952) 147–149.

[13] P. Pérez-Aros, E. Vilches: *Moreau envelope of supremum functions with applications to infinite and stochastic programming*, SIAM J. Optim. 31/3 (2021) 1635–1657.

[14] A. Shapiro, D. Dentcheva, A. Ruszczyński: *Lectures on Stochastic Programming – Modeling and Theory*, MOS-SIAM Series on Optimization Vol. 28, Society for Industrial and Applied Mathematics, Philadelphia (2021).

[15] X. A. Sun, A. J. Conejo: *Robust Optimization in Electric Energy Systems*, International Series on Opererations Research and Management Science Vol. 313, Springer, Cham (2021).

[16] W. Wiesemann, D. Kuhn, M. Sim: *Distributionally robust convex optimization*, Oper. Res. 62/6 (2014) 1358–1376.