Generalized Nash Equilibrium Problems with Partial Differential Operators: Theory, Algorithms, and Risk Aversion

Deborah Gahururu, Michael Hintermüller, Steven-Marian Stengl, and Thomas M. Surowiec

Abstract PDE-constrained (generalized) Nash equilibrium problems (GNEPs) are considered in a deterministic setting as well as under uncertainty. This includes a study of deterministic GNEPs with nonlinear and/or multi-valued operator equations as forward problems and PDE-constrained GNEPs with uncertain data. The deterministic nonlinear problems are analyzed using the theory of generalized convexity for set-valued operators, and a variational approximation approach is proposed. The stochastic setting includes a detailed overview of the recently developed theory and algorithms for risk-averse PDE-constrained optimization problems. These new results open the way to a rigorous study of stochastic PDE-constrained GNEPs.

Keywords Generalized Nash equilibrium problems · PDE-constrained optimization · L-convexity · Set-valued analysis · Fixed-point theory · Risk-averse optimization · Coherent risk measures · Stochastic optimization · Method of multipliers

---

D. Gahururu · T. M. Surowiec
Philipps-Universität Marburg, Faculty of Mathematics and Computer Science, Marburg, Germany
e-mail: gahururu@mathematik.uni-marburg.de; surowiec@mathematik.uni-marburg.de

M. Hintermüller (✉)
Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany
e-mail: michael.hintermueller@wias-berlin.de

S.-M. Stengl
Weierstrass Institute, Berlin, Germany
e-mail: steven-marian.stengl@wias-berlin.de

1 Introduction

Many applications and areas in science study phenomena sharing the common requirement of minimizing more than one objective simultaneously. In general, the solution of these problems has to address conflicting interests of the involved agents. Hence, we turn our attention to modeling a degree of competition and noncooperative behavior leading to *Nash games*. This concept has been successfully applied to a variety of applications in economics and in the context of networks, see [9, 11] and additionally [35] for the combinatorial branch of optimization. In many practical cases, the actions of the players in these games are restricted by equilibrium constraints establishing a reinforced linkage between the diverging interests. As we know from the mathematical treatment of optimal control and design problems, this coupling is usually resolved as an operator equation. However, in the context of partial differential equation (PDE)-constrained optimization, this type of concept has not yet been frequently studied.

We start by motivating N agent *games*. In this context, mathematically speaking, a set of N agents (or players) solve each an individual minimization problem to find their respective optimal strategy. For player i, this reads as

$$\text{minimize}_{u_i \in U_{\text{ad}}^i} \mathcal{J}_i(u_i, u_{-i}) \text{ over } u_i \in U_i,$$

where $U_{\text{ad}}^i \subset U_i$, with U_i a Banach space, is the set of *feasible strategies*. The functional $\mathcal{J}_i$ is specific for the player and involves his strategy u_i as well as the (given) strategies of all other players denoted as u_{-i}. Here and in the following, the combined vector of *all* strategies is usually denoted as $u = (u_i, u_{-i})$ without any permutation of components. A vector $u \in U$ with $U = U_1 \times \cdots \times U_N$ is called a *Nash equilibrium* if every strategy chosen by an agent is his optimal choice given the strategies of the other agents. This yields

$$u_i \in \text{argmin}_{u_i' \in U_{\text{ad}}^i} \left\{ \mathcal{J}_i(u_i', u_{-i}) \text{ over } u_i' \in U_i \right\} \text{ for all } i = 1, \ldots, N. \qquad (1.1)$$

The problem of finding such a strategy vector is then called a *Nash equilibrium problem* (NEP). In this setting, the influence of the other players' actions is limited to the objectives, whereas the strategy sets remain unchanged. Allowing the other players to also influence the set of feasible strategies leads to a set-valued strategy mapping $C_i : U_{\text{ad}}^{-i} \rightrightarrows U_{\text{ad}}^i$ in the underlying optimization problems. A Nash equilibrium is then a point $u \in U_{\text{ad}}$ with $U_{\text{ad}} = U_{\text{ad}}^1 \times \ldots \times U_{\text{ad}}^N$ satisfying

$$u_i \in \text{argmin}_{u_i' \in C_i(u_{-i})} \left\{ \mathcal{J}_i(u_i', u_{-i}) \text{ over } u_i' \in U_i \right\} \text{ for all } i = 1, \ldots, N.$$

Finding a solution for the latter type of problem is also known as *Generalized Nash equilibrium problem* (GNEP). Correspondingly, we assume the strategy mapping to be structured as

$$C_i(u_{-i}) = \left\{ u_i' \in U_{\text{ad}}^i : g(u_i', u_{-i}) \in K \right\}$$

with $g : U \to X$ and $K \subseteq X$ a nonempty, closed, convex subset of some Banach space X. In principle, it is possible to incorporate several mappings $g_i : U \to X_i$, but we want to keep our presentation concise. Therefore, in our context, a (GNEP) is given by

$$u_i \in \operatorname{argmin}\{\mathcal{J}_i(u_i', u_{-i}) \text{ subject to } u_i' \in U_{\text{ad}}^i \text{ and } g(u_i', u_{-i}) \in K\} \qquad (1.2)$$

for all $i = 1, \dots, N$. Concerning the general constraint, we are particularly interested in constraints on the *state* variable y that is generated through a continuous *solution mapping* $S : U \to Y$ involving the entirety of the players strategies via $y = S(u)$. Here, the set Y is again a Banach space. The origin of this operator might be a PDE or the minimization of an underlying parametrized optimization problem. Moreover, we assume in our setting that the players' objectives are separable of the type

$$\mathcal{J}_i(u_i, u_{-i}) = J_i^1(S(u_i, u_{-i})) + J_i^2(u_i).$$

Here J_i^1 only depends on the state, e.g., by a data-fitting, respectively, tracking-type term, and J_i^2 only on the control, e.g., in the form of a regularization or control cost. Note that by this setting a coupling between the players is established via the objectives. The dependence of the feasible sets occurs through the presence of a state constraint $\mathcal{G}(y) \in K$, which might stem from a physical or technical consideration. Hence, a (GNEP) in our setting has the general form

$$\begin{aligned} &\text{minimize}_{u_i, y} \quad J_i^1(y) + J_i^2(u_i) \text{ over } u_i \in U_i, \, y \in Y \\ &\text{subject to } u_i \in U_{\text{ad}}^i \text{ and } \mathcal{G}(y) \in K \text{ with } y = S(u_i, u_{-i}). \end{aligned} \qquad (1.3)$$

Here, the continuous mapping $\mathcal{G} : Y \to X$, together with the set K, models the state constraint, leading to the relation $g = \mathcal{G} \circ S$. This model is flexible enough to allow for a wide variety of different mathematical and practical applications. However, some aspects discussed hereafter are more conveniently described using the more abstract setting of (1.2) rather than (1.3). We will, hence, switch between these formulations keeping their formal relation in mind.

As previously mentioned, the operator S may originate from a broad variety of problems including (possibly nonlinear) PDEs, Vis, or complementarity problems. Throughout, we assume the solution mapping to be a singleton, meaning that given u the state $y = y(u)$ is unique. This does not need to be the case in general. Our model may thus be seen as closely related to multi-leader-follower games (MLFG) that are investigated within the scope of this report, as well.

Mathematical games involve a broad variety of challenges, including existence, characterization of equilibria via first-order systems, as well as numerical analysis and solvers. Moreover, in many applications, problem data are uncertain, occurring, e.g., as random parameters. This gives rise to risk-related formulations of the involved PDE-constrained minimization as well as (G)NEP. In this chapter, we study in particular risk-averse agents by modeling appropriate individual objectives.

2 Nash Games Involving Nonlinear Operator Equations

We study the following Nash game with a linear operator equations and compare [21]:

$$
\begin{aligned}
\text{minimize} \quad & J_i^1(y) + J_i^2(u_i) \text{ over } u_i \in U_i, \, y \in Y, \\
\text{subject to } & u_i \in U_{\text{ad}}^i \text{ and } Ay = b + Bu \text{ in } W.
\end{aligned}
\tag{2.1}
$$

Here, Y is as before, W a Banach space, $b \in W$ fixed, $A \in \mathcal{L}(Y, W)$ an invertible, bounded linear operator, and $B \in \mathcal{L}(U, W)$ a bounded, linear operator involving the strategies of all players at once. This motivates the solution operator $S(u) = A^{-1}(b + Bu)$ of the state equation $Ay = b + Bu$.

First, we study existence of an equilibrium of (2.1). Here, the coupling of the minimization problems of the individual agents prevents using a technique associated with a single minimization problem. Rather we need to invoke fixed-point theory for set-valued operators. For this, we reformulate (1.1) as

$$
u \in \mathcal{B}(u), \tag{2.2}
$$

with $\mathcal{B}(u) := \Pi_{i=1}^N \mathcal{B}_i(u_{-i})$, and $\mathcal{B}_i(u_{-i}) = \operatorname{argmin} \left\{ \mathcal{J}_i(u_i', u_{-i}) : u_i' \in U_{\text{ad}}^i \right\}$. Here, the *best response mapping* $\mathcal{B} : U_{\text{ad}} \rightrightarrows U_{\text{ad}}$ assigns to every given strategy the Cartesian product of all players' feasible strategies yielding the optimal value. The existence proof of a solution to (2.2) uses a result of Kakutani, Fan and Glicksberg:

Theorem 2.1 (cf. [14]) *Given a closed point-to-(nonvoid)-convex-set mapping $\Phi : Q \rightrightarrows Q$ of a convex compact subset Q of a convex Hausdorff linear topological space into itself, then there exists a fixed point $x \in \Phi(x)$.*

Two assumptions are crucial in the above theorem: (i) the convexity assumption on the values of the mapping and (ii) the compactness of the underlying set. In our situation, (i) becomes a topological condition regarding the set of minimizers for the players' optimization problems. This property is guaranteed when the (reduced) objective functional is convex. Concerning (ii), in finite dimensions, the compactness is guaranteed by closedness and boundedness. In our infinite-dimensional setting, however, this condition is usually not fulfilled with respect to the strong topology. Hence, we require a transition to the weak topology leading to a strengthened condition on the closedness of the graph of the operator.

In order to apply Theorem 2.1, let J_i^1, J_i^2 be convex, continuous, functionals. Moreover, let J_i^2 or S be completely continuous on their respective domains. Additionally, let U_i be a reflexive, separable Banach space and U_{ad}^i a nonempty, closed, and bounded subset of U_i. Then, the latter is also compact with respect to the weak topology. These conditions guarantee the existence of an equilibrium by applying the theorem.

We next come to the (GNEP) in [21], which reads

$$\text{minimize} \quad J_i^1(y) + J_i^2(u_i) \text{ over } u_i \in U_i, \, y \in Y,$$

$$\text{subject to } u_i \in U_{\text{ad}}^i \text{ and } y \in K \text{ with} \tag{2.3}$$

$$Ay = b + Bu \text{ in } W,$$

with a continuous embedding $Y \hookrightarrow X$. Let

$$C_i(u_{-i}) := \left\{ u_i' \in U_{\text{ad}}^i : S(u_i', u_{-i}) \in K \right\}$$

denote the associated set-valued strategy map, with C again the Cartesian product. This setting adds another difficulty to the existence proof, as we are now confronted with moving sets of feasible strategies. Hence, the selection of sequences in the range of the operator to prove the closedness property of the best response map becomes an issue. To address this challenge, we notice that the condition restricting the players' feasible strategies is the same for all players. Hence, one is able to formulate the overall set of feasible strategies as

$$\mathcal{F} = \{u \in U_{\text{ad}} : S(u) \in K\}.$$

It is worth noting that the set $\mathcal{F}$ characterizes the whole strategy mapping via

$$u_i' \in C_i(u_{-i}) \Leftrightarrow (u_i', u_{-i}) \in \mathcal{F}$$

for all $i = 1, \ldots, N$, which implies in particular $\text{Fix}(C) = \mathcal{F}$, where $\text{Fix}(\cdot)$ denotes the set of fixed points of a map. In fact, this observation applies already to the more general setting of (1.2) and allows us to introduce the strengthened solution concept of *variational equilibria*. It relates to a strategy vector $u \in \mathcal{F}$ solving the fixed-point problem

$$u \in \widehat{\mathcal{B}}(u), \tag{2.4}$$

with $\widehat{\mathcal{B}} : \mathcal{F} \to \mathcal{F}$, and $\widehat{\mathcal{B}}(u) = \text{argmin} \left\{ \sum_{i=1}^N \mathcal{J}_i(u_i', u_{-i}) \text{ over } u' \in \mathcal{F} \right\}$. In this formulation, only a *single* minimization process occurs. It is straightforward to prove that every variational equilibrium is also a Nash equilibrium. Consequently, providing existence for the operator $\widehat{\mathcal{B}}$ is sufficient. To apply Theorem 2.1, we note that due to the linearity of S the joint set of feasible strategies is convex as well. If a (GNEP) has in addition only convex objectives, then it is referred to as a *jointly convex Nash game*.

Nonlinear PDEs lead to an underlying operator equation of the type

$$A(y) = b + B(u) \text{ in } W,$$

with a nonlinear operator $A : Y \to W$ and again a bounded linear $B : U \to W$. Now the solution mapping $S : U \to Y$ is nonlinear. In contrast to the previously discussed case, convexity of the reduced objectives is not necessarily fulfilled. Of course, the same holds in the generalized case for values of the strategy set C as well as for the joint set of strategy vectors $\mathcal{F}$. Hence, the existence proof becomes a very delicate task. One option to proceed is the identification of combinations of objectives and operator equations that still guarantee the required convexity conditions. In this context, it is interesting to discuss the necessary structure first for mere optimization problems and then for Nash games. If not otherwise stated, the subsequent results of the following subsection will be made available in [19] together with their proofs.

2.1 On the Convexity of Optimal Control Problems Involving Nonlinear Operator Equations

In the following, we investigate *generalized* operator equations of the type

$$w \in A(y) \text{ in } W.$$

This setting allows us to treat also variational inequalities (VIs). Here, $w \in W$ is a given control and $y \in Y$ the associated state. To ensure well-posedness, we assume that the set-valued operator $A : Y \rightrightarrows W$ has a single-valued inverse $A^{-1} : W \to Y$ with the entire space W as its domain. Moreover, associated with Y and W, let $K \subseteq Y$, respectively, $K_W \subseteq W$ denote nonempty, closed, and convex cones. These cones induce preorder relations $\leq_K$ and $\leq_{K_W}$ on their respective spaces by $y_0 \leq_K y_1 :\Leftrightarrow y_1 - y_0 \in K$ for $y_0, y_1 \in K$ (and analogously for W). Using these relations, it is possible to generalize the convexity notion from functionals to operators, and further even to set-valued operators between Banach spaces, cf. [5, Subsection 2.3.5].

Definition 2.2 Let X_1, X_2 be topological vector spaces with $L \subseteq X_2$ a nonempty closed, convex cone inducing a preorder relation as described above. A set-valued mapping $\Phi : X_1 \rightrightarrows X_2$ is called *L-convex*, if for all $t \in (0, 1)$ and $x_0, x_1 \in X_1$ the relation

$$t\Phi(x_1) + (1 - t)\Phi(x_0) \subseteq \Phi(tx_1 + (1 - t)x_0) + L$$

holds. Additionally, Φ is called *L-concave* if it is $(-L)$-convex.

Our next aim is to identify conditions on the operator A that guarantee that the solution operator $A^{-1} : W \to Y$ is L-convex.

Theorem 2.3 *Let Y, W be Banach spaces, both equipped with closed and convex cones $L \subseteq Y$ and $L_W \subseteq W$, respectively. Let $A : Y \rightrightarrows W$ be a set-valued operator fulfilling the following assumptions:*

(i) *The operator A is L_W-concave in the sense of Definition 2.2.*
(ii) *The mapping $A^{-1} : W \to Y$ is single-valued with domain $\mathrm{dom}\, A = W$, and it is L_W-L-isotone (compare also to [4, Section 1.2]), i.e.,*

$$\text{for } w_1, w_0 \in W \text{ with } w_2 \geq_{L_W} w_1 \text{ it holds that } A^{-1}(w_2) \geq_L A^{-1}(w_1).$$

Then, the mapping $A^{-1} : W \to Y$ is L-convex.

We illustrate the previous Theorem 2.3 by two examples.

Example Let $d \in \mathbb{N}\backslash\{0\}$ and $D \subseteq \mathbb{R}^d$ be an open, bounded domain with Lipschitz boundary. Consider the operator

$$A(y) := -\Delta y + N(y) \tag{2.5}$$

on the Sobolev space $Y = H_0^1(D)$ with $W = H^{-1}(D)$. Let N be a superposition operator $N : L^2(D) \to L^2(D)$ induced by a concave, nondecreasing function on $\mathbb{R}$. We set $L := \left\{\varphi \in H_0^1(D) : \varphi \geq 0 \text{ a.e. on } D\right\}$ together with $L_W := L^+$ with

$$L^+ = \left\{\xi \in H^{-1}(D) : \langle \xi, \varphi \rangle_{H^{-1}, H_0^1} \geq 0 \text{ for all } \varphi \in H_0^1(D) \text{ with } \varphi \geq 0 \text{ a.e. on } D\right\}.$$

Then, A is L_W-concave: Indeed, let $t \in (0, 1)$ and $y_0, y_1 \in H_0^1(D)$ and $\varphi \in L$ be arbitrarily chosen; then we have

$$\langle t A(y_1) + (1 - t)A(y_0) - A(t y_1 + (1 - t)y_0), \varphi \rangle_{H^{-1}, H_0^1}$$
$$= (t N(y_1) + (1 - t)N(y_0) - N(t y_1 + (1 - t)y_0), \varphi)_{L^2(D)} \leq 0,$$

showing the concavity of A. Moreover, the operator A is invertible and isotone in the L_W-L-sense. The first property can be deduced from the monotonicity of the operator N together with the coercivity of the Laplacian. To see the latter, choose $w_0, w_1 \in W$ with $w_0 \leq_{L_W} w_1$, and let $y_0, y_1 \in Y$ be the solution of $w_j = A(y_j)$ for $j = 0, 1$. Testing the difference of the equations by $(y_0 - y_1)^+$ yields

$$0 \geq -\|\nabla(y_0 - y_1)^+\|_{L^2(D)}^2 - (N(y_0) - N(y_1), (y_0 - y_1)^+)_{L^2(D)}$$
$$= \langle A(y_1) - A(y_0), (y_0 - y_1)^+ \rangle_{H_0^1, H^{-1}} = \langle w_1 - w_0, (y_0 - y_1)^+ \rangle_{H_0^1, H^{-1}} \geq 0,$$

which implies $y_1 \geq y_0$ a.e. and hence the isotonicity of A^{-1}, which gives us finally the L-convexity of the solution operator A^{-1}.

In the previous example, (2.5) relates to semilinear elliptic PDEs and hence addresses a constraint that has been widely discussed in the optimal control literature (cf. [42] for a general overview and [7, 30] for more recent research activities). An extension to semilinear parabolic equations is possible; see, e.g., [31, Chapter 3, Section 2]. Theorem 2.3 can be applied to VIs as well; see [32, Lemma 4.1] for a first result. In contrast, here we provide a more general result.

Example Let Y be a reflexive vector lattice with order cone L, i.e., Y is a reflexive Banach space and L a nonempty, closed, and convex cone with $L \cap (-L) = \{0\}$, and consider an L^+-concave, semicontinuous, and strongly monotone operator $A : Y \rightrightarrows Y^*$. Moreover, assume A to be strictly T-monotone, i.e., $\langle A(y+z) - A(y), (-z)^+ \rangle < 0$ for z with $(-z)^+ \neq 0$. Let $M \subseteq Y$ be a nonempty, closed, convex set, and lower bounded, i.e., $M + L \subseteq M$ and for all $y_0, y_1 \in M$ and $\min(y_0, y_1) \in M$. Moreover, let $w \in Y^*$ be given. We consider the following VI:

$$\text{Find } y \in M \ : \ w \in A(y) + N_M(y).$$

Then, one can show that the associated solution operator $S : Y^* \to Y$ is L-convex.

These examples illustrate the power of the proposed concept, which allows us to next consider optimization problems of the type

$$\begin{aligned}
&\text{minimize} \quad J^1(y) + J^2(u) \text{ over } u \in U, y \in Y, \\
&\text{subject to } u \in U_{\text{ad}} \text{ and } y \in K \text{ with} \hspace{2cm} (2.6) \\
&\qquad b + Bu \in A(y) \text{ in } W,
\end{aligned}$$

which may represent a model for a single agent's decision process. In order to guarantee the convexity of (2.6), we assume the convexity of both parts J^1 and J^2, respectively. Additionally, we assume the isotonicity of J^1 on Y, i.e., $y_0 \leq_L y_1 \Rightarrow J^1(y_0) \leq J^1(y_1)$. Considering single-valuedness, the L-convexity of the solution operator $S(u) := A^{-1}(b + B(u))$ reads $S(tu_1 + (1-t)u_0) \leq_L tS(u_1) + (1-t)S(u_0)$. Hence, $J^1 \circ S$ is convex and so is the entire objective as well. For a nonempty, closed, convex set $K \subseteq Y$ with $K - L \subseteq K$, the indicator functional $i_K : Y \to [0, +\infty]$ is isotone and convex. Thus, the convexity of the set of feasible controls in (2.6) can be stated as the following intersection of closed, convex sets:

$$\{u \in U_{\text{ad}} : S(u) \in K\} = U_{\text{ad}} \cap \{u \in U : i_K(S(u)) \leq 0\}.$$

Under these conditions, the convexity of the optimization problem (2.6) is guaranteed. We illustrate this by the following optimization of doping profiles; cf. [28].

Example Let $D \subseteq \mathbb{R}^2$ be a given, bounded, open domain with Lipschitz boundary and $D_o \subseteq D$ an open subset. For a function $z \in L^2(\Omega)$, we denote $z^{2+} :=$

$\max(0, z)^2$. Consider

$$\min_{u \in U_{ad}} \frac{1}{2} \int_{D_o} (S(u) + 1)^{2+} dx + \frac{\alpha}{2} \int_D u^2 dx, \tag{2.7}$$

where $S : L^2(D) \to H^1(D)$ is the solution operator of the following PDE:

$$-\kappa \Delta y + \sinh(y) = -Bu - b \text{ in } D, \ \kappa \frac{\partial y}{\partial n} = 0 \text{ on } \partial D,$$

with B the (linear) solution operator of the PDE

$$-r \Delta d + d = u \text{ in } D, \ r \frac{\partial d}{\partial n} = 0 \text{ on } \partial D,$$

and $U_{ad} := \{u \in L^2(D) : 0 \leq u \leq 1 \text{ a.e. on } D\}$. Note that by the use of the Trudinger–Moser inequality (cf. [34]), the function $\sinh(y)$ lies in $L^2(D)$ for $y \in H_0^1(D)$. Assume further that $b \geq 0$ a.e. on D. Then the solution operator is L-convex. To see this, define the auxiliary operator $A : H^1(D) \to H^{-1}(D)$, $\langle A(y), w \rangle_{H^{-1}, H^1} := (\nabla y, \nabla w)_{L^2} + (N(y), w)_{L^2}$, with

$$N(y) = \begin{cases} y, & \text{if } y \geq 0 \\ \sinh(y), & \text{else} \end{cases}$$

as a superposition operator. Recalling the result corresponding to (2.5), we see that the operator N is induced by a monotone and concave function on $\mathbb{R}$. Hence, the solution map is L-convex. The solution operator of the auxiliary problem and S coincide, because both operators are sign preserving. Since $u \geq 0$ a.e. by feasibility, we get $Bu \geq 0$ a.e. and together with $b \geq 0$ a.e. on D the nonnegativity of the solutions. Hence, the operators $\sinh$ and N coincide. Thus, we see that S is indeed L-convex on U_{ad}. Moreover, the objective is convex and isotone yielding the convexity of (2.7).

We would now like to derive first-order optimality conditions for (2.6). For this purpose, we extend the subdifferential concept from convex and nonsmooth analysis to vector-valued operators. For an element $y^* \in L^+$ with

$$L^+ := \{z^* \in Y^* : \langle z^*, y \rangle \geq 0 \text{ for all } y \in L\},$$

we define the *subdifferential* of the solution operator $S : U \to Y$ in direction y^* as

$$\partial S(u)(y^*) := \partial \langle y^*, S(\cdot) \rangle (u). \tag{2.8}$$

Due to the L-convexity of S also the functional $u \mapsto \langle y^*, S(u) \rangle$ is convex. Hence, the above expression (2.8) is well defined and reads as a scalarizing formulation;

compare [33, Theorem 1.90]. Note that this object is closely linked to the (Fréchet) coderivative (cf. [33, Definition 1.32], which is defined for a set-valued operator $F : X_1 \rightrightarrows X_2$ as

$$D^* F(x_1, x_2)(x_2^*) := \left\{ x_1^* \in X_1^* : (x_1^*, -x_2^*) \in N_{\text{gph}(F)}(x_1, x_2) \right\},$$

where $N_{\text{gph}(F)}(x_1, x_2)$ denotes the (Fréchet) normal cone of $\text{gph}(F)$ in $(x_1, x_2) \in \text{gph}(F)$, the graph of F; see [5] for more details. In the case of a nonempty, closed, convex set, the Fréchet normal cone and its corresponding notion from convex analysis coincide. Using the mapping $S_L : U \rightrightarrows Y$ defined by $S_L(u) := S(u) + L$, we obtain for our notation in (2.8) the equivalent formulation

$$\partial S(u)(y^*) = \{u^* \in U^* : (u^*, -y^*) \in N_{\text{gph}(S_L)}(u, S(u))\},$$

where we use $y^* \in K^+$. This concept allows for the following type of chain rule. In its formulation, $\mathcal{D}$ denotes the set of arguments of a set-valued map with nonempty image, and core the core of a set; see, e.g., [5, Definition 2.72] and [6, Subsection 4.1.3] for definitions and details.

Theorem 2.4 *Let U, Y be Banach spaces, the latter one equipped with a closed, convex cone L. Let $f_2 : U \rightarrow \mathbb{R} \cup \{+\infty\}$ and $f_1 : Y \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex, proper, lower semicontinuous functionals, and moreover let f_1 be L-isotone. Let the operator $S : U \rightarrow Y$ be L-convex. Then, the functional $f_1 \circ S + f_2 : U \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex. Furthermore, consider $u \in \mathcal{D}(\partial f_2)$ with $S(u) \in \mathcal{D}(\partial f_1)$ and let one of the following two conditions hold:*

(i) *Let S be locally bounded and the following constraint qualification hold*

$$0 \in \text{core} \left(\text{dom} f_2 \times \text{dom} f_1 - \text{gph}(S) \right).$$

(ii) *Let S be semicontinuous and the following constraint qualification hold*

$$0 \in \text{core} \left(S \left(\text{dom} f_2 \right) - \text{dom} f_1 \right).$$

Then, the following chain rule holds for the subdifferential of the composed objective:

$$\partial(f_1 \circ S + f_2)(u) = \partial S(u) \left(\partial f_1(S(u)) \right) + \partial f_2(u).$$

The proposed chain rule in Theorem 2.4 as well as the proof and the other results of Sect. 2 will be made available in [19]. Using the functionals $f_2 = J^2 + i_{U_{\text{ad}}}$ and

$f_1 = J^1 + i_C$, we obtain the first-order system

$$
\begin{aligned}
-q &\in \partial J^2(u) + N_{U_{\mathrm{ad}}}(u), \\
y^* &\in \partial J^1(y) + N_K(y), \\
q &\in \partial S(u)(y^*).
\end{aligned}
\tag{2.9}
$$

Theorem 2.4 enables one to derive necessary and sufficient optimality conditions even for constraints involving PDEs, VIs, or complementarity problems admitting a nonsmooth solution operator. Of course, not all optimal control problems will fit into the above framework and might not meet the assumptions required in Theorem 2.1. Hence, it might be worthwhile investigating the use of more general fixed-point results. One possibility in this direction is the Eilenberg–Montgomery Theorem (cf. [8]) where a weaker topological assumption replaces convexity. The application of this result still requires a characterization of the solution set for the players' optimization problems. This, however, is ongoing research.

3 Nash Games Using Penalization Techniques

The direct application of the nonsmooth approach in the previous section may be delicate for many Nash games. We therefore draw our attention to a characterization of first-order conditions for (1.3) involving a continuously differentiable solution operator. Indeed, let $A : Y \to W$ be an invertible, continuously differentiable operator with an everywhere invertible derivative. In the following, let K denote a nonempty, closed convex *cone*, and $\mathcal{G}$ a constraint map. The first-order system for a Nash equilibrium of the game associated with

$$
\begin{aligned}
&\text{minimize } J_i^1(y) + J_i^2(u_i) \text{ over } u_i \in U_i, \, y \in Y \text{ subject to} \\
&\quad u_i \in U_{\mathrm{ad}}^i \text{ and } \mathcal{G}(y) \in K \text{ with} \\
&\quad A(y) = b + Bu
\end{aligned}
\tag{3.1}
$$

for $i = 1, \ldots, N$ can be derived by the proposition of a constraint qualification of Robinson–Zowe–Kurcyusz type (RZK) (see [45]). In this setting, it reads

$$
\left(D\mathcal{G}(y) \circ DA(y)^{-1} \circ B_i \right) U_{\mathrm{ad}}^i - K(\mathcal{G}(y)) = X \text{ for all } i = 1, \ldots, N.
\tag{3.2}
$$

The first-order system then becomes

$$
\begin{aligned}
0 &= \partial_i J_i^2(u_i) + B_i^* p_i + \lambda_i && \text{in } U_i^*, \\
A(y) &= b + Bu && \text{in } W, \\
DA(y)^* p_i &= \partial_y J_i^1(y) - D\mathcal{G}(y)^* \mu_i && \text{in } Y^*, \\
\lambda_i &\in N_{U_{\mathrm{ad}}^i}(u_i) && \text{in } U_i^*, \\
X^* \supseteq K^+ \ni \mu_i &\perp \mathcal{G}(y) \in K \subseteq X && \text{for all } i = 1, \dots, N.
\end{aligned}
\tag{3.3}
$$

In the case of a variational equilibrium, the single non-decoupling optimization process leads to a (possibly weaker) constraint qualification formulated as

$$
\left(D\mathcal{G}(y) \circ DA(y)^{-1} \circ B \right) U_{\mathrm{ad}}(u) - K(\mathcal{G}(y)) = X.
\tag{3.4}
$$

This leads to a special instance of (3.3) where all multipliers $\mu_i \in X^*, i = 1, \dots, N$, coincide, i.e., $\mu_i = \mu$ for all $i \in \{1, \dots, N\}$ in (3.3). In many situations involving function spaces, higher regularity of the state is needed to guarantee the constraint qualification. This on the other hand leads to a reduced regularity of the multiplier(s) $\mu_{(i)}$ and subsequently also of the adjoint states p_i in practice. The above results of the subsequent ones in this section can be found in [18], if not stated otherwise.

3.1 Γ-Convergence

Next we use the notion of Γ-*convergence* to approximate our state-constrained Nash game by a sequence of simpler Nash games with a weakened form of the state constraint.

First we introduce a unified view on the different notions of equilibria discussed here.

Definition 3.1 Let a Banach space U and a functional $\mathcal{E} : U \times U \to \overline{\mathbb{R}}$ be given. A point $u \in U$ is called equilibrium, if

$$
\mathcal{E}(u, u) \leq \mathcal{E}(u', u) \text{ holds for all } u' \in U.
$$

The first component in the functional fulfills the task of a control variable, whereas the second one acts as a parameter and hence establishes a feedback mechanism. Note that the dependence of the domain of the reduced functional $\mathcal{E}(\cdot, u)$ on u is possible. Recalling the definition of the strategy mapping C as $C(u) = \prod_{i=1}^N C_i(u_{-i})$ with $C_i(u_{-i}) = \{u_i' \in U_{\mathrm{ad}}^i : g(u_i', u_{-i}) \in K\}$ and $g = \mathcal{G} \circ S$ as the composition of state constraint and solution operator, we reobtain by the choice

of functionals

$$
\mathcal{E}(u', u) = \sum_{i=1}^{N} \mathcal{J}_i(u'_i, u_{-i}) + i_{C(u)}(u') = \sum_{i=1}^{N} \left(\mathcal{J}_i(u'_i, u_{-i}) + i_{C_i(u_{-i})}(u'_i) \right)
$$

$$
= \sum_{i=1}^{N} \left(\mathcal{J}_i(u'_i, u_{-i}) + i_{U^i_{\mathrm{ad}}}(u_i) + i_K(g(u'_i, u_{-i})) \right)
$$

$$(3.5)$$

and

$$
\widehat{\mathcal{E}}(u', u) = \sum_{i=1}^{N} \mathcal{J}_i(u'_i, u_{-i}) + i_{\mathcal{F}}(u')
$$

$$
= \sum_{i=1}^{N} \left(\mathcal{J}(u'_i, u_{-i}) + i_{U^i_{\mathrm{ad}}}(u'_i) \right) + i_K(g(u'_i, u_{-i}))
$$

$$(3.6)$$

the notion of Nash, respectively, variational equilibria. Our aim now is to generalize Γ-convergence to equilibrium problems of the above form.

Definition 3.2 Let U be a Banach space and let $\mathcal{T}$ denote either the strong or weak topology on U. A sequence of functionals $\mathcal{E}_n : U \times U \to \overline{\overline{\mathbb{R}}}$ is called Γ-*convergent* to a functional $\mathcal{E} : U \times U \to \overline{\overline{\mathbb{R}}}$ if the following two conditions hold:

(i) For all sequences $u_n \xrightarrow{\mathcal{T}} u$, it holds $\mathcal{E}(u, u) \leq \liminf_{n\to\infty} \mathcal{E}_n(u_n, u_n)$.
(ii) For all $u' \in U$ and all sequences $u_n \xrightarrow{\mathcal{T}} u$, there exists a sequence $u'_n \xrightarrow{\mathcal{T}} u'$ such that $\mathcal{E}(u', u) \geq \limsup_{n\to\infty} \mathcal{E}_n(u'_n, u_n)$.

Of course, it is as well possible to combine the strong and weak topology in Definition 3.2. Note that the classical notion of Γ-convergence for a minimization problem is a special case of the above. The following convergence result holds true.

Proposition 3.3 *Let $\mathcal{E}_n$ be a Γ-convergent sequence of functionals as in Definition 3.1 with limit $\mathcal{E}$. Then, every accumulation point of a sequence of corresponding equilibria $(u_n)_{n\in\mathbb{N}}$ is an equilibrium of the limit.*

Our intention is to address the state constraint by applying a *penalization technique*. Therefore, the constraint $g(u) \in K$ encoded in the indicator function is substituted by a continuously differentiable *penalty function* $\beta : X \to [0, +\infty)$,

$$
\beta(x) = 0 \text{ if and only if } x \in K,
$$

scaled by a penalty parameter $\gamma > 0$. This leads to the formulation of the penalized functionals corresponding to the (GNEP) as

$$\mathcal{E}_\gamma(v, u) = \sum_{i=1}^{N} \Big(\mathcal{J}_i(v_i, u_{-i}) + \gamma \beta(g(v_i, u_{-i})) \Big) + i_{U_{\mathrm{ad}}}(v),$$

as well as to the variational equilibrium problem

$$\widehat{\mathcal{E}}_\gamma(v, u) = \sum_{i=1}^{N} \mathcal{J}_i(v_i, u_{-i}) + \gamma \beta(g(v)) + i_{U_{\mathrm{ad}}}(v).$$

Using the definition of the state as well as the composition $g = \mathcal{G} \circ S$, this leads to the *penalized Nash game*

$$\begin{aligned} \text{minimize} \quad & J_i^1(u_i) + J_i^2(y) + \gamma \beta(\mathcal{G}(y)) \text{ over } u_i \in U_i, \, y \in Y \\ \text{subject to } & u_i \in U_{\mathrm{ad}}^i \text{ with } A(y) = b + Bu, \end{aligned} \quad (3.7)$$

and in a similar fashion to the *penalized variational equilibrium problem*

$$\text{minimize} \quad \sum_{i=1}^{N} \Big(J_i^1(y_i) + J_i^2(u_i') \Big) + \gamma \beta(\mathcal{G}(y)) \text{ over } u_i' \in U_i, \, y_i \in Y \text{ and } y \in Y$$

subject to $u_i' \in U_{\mathrm{ad}}^i$ and $A(y_i) = b + B(u_i', u_{-i})$ as well as

$$A(y) = b + Bu'. \quad (3.8)$$

The definition of the states y_i and y comes from the presence of the terms $S(u_i', u_{-i})$ in the state-related functionals J_i^1 and of the expression $S(u')$ occurring in $\beta \circ \mathcal{G}$ for the penalization of the constraint $u' \in \mathcal{F}$. Moreover, we assume in the terms of the abstract setting (1.2) that the functionals $u \mapsto \mathcal{J}_i(u_i, u_{-i})$ are continuous with respect to the strong topology on U_i and the weak one on U_{-i}, i.e., for all sequences $u_i^n \to u_i$ and $u_{-i}^n \rightharpoonup u_{-i}$ it holds that $\mathcal{J}_i(u_i^n, u_{-i}^n) \to \mathcal{J}_i(u_i, u_{-i})$. This condition can usually be guaranteed for a wide variety of applications as in the setting of (1.3) by complete continuity of the solution map S together with continuity of the mappings J_1^i on Y and J_2^i on U_i. With these conditions at hand, it is possible to derive the Γ-convergence of (3.6) and by proposing $\mathrm{dom}(C) = U_{\mathrm{ad}}$ also the Γ-convergence of (3.5).

Turning to the derivation of a first-order system for the penalized problems, we assume for convenience that $J_i^1, J_i^2, i = 1, \ldots, N$, are all continuously

differentiable. In *both equilibrium cases*, this leads to the following system:

$$
\begin{aligned}
0 &= \partial_i J_i^2(u_i) + B_i^* p_i + \lambda_i & &\text{in } U_i^*, \\
A(y) &= b + Bu & &\text{in } W, \\
DA(y)^* p_i &= DJ_i^1(y) - D\mathcal{G}(y)^* \mu & &\text{in } Y^*, \\
\lambda_i &\in N_{U_{\mathrm{ad}}^i}(u_i) & &\text{in } U_i^*, \\
\mu &= -\gamma D\beta(\mathcal{G}(y)) & &\text{in } X^*.
\end{aligned}
\tag{3.9}
$$

In fact, for a jointly convex game, the first-order system would not only be necessary, but also sufficient implying the equivalence of the two penalized equilibrium problems. Assuming for the moment that at least the functionals J_i^2 are strongly convex, we find the strong monotonicity of the first derivative $\partial_i J_i^2 : U_i \to U_i^*$ and hence the unique solvability of the VI

$$
\text{Find } u_i \in U_i \ : \ u_i^* \in \partial J_i^2(u_i) + N_{U_{\mathrm{ad}}}(u_i),
$$

given an arbitrary $u_i^* \in U_i^*$. This problem admits a Lipschitz-continuous solution operator denoted by $P_i : U_i^* \to U_i$. In the simplest case of $J_i^2(u_i) = \frac{1}{2}\|u_i\|_{U_i}^2$ for a separable Hilbert space U_i, this map reads as a composition with the projection mapping on U_{ad}. Often, the system can be rewritten as a fixed-point problem

$$
u = T(u)
$$

with $T : U_{\mathrm{ad}} \to U_{\mathrm{ad}}$ defined by $T(u) = (T_1(u), \ldots, T_N(u))$ and

$$
T_i(u) = P_i(-B_i^* p_i) \text{ with } p_i = p_i(y) = DA(y)^{-*}\left(\partial_y J_i^1(y) + \gamma D\mathcal{G}(y)^* D\beta(\mathcal{G}(y))\right),
$$

and $y = S(u) = A^{-1}(b + Bu)$. Since this is a fixed-point problem involving only a single-valued operator—in contrast to the formulation for Nash and variational equilibria—the existence question does not suffer from a lack of topological characterization of its values and can thus be treated with classical Schauder-type results, cf. [44, Theorem IV.7.18]. Using the described penalization technique, one is hence able to propose a generalized solution concept that is also suitable for a numerical treatment of the state constraint by motivating a *path-following technique*. The idea is to observe the solution(s) of the above first-order system for a range of penalty parameters $\gamma \in [\gamma_{\min}, +\infty)$ leading to the *path*

$$
\begin{aligned}
\mathcal{P} = \Big\{ &(\gamma, u^\gamma, y^\gamma, p^\gamma, \mu^\gamma, \lambda^\gamma) \in [\gamma_{\min}, +\infty) \times U \times Y \times (W^*)^N \times X \times U^* \\
&\text{such that } (u^\gamma, y^\gamma, p^\gamma, \mu^\gamma, \lambda^\gamma) \text{ solves } (3.9) \Big\}.
\end{aligned}
$$

From the numerical viewpoint, it is interesting to study the behavior of the solutions of (3.9) for $\gamma \to +\infty$. As a first step toward a path analysis, we study the boundedness of the path. This is next done in the fully abstract setting only.

Lemma 3.4 *Let the mappings $v \mapsto \partial_i \mathcal{J}_i(v_i, v_{-i})$ (in the fully abstract setting) be bounded for all $i = 1, \ldots, N$ (i.e., images of bounded sets are bounded). If additionally the RZK condition (3.4) holds, then the path $\mathcal{P}$ is bounded.*

Using this result, it is straightforward to utilize reflexivity and the Banach–Alaoglu theorem to obtain the existence of weakly and weakly* converging subsequences. The next result guarantees that the corresponding limits are the desired solutions.

Theorem 3.5 *Let the condition (3.4) as well as the boundedness condition of Lemma 3.4 be fulfilled, and let moreover the following additional assumptions hold:*

(i) *The first derivatives of the objectives $\mathcal{J}_i$ with respect to the players' strategy satisfy for every weakly convergent sequence $u_i^n \rightharpoonup u_i^*$ in U the property*

$$\langle \partial_i \mathcal{J}_i(u_i^*, u_{-i}^*), u_i^* \rangle_{U_i, U_i^*} \leq \limsup_{n \to +\infty} \langle \partial_i \mathcal{J}_i(u_i^n, u_{-i}^n), u_i^n \rangle_{U_i, U_i^*}.$$

(ii) *The mapping $g : U \to X$ is strongly continuous and uniformly Fréchet differentiable on every bounded set, i.e., on very bounded subset $M \subseteq U$ holds that*

$$\lim_{\|h\|_X \to 0} \sup_{u \in M} \frac{\|g(u + h) - g(u) - Dg(u)h\|_X}{\|h\|_U} = 0.$$

Then, every path has a limiting point $(u^*, q^*, \lambda^*, \mu^*)$ along a subsequence, and every limiting point fulfills the necessary first-order condition for a Nash equilibrium (resp. variational equilibrium).

Together with the existence for solutions to the first-order system for the penalized system (3.9), the combined fulfillment of the conditions guarantees the existence of a point fulfilling the first-order system for (VEP) and hence especially for (GNEP).

This procedure sketches the numerical treatment of the (GNEP) problem (2.4). Besides identifying a suitable algorithm to solve the system (3.9), also an adaptive parameter update technique is needed; compare [21] for the latter. Here take a highly related approach leading to the definition of the *value functions*

$$\mathcal{W}_\gamma(u^\gamma) = \inf_{u' \in U_{ad}} \mathcal{E}_\gamma(u', u^\gamma) = \inf_{u' \in U_{ad}} \sum_{i=1}^{N} \left(\mathcal{J}_i(u_i', u_{-i}^\gamma) + \gamma \beta(g(u_i', u_{-i}^\gamma)) \right)$$

$$= \sum_{i=1}^{N} \inf_{u_i' \in U_{ad}^i} \left(\mathcal{J}_i(u_i', u_{-i}^\gamma) + \gamma \beta(g(u_i', u_{-i}^\gamma)) \right)$$

$$(3.10)$$

and analogously for the penalized (VEP)

$$\widehat{\mathcal{W}}_\gamma(u^\gamma) = \inf_{u' \in U_{\mathrm{ad}}} \left(\sum_{i=1}^N \mathcal{J}_i(u'_i, u^\gamma_{-i}) + \gamma \beta(g(u')) \right). \tag{3.11}$$

One observes that $\mathcal{E}_\gamma(u^\gamma, u^\gamma) - \mathcal{W}_\gamma(u^\gamma) \geq 0$ and $\widehat{\mathcal{E}}_\gamma(u^\gamma, u^\gamma) - \widehat{\mathcal{W}}_\gamma(u^\gamma) \geq 0$, with equality only if u^γ is a solution of the penalized Nash game, respectively, (VEP). Using the defined value functionals, we seek to evaluate the effect of an increase of γ on the behavior of our solution. Therefore, we consider the functional $\tilde{\gamma} \mapsto \mathcal{W}_{\tilde{\gamma}}(u_\gamma)$ respectively $\tilde{\gamma} \mapsto \widehat{\mathcal{W}}_{\tilde{\gamma}}(u_\gamma)$. For a local description of the behavior, we extract first-order information by providing bounds for the upper and lower limits for the directional derivative of the proposed functionals.

Lemma 3.6 *Let J_i^1, J_i^2 be continuous functionals, and let the best response mapping with respect to the penalty parameter $\tilde{\gamma}$, i.e.,*

$$\tilde{\gamma} \mapsto \mathcal{B}^{\tilde{\gamma}}(u^\gamma) = \mathrm{argmin}\left\{ \sum_{i=1}^N \left(\mathcal{J}_i(u'_i, u^\gamma_{-i}) + \tilde{\gamma} \beta(g(u'_i, u^\gamma_{-i})) \right) \text{ over } u' \in U_{\mathrm{ad}} \right\} \text{ and}$$

$$\tilde{\gamma} \mapsto \widehat{\mathcal{B}}^{\tilde{\gamma}}(u^\gamma) = \mathrm{argmin}\left\{ \sum_{i=1}^N \mathcal{J}_i(u'_i, u^\gamma_{-i}) + \tilde{\gamma} \beta(g(u')) \text{ over } u' \in U_{\mathrm{ad}} \right\}$$

be nonempty-valued. Let $u^\gamma \in U_{\mathrm{ad}}$ be an equilibrium for the penalized (GNEP) in (3.7), respectively, (VEP) in (3.8). Then, the difference quotients satisfy

$$0 \leq \liminf_{\eta \searrow 0} \frac{\mathcal{W}(\gamma + \eta) - \mathcal{W}(\gamma)}{\eta} \leq \limsup_{\eta \searrow 0} \frac{\mathcal{W}(\gamma + \eta) - \mathcal{W}(\gamma)}{\eta} \leq N\beta(g(u^\gamma)) \text{ and}$$

$$0 \leq \liminf_{\eta \searrow 0} \frac{\widehat{\mathcal{W}}(\gamma + \eta) - \widehat{\mathcal{W}}(\gamma)}{\eta} \leq \limsup_{\eta \searrow 0} \frac{\widehat{\mathcal{W}}(\gamma + \eta) - \widehat{\mathcal{W}}(\gamma)}{\eta} \leq \beta(g(u^\gamma)).$$

If, moreover, the best response map $\tilde{\gamma} \mapsto \mathcal{B}^{\tilde{\gamma}}(u^\gamma)$, respectively, $\tilde{\gamma} \mapsto \widehat{\mathcal{B}}^{\tilde{\gamma}}(u^\gamma)$, is single-valued and continuous, then the functional $\mathcal{W}$, respectively $\widehat{\mathcal{W}}$, is even differentiable with $\mathcal{W}'(\gamma) = N\beta(g(u^\gamma))$, respectively $\widehat{\mathcal{W}}'(\gamma) = \beta(g(u^\gamma))$.

Hence, the composition of the penalty and the state constraint serves as a way to adjust the penalty parameter for each step of the path-following procedure by

$$\gamma \mapsto \gamma + \max\left(\frac{\pi_{\mathrm{path}}}{\beta(g(u^\gamma))}, \varepsilon \right)$$

with a fixed parameter $\pi_{\mathrm{path}} > 0$. Using this technique, strong violations of the state constraint resulting in a big penalty term induce a more timid update, whereas low values cause a more aggressive behavior. The update is safeguarded with a fixed

upper bound $\varepsilon > 0$ for the case of very low values of the penalty functional. If the value is zero, then the algorithm terminates since it has found a solution of the original (GNEP), respectively, (VEP). The results of Sect. 3 together with the corresponding proofs and details will be made available in [18].

With this outline of an algorithm, we end the discussion of deterministic Nash equilibria and turn our attention to the case involving uncertainties.

4 PDE-Constrained GNEPs Under Uncertainty

4.1 Motivation

Most real-world problems in the natural sciences, engineering, economics, and finance are subject to uncertainty. This inherent stochasticity arises from a number of unavoidable factors, which range from noisy measurements and data acquisition to ambiguity in the choice of model and its underlying exogenous parameters. Consequently, we must incorporate random parameter into our mathematical models. Within the framework of PDE-constrained decision problems, we are then confronted with the task of optimizing systems of random partial differential equations.

In order to ensure these new infinite-dimensional stochastic decision problems yield robust solutions to outliers or potentially catastrophic events, we appeal to the theory of risk-averse optimization, which has been widely developed over the last several decades within the (finite dimensional) stochastic programming community, see, e.g., [41] and many references therein. Furthermore, using risk models in the context of Nash equilibrium problems allows us to model the preferences of the agents more accurately by assuming they have well-defined risk preferences.

Nevertheless, the literature on risk-averse PDE-constrained optimization was extremely scarce until recently [13, 24–28]. Therefore, in order to tackle risk-averse PDE-constrained GNEPs, it has been necessary to first develop the theory, approximation, and algorithms for the optimization setting. These results can now be leveraged for the NEP and ultimately GNEP setting.

In what follows, we will first present the recent theory of risk-averse PDE-constrained optimization in which the risk preferences of the individual agents are modeled by convex risk measures. Following this, we will apply the theory to a model risk-averse PDE-constrained Nash equilibrium problem. This will more clearly delineate the differences between the optimization and game-theoretic frameworks. We then present the recent approach in [25] for smoothing nonsmooth risk measures that is interesting from a theoretical perspective, but also useful for gradient-based optimization algorithms. In particular, we will see that epi-regularization of risk measures is an essential component of the primal–dual risk minimization algorithm recently developed in [27].

4.2   Additional Notation and Preliminary Results

In addition to the notation introduced above, we recall several further concepts necessary for the coming discussions. Unless otherwise stated, these are considered standing assumptions in the text below.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space where Ω is an arbitrary set of outcomes, $\mathcal{F} \subseteq 2^{\Omega}$ is the associated σ-algebra of events, and the set function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure. We employ the standard abbreviations "a.e." and "a.a." for "almost everywhere" and "almost all" with respect to $\mathbb{P}$, respectively. If necessary, we will append these by $\mathbb{P}$ and write $\mathbb{P}$-a.e. or $\mathbb{P}$-a.a. As $\mathcal{F}$ is fixed, we write "$\mathcal{F}$-measurable" simply as "measurable" if clear in context. Since we will often deal with Banach-space-valued random terms, we recall that a random element X in a Banach space $\mathcal{X}$ is a measurable mapping $X : \Omega \to \mathcal{X}$, where $\mathcal{X}$ is endowed with the Borel σ-algebra. We denote expectation by $\mathbb{E}[X]$.

We assume that the control space U is a real reflexive Banach space and denote the set of admissible decisions by $U_{\mathrm{ad}} \subset U$. The latter is assumed to be a nonempty, closed, and convex set. In the context of Nash equilibrium problems, U_{ad} is assumed to be bounded as well. The physical domain for the deterministic PDE solutions will be denoted by $D \subset \mathbb{R}^d$. We assume that D is an open and bounded set with Lipschitz boundary ∂D. The associated state space for the deterministic solutions will be denoted by $V := H^1(D)$ (or $H_0^1(D)$), where $H^1(D)$ is the usual Sobolev space of $L^2(D)$-functions with weak derivatives in $L^2(D)$ [1].

The natural function-space setting for solutions of random PDEs is in classical Bochner spaces, cf. [17]. We recall that the Bochner space $L^p(\Omega, \mathcal{F}, \mathbb{P}; W)$ comprises all measurable functions that map Ω into some Banach space W with p finite moments for $p = [1, \infty)$. When $p = \infty$, $L^\infty(\Omega, \mathcal{F}, \mathbb{P}; W)$ is the space of all essentially bounded W-valued measurable functions. The norms are given by

$$\|v\|_{L^p(\Omega, \mathcal{F}, \mathbb{P}; W)} = \mathbb{E}\left[\|v\|_W^p\right]^{1/p} \text{ for } p \in [1, \infty)$$

$$\|v\|_{L^\infty(\Omega, \mathcal{F}, \mathbb{P}; W)} = \operatorname*{ess\,sup}_{\omega \in \Omega} \|v(\omega)\|_W.$$

When $W = \mathbb{R}$, we set $L^p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}) = L^p(\Omega, \mathcal{F}, \mathbb{P})$. In our optimization and equilibrium settings, the random objective maps U into $\mathcal{X} := L^p(\Omega, \mathcal{F}, \mathbb{P})$ for some $p \in [1, \infty)$. Whenever it is clear, we simply write $\mathcal{X}$.

As discussed in Sect. 4.1, we model risk-averse behavior by means of risk measures. There is a vast literature on the subject of risk measures and their usage in optimization. In our models, the individual agents' problems are assumed to take the form:

$$\min_{u \in U_{\mathrm{ad}}} \mathcal{R}[\mathcal{J}(S(u))] + \wp(u),$$

where $\mathcal{R}$ is a nonlinear, typically nonsmooth, functional on $\mathcal{X}$. We refer the interested reader to [41, Chap. 6.] and the references therein as a starting point. For our purposes, it will suffice to introduce two general classes of risk measures here, each of which follows the standard axiomatic approach as in [3, 12, 39]. We start by recalling the definition of a regular measure of risk as suggested by Rockafellar and Uryasev in [39]. The conditions below were postulated as minimal regularity properties for risk measures in the context of optimization. A functional $\mathcal{R} : \mathcal{X} \to \overline{\mathbb{R}}$ where $\overline{\mathbb{R}} := (-\infty, \infty]$ is a regular measure of risk provided is proper, closed, convex and satisfies $\mathcal{R}[C] = C$ for all constant random variables $C \in \mathbb{R}$, and $\mathcal{R}$ is risk averse: $\mathcal{R}[X] > \mathbb{E}[X]$ for all nonconstant $X \in \mathcal{X}$. Therefore, the expected value is not a regular measure of risk in this setting. This is reasonable from the perspective that setting $\mathcal{R} = \mathbb{E}$ would indicate neutrality to risk and not yield a robust solution.

Perhaps the most well-known risk measures are the coherent risk measures. These were introduced in a systematic way in [3] as a means of axiomatizing the behavior of risk-averse decision makers. The risk measure $\mathcal{R}$ is coherent provided:

(C1) *Subadditivity*: If $X, X' \in \mathcal{X}$, then $\mathcal{R}[X + X'] \leq \mathcal{R}[X] + \mathcal{R}[X']$.
(C2) *Monotonicity*: If $X, X' \in \mathcal{X}$ and $X \geq X'$ almost surely, then $\mathcal{R}[X] \geq \mathcal{R}[X']$.
(C3) *Translation equivariance*: If $C \in \mathbb{R}$ and $X \in \mathcal{X}$, then $\mathcal{R}[X + C] = \mathcal{R}[X] + C$.
(C4) *Positive homogeneity*: If $C \in [0, \infty)$ and $X \in \mathcal{X}$, then $\mathcal{R}[CX] = C\mathcal{R}[X]$.

A rather popular coherent risk measure is the conditional or average value at risk (CVaR or AVaR). Given a risk or confidence level $\beta \in (0, 1)$, the average value at risk of a random variable X is the average of the associated quantiles $F_\alpha^{-1}(X)$ over $\alpha \in (\beta, 1)$. Here, we have

$$F_\alpha^{-1}(X) = \text{VaR}_\beta(X) := \inf \{x \in \mathbb{R} \, : \, F_X(x) \geq \beta\},$$

i.e., the value at risk of X at confidence level β, and

$$\text{AVaR}_\beta(X) := \frac{1}{1 - \beta} \int_\beta^1 \text{VaR}_\alpha(X) \, d\alpha.$$

This gives a measure of the tail of the distribution of X. It is particularly well suited in the context of risk-averse optimization as a means of accounting for tail events. CVaR can be written in several ways; for optimization, we use

$$\text{AVaR}_\beta(X) = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{1 - \beta} \mathbb{E}[(X - t)_+] \right\}, \tag{4.1}$$

where $(x)_+ := \max\{0, x\}$ [38]; the (smallest) minimizer in (4.1) is $\text{VaR}_\beta(X)$.

As shown in [25, Thm 1], the only coherent risk measures that are continuously Fréchet differentiable are expectations. Therefore, regardless of how smooth the objective or control state mappings are, any risk-averse PDE-constrained opti-

mization problem using coherent regular risk measures is an infinite-dimensional nonsmooth optimization problem.

4.3 Risk-Averse PDE-Constrained Optimization: Theory

We now focus on developing the theory for the "single-player" setting. We start by considering the following abstract optimization problem:

$$\min_{u \in U_{\mathrm{ad}}} \mathcal{R}[\mathcal{J}(S(u))] + \wp(u). \tag{4.2}$$

Here, $u \in U$ represents the decision variable (controls, parameters, designs, etc.), U_{ad} is the associated feasible set, $\wp$ is a deterministic cost function, $\mathcal{R}$ is a risk measure as in Sect. 4.2, $\mathcal{J}$ is a random objective in the form of a general superposition operator, and $S(u)$ is the solution mapping for the random PDE.

As motivation for the chosen setting, we recall the class of random PDEs considered in [28] (in strong form): For $u \in U$ and $\mathbb{P}-$a.e. $\omega \in \Omega$, $y = S(u)$ solves

$$-\nabla \cdot (\kappa(\omega)\nabla y(\omega)) + c(\omega)y(\omega) + N(y(\omega), \omega) = [B(\omega)u] + b(\omega), \quad \text{in } D$$

$$\kappa(\omega)\frac{\partial y}{\partial n}(\omega) = 0, \quad \text{on } \partial D. \tag{4.3}$$

Here, we assume κ, c, b are random elements in an appropriate Bochner space and the operator N is a potentially nonlinear maximal monotone operator. $B(\omega)$ maps u into the image space of the differential operator.

Returning to the abstract setting, it was shown in [26] that a number of basic regularity assumptions need to be imposed on $\mathcal{R}, \mathcal{J}, S, \wp$, and U_{ad} in order to prove the existence of a solution and derive optimality conditions for (4.2). The inclusion of stochasticity and the nonlinearity and nonsmoothness of $\mathcal{R}$ add a further level of complexity not seen in deterministic problems. We impose the following conditions on S and $\mathcal{J}$ throughout.

Assumption 4.1 (Properties of the Solution Map) It holds that

1. $S(u) : \Omega \to V$ is strongly $\mathcal{F}$-measurable for all $u \in U_{\mathrm{ad}}$.
2. There exists an increasing function $\rho : [0, \infty) \to [0, \infty)$ and $C \in L^q(\Omega, \mathcal{F}, \mathbb{P})$ with $C \geq 0, q \in [1, \infty]$ such that

$$\|S(u)\|_V \leq C\rho(\|u\|_U) \quad \mathbb{P}-\text{a.e.} \quad \forall u \in U_{\mathrm{ad}}.$$

3. If $u_n \rightharpoonup u$ in U_{ad}, then $S(u_n) \rightharpoonup S(u)$ in V $\mathbb{P}-$a.e.

Each of these assumptions is minimal. For example, if $S(u)$ is not measurable, then $\mathcal{R} \circ \mathcal{J} \circ S$ is meaningless. The second assumption can be seen as an integrability requirement. Since $\mathcal{J}$ is typically a nonlinear operator, it is essential for S to possess such properties. The latter condition appears to be the weakest condition needed (along with the assumption on $\mathcal{R}$, $\mathcal{J}$, etc. below) to prove the existence of a solution. As shown in [24, Sec. 2.2], Assumption 4.1 implies:

1. $S(u) \in L^q(\Omega, \mathcal{F}, \mathbb{P}; V)$ for all $u \in U_{\text{ad}}$.
2. By letting

$$\mathcal{V} := L^q(\Omega, \mathcal{F}, \mathbb{P}; V),$$

   we have $S(u_n) \rightharpoonup S(u)$ in $\mathcal{V}$ for any $\{u_n\} \subset U_{\text{ad}}$ such that $u_n \rightharpoonup u$.

Furthermore, in order to derive optimality conditions, S needs to be continuously differentiable.

Assumption 4.2 There exists an open set $W \subseteq U$ with $U_{\text{ad}} \subseteq W$ such that the solution map $u \mapsto S(u) : W \to \mathcal{V}$ is continuously Fréchet differentiable.

The results in [24] indicate that we could slightly weaken this to Hadamard directional differentiability, which would allow us to consider risk-averse control of random elliptic variational inequalities in the future.

Continuing, we will assume that the random objective $\mathcal{J}$ is the result of a superposition of some possibly random integral functional J and an element $y \in \mathcal{V}$. The necessary, and in part sufficient, conditions needed for J are given below.

Assumption 4.3 (Properties of $J : V \times \Omega \to \mathbb{R}$) It holds that

1. J is a Carathéodory function, i.e., $J(\cdot, \omega)$ is continuous for $\mathbb{P}-$a.e. $\omega \in \Omega$ and $J(u, \cdot)$ is measurable for all $v \in V$.
2. If $1 \leq p, q < \infty$, then there exists $a \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ with $a \geq 0$ $\mathbb{P}-$a.e. and $c > 0$ such that

$$|J(v, \omega)| \leq a(\omega) + c\|v\|_U^{q/p}. \tag{4.4}$$

   If $1 \leq p < \infty$ and $q = \infty$, then the uniform boundedness condition holds: for all $c > 0$, there exists $\gamma = \gamma(c) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$|J(v, \omega)| \leq \gamma(\omega) \quad \mathbb{P}-\text{a.e.} \quad \forall v \in V, \ \|v\|_V \leq c. \tag{4.5}$$

3. $J(\cdot, \omega)$ is convex for $\mathbb{P}-$a.e. $\omega \in \Omega$.

It follows from a well-known result due to Krasnosel'skii, see, e.g., [29], [43, Thm 19.1], see also Theorem 4 in [15], that Assumption 4.3.1–2 guarantees $\mathcal{J} : \mathcal{V} \to L^p(\Omega, \mathcal{F}, \mathbb{P})$ continuously. These are necessary and sufficient and cannot be weakened. For several examples of objectives that satisfy Assumption 4.3, we refer to [26, Sec. 3.1]. Finally, the convexity assumption guarantees Gâteaux directional

differentiability. If this is not available, then additional assumptions must be made on the partial derivatives of J with respect to u. We gather the related main statements on $\mathcal{J}$ from [26] here for the reader's convenience.

Theorem 4.4 (Continuity and Gâteaux Differentiability of $\mathcal{J}$) *Let Assumption* 4.3*.1–2 hold. Then* $\mathcal{J} : \mathcal{V} \to L^p(\Omega, \mathcal{F}, \mathbb{P})$ *is continuous. Furthermore, if Assumption* 4.3*.1–3 holds, then* $\mathcal{J}$ *is Gâteaux directionally differentiable.*

Since the objective functional in (4.2) is of the form $\mathcal{R} \circ \mathcal{J} \circ S$, Theorem 4.4 is not strong enough to guarantee the necessary smoothness properties of $\mathcal{J}$ as a nonlinear operator from $\mathcal{V}$ into $L^p(\Omega, \mathcal{F}, \mathbb{P})$ that would provide us with first-order optimality conditions. This requires further regularity conditions. The weakest type of directional differentiability that allows a chain rule is Hadamard directional differentiability, cf. [40]. In the current setting, this can be demonstrated if $\mathcal{J}$ is locally Lipschitz, see [26, Cor. 3.10]. For the development of function-space-based optimization algorithms, in particular the convergence analysis, we generally need continuous Fréchet differentiability. This can be proven provided the partial derivatives of $\partial_u J(\cdot, \omega)$ satisfy a Hölder continuity condition, see [26, Thm. 3.11].

We now have a sufficient amount of structure to prove existence of optimal solutions to (4.2). The following lemma is essential.

Lemma 4.5 (Weak Lower-Semicontinuity of the Composite Objective) *Let Assumptions* 4.1 *and* 4.3 *hold. If* $\mathcal{R} : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ *is proper, closed, monotonic, convex, and subdifferentiable at* $\mathcal{J}(S(u))$ *for some* $u \in U_{\mathrm{ad}}$, *then the composite functional* $(\mathcal{R} \circ \mathcal{J} \circ S) : U_{\mathrm{ad}} \to \mathbb{R}$ *is weakly lower semicontinuous at* $u \in U_{\mathrm{ad}}$.

Using Lemma 4.5, we can now prove existence of solutions.

Theorem 4.6 (Existence of Optimal Solutions) *Let Assumptions* 4.1*,* 4.2*, and* 4.3 *hold. Let* $\mathcal{R} : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ *be a proper, closed, convex, and monotonic risk measure, and let* $\wp : U \to \overline{\mathbb{R}}$ *be proper, closed, and convex. Finally, suppose either* U_{ad} *is bounded or* $u \mapsto \mathcal{R}(\mathcal{J}(S(u))) + \wp(u)$ *is coercive. Then,* (4.2) *has a solution.*

Next, we can also derive a general first-order optimality condition. The essential point here is the regularity condition on $\mathcal{R}$, which guarantees the composite reduced objective function $\mathcal{R} \circ \mathcal{J} \circ S$ is Hadamard directionally differentiable. The standard regularity assumptions: finiteness or $\mathrm{int}\,\mathrm{dom}\,\mathcal{R} \neq \emptyset$ are considerably mild given the types of risk measures used in practice.

Theorem 4.7 (A General Optimality Condition) *Suppose that in addition to the assumptions of Theorem* 4.6*, the risk measure* $\mathcal{R}$ *is either finite on* $L^1(\Omega, \mathcal{F}, \mathbb{P})$ *or* $\mathrm{int}\,\mathrm{dom}\,\mathcal{R} \neq \emptyset$. *Moreover, assume that* $\mathcal{J} : \mathcal{V} \to L^p(\Omega, \mathcal{F}, \mathbb{P})$ *is locally Lipschitz and* $\wp$ *is Gâteaux directionally differentiable. Then for any optimal solution* $u^\star$

to (4.2), *the following first-order optimality condition holds:*

$$\sup_{\vartheta \in \partial \mathcal{R}(\mathcal{J}(S(u^\star)))} \mathbb{E}[\mathcal{J}'(S(u^\star); S(u^\star)'\delta u)\,\vartheta] + \wp'(u^\star; \delta u) \geq 0, \ \ \forall \delta u \in T_{U_{\mathrm{ad}}}(u^\star),$$

(4.6)

where $T_{U_{\mathrm{ad}}}(u^\star)$ is the contingent cone to U_{ad} at $u^\star$, which is defined by

$$T_{U_{\mathrm{ad}}}(u^\star) := \left\{ d \in U \mid \exists \, \tau_k \downarrow 0, \ \exists \, d_k \to d \text{ in } U \ : z^\star + \tau_k d_k \in U_{\mathrm{ad}} \, \forall k \right\}.$$

For illustration of (4.6), let $p = 2$, $U = L^2(D)$, $S(u^\star) = \mathbf{A}^{-1}(\mathbf{B}u^\star + \mathbf{b})$ and

$$J(y, \omega) = J(y) := \frac{1}{2}\|y - y_d\|_{L^2(D)}^2 \text{ and } \wp = \frac{\nu}{2}\|u\|_{L^2(D)}^2,$$

where $\mathbf{A}^{-1}$ is a linear isomorphism from $\mathcal{V}^*$ into $\mathcal{V}$, $\mathbf{B} \in \mathcal{L}(U, \mathcal{V}^*)$, and $\mathbf{b} \in \mathcal{V}^*$; then (4.6) unfolds into a somewhat more familiar form: If $u^\star$ is an optimal solution of (4.2), then there exists an adjoint state $p^\star \in \mathcal{V}^*$ and a subgradient $\vartheta^\star \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$\left(u^\star - \frac{1}{\nu}\mathbb{E}[\mathbf{B}^* p^\star \vartheta^\star], u - u^\star \right)_U \geq 0, \ \ \forall u \in U_{\mathrm{ad}},$$

$$\mathcal{R}[X] - \mathcal{R}[\mathcal{J}(y^\star)] - \mathbb{E}[\vartheta^\star(X - \mathcal{J}(y^\star))] \geq 0, \ \ \forall X \in L^1(\Omega, \mathcal{F}, \mathbb{P}),$$

(4.7)

$$\mathbf{A}y^\star - \mathbf{B}u^\star + \mathbf{b} = 0,$$

$$\mathbf{A}^* p^\star - y_d + y^\star = 0.$$

This provides us with the interesting fact that the optimal control is the projection onto U_{ad} of the expectation of adjoint term $\mathbf{B}^* p^\star$, where the expectation has been adjusted according to the risk preference expressed in $\mathcal{R}$ via the subgradient $\vartheta^\star$. The latter is often referred to as the "risk indicator" in the literature for obvious reasons. In the case of AVaR$_\beta$, the numerical experiments in [24] indicate that $\mathbb{P}(\mathrm{supp}\,\vartheta^\star) = 1 - \beta$. Therefore, the majority of support is used to treat tail events. Note also that when designing first-order methods for such problems, this fact allows a significant reduction in the number of PDEs solved per iteration required to calculate the reduced gradient.

For a more challenging example, we recall the setting from [28] in (4.3) in more detail. Among the most difficult aspects of the assumptions used to prove existence of a solution and derive optimality conditions are the conditions placed on the solution mapping S. In [28], we postulate several verifiable assumptions. To this aim, we suppose that $S(u)$ is the solution of a general parametric operator equation: For each $u \in U$, find $y(\omega) = [S(u)](\omega) \in U$ such that

$$e(y, u; \omega) := \mathbf{A}(\omega)y + \mathbf{N}(y, \omega) - \mathbf{B}(\omega)u - \mathbf{b}(\omega) \ni 0 \quad \text{for a.a. } \omega \in \Omega. \quad (4.8)$$

We impose the following assumptions on the operators.

Assumption 4.8 (Pointwise Characterization of the Problem Data in (4.8)*)*

1. Let $\mathbf{A} : \Omega \to \mathcal{L}(V, V^*)$ satisfy $\mathbf{A}(\omega)$ is monotone for a.a. $\omega \in \Omega$ and there exists $\gamma > 0$ and a random variable $C : \Omega \to [0, \infty)$ with $C > 0$ a.e. such that

$$\langle \mathbf{A}(\cdot)y, y \rangle_{U^*, U} \geq C \|y\|_V^{1+\gamma} \quad \text{a.e.} \ \forall\, y \in V. \tag{4.9}$$

2. Let $\mathbf{b} : \Omega \to V^*$.
3. Let $\mathbf{N} : V \times \Omega \rightrightarrows V^*$ satisfy $\mathbf{N}(\cdot, \omega)$ is maximal monotone with $\mathbf{N}(0, \omega) = \{0\}$ for a.a. $\omega \in \Omega$.
4. Let $\mathbf{B} : \Omega \to \mathcal{L}(U, V^*)$ be completely continuous for a.a. $\omega \in \Omega$.

Since these conditions are taken to be pointwise in ω, they can be viewed as the minimal data assumptions that are imposed when considering optimization of elliptic semilinear equations. The following assumption is essential for measurability issues. It is unclear if it can be weakened. Ultimately, the coefficients and mappings used to define $\mathbf{A}$, $\mathbf{N}$, etc. will dictate the integrability of $S(u)$.

Assumption 4.9 (Measurability and Integrability of the Operators in (4.3)*)* Let Assumption 4.8 hold and suppose there exists s, $t \in [1, \infty]$ with

$$1 + \frac{1}{\gamma} \leq s < \infty \quad \text{and} \quad t \geq \frac{s}{\gamma(s-1)-1}$$

such that $\mathbf{A}(\cdot)y \in L^s(\Omega, \mathcal{F}, \mathbb{P}; V^*)$ for all $y \in V$, $\mathbf{N}(\cdot, \omega)$ is single-valued and continuous for a.a. $\omega \in \Omega$ and $\mathbf{N}(y, \cdot) \in L^s(\Omega, \mathcal{F}, \mathbb{P}; V^*)$ for all $y \in V$, $\mathbf{B} \in L^s(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{L}(U, V^*))$, $\mathbf{b} \in L^s(\Omega, \mathcal{F}, \mathbb{P}; V^*)$ and $C^{-1} \in L^t(\Omega, \mathcal{F}, \mathbb{P})$.

Finally, we require assumptions on $\mathbf{N}$ to derive optimality conditions.

Assumption 4.10 (Differentiability of $\mathbf{N}(\cdot, \omega)$*)* In addition to Assumption 4.9, we assume that $\mathbf{N}(\cdot, \omega)$ is single-valued and continuously Fréchet differentiable from V into V^* for a.a. $\omega \in \Omega$ with partial derivative $\mathbf{N}'(y, \omega)$, which defines a bounded, nonnegative linear operator from V into V^* a.e. for all $y \in V$. Moreover, we assume that $\mathbf{A}$ and $y \mapsto \mathbf{N}(y, \cdot)$ are continuous maps from $\mathcal{V}$ into $L^s(\Omega, \mathcal{F}, \mathbb{P}; V^*)$ and $y \mapsto \mathbf{N}'(y, \cdot)$ is a continuous map from $\mathcal{V}$ into $L^{qs/(q-s)}(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{L}(V, V^*))$.

We gather the main results in [28, Sec. 2.3] here for the reader's convenience.

Theorem 4.11 (Properties of the Solution Mapping $S(u)$**)** *Under the standing assumptions, the following statements hold.*

1. *If Assumption* 4.8 *holds, then* $\mathbf{A}(\omega) + \mathbf{N}(\cdot, \omega)$ *is surjective from V into V^* for a.a. $\omega \in \Omega$. In particular, there exists a unique solution $S(u)$ to* (4.8) *such that* $[S(u)](\omega) \in V$ *for a.a. $\omega \in \Omega$.*

2. *If in addition Assumption* 4.9 *holds and we let*

$$q := \frac{s\gamma}{1 + s/t}, \tag{4.10}$$

then $S(u) \in \mathcal{V} := L^q(\Omega, \mathcal{F}, \mathbb{P}; V)$ *for all* $u \in V$. *Furthermore, if* $u_k \rightharpoonup u$ *in* U, *then* $S(u_k) \to S(u)$ *in* V *a.e. and* $S(u_k) \to S(u)$ *in* $\mathcal{V}$, *i.e.,* S *is completely continuous.*

3. *If in addition Assumption* 4.10 *holds, then* $u \mapsto S(u)$ *is continuously Fréchet differentiable from* U *into* $\mathcal{V}$.

We now return to a concrete example and cast (4.3) in the form (4.8).

Example Define the linear elliptic operator $\mathbf{A}(\omega)$ by

$$\langle \mathbf{A}(\omega)y, v \rangle_{V^*, V} = \int_D \{\kappa(\omega, x)\nabla y(x) \cdot \nabla v(x) + c(\omega, x)y(x)v(x)\} \, \mathrm{d}x,$$

for $y, v \in \mathcal{V}$. Analogously, we let $\mathbf{N}(\cdot, \omega)$ be the nonlinear operator given by

$$\langle \mathbf{N}(y, \omega), v \rangle_{V^*, V} = \int_D N(y(x), \omega, x)v(x) \, \mathrm{d}x,$$

where $N : \mathbb{R} \times \Omega \times D \to \mathbb{R}$. The right-hand side can be defined by

$$\langle \mathbf{B}(\omega)u, v \rangle_{V^*, V} = \int_D [B(\omega)u](x)v(x) \, \mathrm{d}x \quad \text{and} \quad \langle \mathbf{b}(\omega), v \rangle_{V^*, V} = \int_D b(\omega, x)v(x) \, \mathrm{d}x,$$

where $B : \Omega \to \mathcal{L}(U, L^2(D))$ and $b \in \mathcal{V}^*$.

Assuming that $\kappa(\omega, \cdot), c(\omega, \cdot) \in L^\infty(D)$ for a.a. $\omega \in \Omega$ and for a.a. $\omega \in \Omega$, $x \in D$, satisfy: there exist $\kappa_0 > 0$ and $c_0 > 0$ such that

$$\kappa_0 \leq \kappa(\omega, x) \text{ and } c_0 \leq c(\omega, x),$$

then the conditions in Assumptions 4.8 and 4.9 on $\mathbf{A}$ are satisfied with $\gamma = 1$, $C = \min\{\kappa_0, c_0\}$, $s = 2$, $t = \infty$. For $\mathbf{N}$, we at least need $N(\cdot, \omega, x) : \mathbb{R} \to \mathbb{R}$ to be continuous and monotonically increasing with $N(0, \omega, x) = 0$ for a.a. $\omega \in \Omega$ and a.a. $x \in D$. This would yield the monotonicity requirement in Assumption 4.8, which would be the case for a nonlinearity of the type: $N(u, \omega, x) = c(\omega, x)(\sinh(u) - u)$. Otherwise, we can obtain continuity via the usual growth conditions of Krasnosel'skii as in, e.g., Theorems 1 and 4 in [15] or the comprehensive monograph [2]. Similarly, if we have $b(\omega, \cdot) \in L^r(D)$ with $r > d/2$ for a.a. $\omega \in \Omega$, then Assumption 4.8.2 holds and if B is, e.g., the canonical embedding operator from $L^2(D)$ into $H^1(D)^*$, then Assumption 4.8.3 also holds. For Assumption 4.9, we could require $b \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}; L^2(D))$ and $\kappa, c \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}; L^\infty(D))$. This assumption would not hold for $\mathbf{N}$ when generated by the hyperbolic sine unless V was replaced by a more regular space,

e.g., $H^2(D)$. However, if $d = 2$ and ∂D is sufficiently regular, then by the Sobolev embedding theorems we could still use $V = H^1(D)$ when N is generated by monotone polynomials of arbitrary degree.

Behind all of these technical details lie the hypotheses imposed by measurable selection theorems, e.g., Filippov's theorem, which generally require the random elements to map into separable spaces. The integrability conditions are then derived using the monotonicity of the operators. Therefore, one should be rather careful when generating new examples from deterministic PDE models as they may not always be well defined in the stochastic setting.

Finally, we conclude this section by noting that many example problems used in the literature consider linear elliptic PDE under uncertainty. This drastically simplifies the measurability, integrability, continuity, and differentiability issues for the solution mapping S. Building on the properties of the solution operator and requirements on the objective functionals J discussed above, one can derive similar measurability, integrability, and (weak) continuity results for the adjoint equations and ultimately an optimality system as in the linear case shown above.

4.4 A Risk-Averse PDE-Constrained Nash Equilibrium Problem

We may now formulate a model risk-averse PDE-constrained Nash equilibrium problem. Using the results of the previous section, we prove existence of a Nash equilibrium and derive optimality conditions. In what follows, we consider the following setting: For each $i = 1, \ldots, N$ $(N > 1)$, we assume:

1. $U^i := L^2(D)$, $U_{\text{ad}}^i := \left\{ v \in U^i \mid a_i \le v \le b_i \text{ a.e. } D \right\}$, $a_i, b_i \in L^2(D) : a_i < b_i$.
2. $J_i(y, \omega) := \frac{1}{2} \| y - y_d^i \|_{L^2(D)}^2$, $y_d^i \in L^2(D)$; $\wp(u) := \frac{v_i}{2} \| u \|_{L^2(D)}^2$ $v_i > 0$.
3. $S : U^1 \times \cdots \times U^N \to \mathcal{V}$ is the solution mapping for the random PDE given by (4.8) under Assumptions 4.8 and 4.9 such that $\mathbf{A}$ is defined as in Example 4.3, i.e., uniformly elliptic with $\gamma = 1$, $C = \min\{\kappa_0, c_0\}$, $s = 2$, $t = \infty$; $\mathbf{N} \equiv 0$; $\mathbf{b} \in \mathcal{V}^*$; and $\mathbf{B} : U^1 \times \cdots \times U^N \to \mathcal{V}^*$ satisfies

$$\mathbf{B}u = \mathbf{B}_1 u_1 + \ldots \mathbf{B}_N u_N,$$

where $\mathbf{B}_i$, $i = 1, \ldots, N$, is defined as in Assumptions 4.8 and 4.9. In particular,

$$S(u) := \mathbf{A}^{-1} \left(\sum_i \mathbf{B}_i u_i + \mathbf{b} \right).$$

4. $\mathcal{R}_i : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is a regular coherent measure of risk, e.g., AVaR$_\beta$.

Under these assumptions, we consider the associated risk-averse PDE-constrained Nash equilibrium problem (NEP) in which the ith player's problem takes the form

$$\min_{u_i \in U^i_{ad}} \mathcal{R}_i(\mathcal{J}_i(S(u_i, u_{-i}))) + \wp(u_i) \text{ over } u_i \in U^i. \qquad (4.11)$$

Using the Kakutani–Fan–Glicksberg fixed-point theorem (Theorem (2.1) above, [14]), we can demonstrate that this problem admits a Nash equilibrium.

Theorem 4.12 (Existence of a Risk-Averse Nash Equilibrium) *The Nash equilibrium problem whose individual players each solve a variant of* (4.11) *admits a solution in the form of a pure strategy Nash equilibrium.*

Proof We need to verify the conditions of Theorem 2.1. Since each U^i is infinite-dimensional, we view each U^i_{ad} as metrizable compact locally convex topological vector spaces as in [20, 21]. This is possible since U^i_{ad} is a norm bounded, closed, and convex set in a separable Hilbert space. Next, we define the best response mappings:

$$\mathcal{B}_i(u_{-i}) := \arg\min_{u_i \in U^i_{ad}} \mathcal{R}_i(\mathcal{J}_i(S(u_i, u_{-i}))) + \wp(u_i) \text{ over } u_i \in U^i.$$

We need to show that each $\mathcal{B}_i$ has nonempty, bounded, and convex images in U^i.

The risk measure $\mathcal{R}_i$ is proper, closed, convex, and monotonic. Since $\mathcal{R}_i$ is defined on all of $L^1(\Omega, \mathcal{F}, \mathbb{P})$, it is finite everywhere and therefore continuous and consequently subdifferentiable; in particular at $\mathcal{J}_i(S(u_i, u_{-i}))$ for any feasible strategy vector (u_i, u_{-i}). The tracking-type functional considered here can easily be shown to satisfy all the necessary assumptions outlined above; see [24] or [26]. Concerning S, we note that for any fixed $u_{-i} \in U^{-i}_{ad}$, we have $\mathbf{B}(0, u_{-i}) = \sum_{j \neq i} \mathbf{B}_j u_j$. The latter term can be taken on the right-hand side of the PDE as a perturbation of $\mathbf{b}$. Clearly, this "new" constant term is in $\mathcal{V}^*$. It light of this, we can readily verify the necessary assumptions for continuity and differentiability with respect to u_i required in Theorem 4.11.

It follows that $\mathcal{R} \circ \mathcal{J} \circ S : U^i \to \mathbb{R}$ is weakly lower semicontinuous (cf. Lemma 4.5). The existence of solutions results from the fact that $\wp$ is coercive and $\mathcal{R}_i \circ \mathcal{J}_i \circ S$ nonnegative (cf. Theorem 4.6). Furthermore, since $\mathcal{R}_i$ is a monotone risk measure, it preserves the pointwise convexity of the integrand $\mathcal{J} \circ S$. Therefore, the set of all optimal solutions is convex and, by hypothesis on U^i_{ad}, bounded. Therefore, we conclude that $\mathcal{B}_i$ has nonempty, convex, bounded images in U^i_{ad}.

Next, define $\mathcal{B} : U^1_{ad} \times \cdots \times U^N_{ad} \rightrightarrows U^1_{ad} \times \cdots \times U^N_{ad}$ by

$$\mathcal{B}(u) := \mathcal{B}_1(u_{-1}) \times \cdots \times \mathcal{B}_N(u_{-N}).$$

Suppose that $(u^k, v^k) \in \text{gph } B$ such that $(u^k, v^k) \rightharpoonup (\bar{u}, \bar{v})$. This means in particular that for all k we have $v_i^k \in \mathcal{B}_i(u_{-i}^k)$, i.e.,

$$(\mathcal{R}_i \circ \mathcal{J}_i \circ S)(v_i^k, u_{-i}^k) + \wp(v_i^k) \leq (\mathcal{R}_i \circ \mathcal{J}_i \circ S)(w, u_{-i}^k) + \wp(w) \quad \forall w \in U_{\text{ad}}^i.$$

In the current setting

$$S(u) = \mathbf{A}^{-1} \left(\sum_i \mathbf{B}_i u_i + \mathbf{b} \right) = \mathbf{A}^{-1} \mathbf{B}_i u_i + \mathbf{A}^{-1} \mathbf{b} + \sum_{j \neq i} \mathbf{A}^{-1} \mathbf{B}_j u_j.$$

As shown in Lemma 2.1 [28], each $\mathbf{B}_i$ is completely continuous from U^i into $L^2(\Omega, \mathcal{F}, \mathbb{P}; V^*) = \mathcal{V}^*$. Therefore, we have $\mathbf{B}_i v_i^k \to \mathbf{B}_i \bar{v}_i$ and $\mathbf{B}_j u_j^k \to \mathbf{B}_j \bar{u}_j$ strongly in $\mathcal{V}^*$ for each i and each $j \neq i$. It immediately follows from that $S(v_i^k, u_{-i}^k) \to S(\bar{v}_i, \bar{u}_{-i})$ and for any $w \in U_{\text{ad}}^i$ $S(w, u_{-i}^k) \to S(w, \bar{u}_{-i})$.

Next, since $\mathcal{R}_i$ and $\mathcal{J}_i$ are continuous on their respective spaces, we have

$$(\mathcal{R}_i \circ \mathcal{J}_i \circ S)(v_i^k, u_{-i}^k) \to (\mathcal{R}_i \circ \mathcal{J}_i \circ S)(\bar{v}_i, \bar{u}_{-i})$$

$$(\mathcal{R}_i \circ \mathcal{J}_i \circ S)(w, u_{-i}^k) \to (\mathcal{R}_i \circ \mathcal{J}_i \circ S)(w, \bar{u}_{-i}).$$

Then due to the weak lower semicontinuity of $\wp$ on U^i, it follows that

$$(\mathcal{R}_i \circ \mathcal{J}_i \circ S)(\bar{v}_i, \bar{u}_{-i}) + \wp(\bar{v}_i) \leq (\mathcal{R}_i \circ \mathcal{J}_i \circ S)(w, \bar{u}_{-i}) + \wp(w) \quad \forall w \in U_{\text{ad}}^i,$$

i.e., $\bar{v}_i \in \mathcal{B}_i(\bar{u}_{-i})$. Hence, the noncooperative game admits a Nash equilibrium. □

Remark 4.13 The previous proof can easily be extended to more complicated PDE models and objective functions. However, for nonlinear operators $\mathbf{N}$, we need to extend the results in Sect. 2 to the stochastic setting.

Given the explicit structure of the current setting, we can also derive optimality conditions for the NEP. Moreover, we can show that this specific problem reduces to a special kind of equilibrium problem in which the risk indicators are determined simultaneously by a single "risk trader."

Theorem 4.14 (Optimality Conditions) *Let $\bar{u}$ be a Nash equilibrium for* (4.11). *Then for each $i = 1, \ldots, N$ there exists a pair $(p_i^\star, \vartheta_i^\star) \in \mathcal{V} \times L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ such that the following conditions hold: $\vartheta^\star \in \partial \mathcal{R}_i[\mathcal{J}_i(y^\star)]$ and*

$$\left(u_i^\star - \frac{1}{v_i} \mathbb{E}[\mathbf{B}_i^* p_i^\star \vartheta_i^\star], w - u_i^\star \right)_{U^i} \geq 0, \ \forall w \in U_{\text{ad}}^i,$$

$$\mathbf{A} y^\star - \mathbf{B} u^\star + \mathbf{b} = 0, \tag{4.12}$$

$$\mathbf{A}^* p_i^\star - y_d^i + y^\star = 0.$$

Proof This follows from Sect. 4.3 and the definition of a Nash equilibrium. □

System (4.12) leads to a useful reformulation. For each i, the adjoint states $p_i^\star$ split into the sum of a joint adjoint state $q^\star := \mathbf{A}^{-*}y^\star$ and a fixed i-dependent term $\tilde{y}_d^i := -\mathbf{A}^{-*}y_d^i$, where $\tilde{y}_d^i$ is now stochastic. Then, for each i, we have

$$\frac{1}{v_i}\mathbb{E}[\mathbf{B}_i^* p_i^\star \vartheta_i^\star] = \frac{1}{v_i}\mathbb{E}[\mathbf{B}_i^*(q^\star + \tilde{y}_d^i)\vartheta_i^\star] = \frac{1}{v_i}\mathbb{E}[\mathbf{B}_i^* q^\star \vartheta_i^\star] + \underbrace{\frac{1}{v_i}\mathbb{E}[\mathbf{B}_i^* \tilde{y}_d^i \vartheta_i^\star]}_{=:\hat{c}_i}.$$

By defining $\mathbf{G}_i u := \frac{1}{v_i}\mathbf{B}_i^*\mathbf{A}^{-*}\mathbf{A}^{-1}\mathbf{B}u$ and $\mathbf{g}_i := \frac{1}{v_i}\mathbf{B}_i^*\mathbf{A}^{-*}\mathbf{A}^{-1}\mathbf{b}$, the variational inequality in (4.12) can be written as

$$(u_i^\star - (\mathbb{E}[\vartheta_i^\star \mathbf{G}_i u^\star] + c_i(\vartheta_i^\star)), v - u_i^\star)_{U^i} \geq 0 \quad \forall v \in U_{\text{ad}}^i,$$

where $c_i(\vartheta_i) := \mathbb{E}[\vartheta_i \mathbf{g}_i] - \hat{c}_i$. Summing over i, we obtain

$$\sum_{i=1}^{N}(u_i^\star - (\mathbb{E}[\vartheta_i^\star \mathbf{G}_i u^\star] + c_i(\vartheta_i^\star)), v_i - u_i^\star)_{U^i} \geq 0 \quad \forall v \in U_{\text{ad}}. \tag{4.13}$$

Conversely, if the previous inequality holds, then by using the variations

$$(v_1^\star, \ldots, v_i, \ldots, v_N^\star) = v \in U_{\text{ad}} = U_{\text{ad}}^1 \times \cdots \times U_{\text{ad}}^N$$

for each $i = 1, \ldots, N$ (leaving only v_i to vary), we recover the individual inequalities. We will refer to (4.13) as the "aggregate player's problem." Letting $\text{Proj}_{U_{\text{ad}}^i}$ denote the metric projection onto U_{ad}^i, this can be formulated as a single nonsmooth equation in the product space $U = U^1 \times \cdots \times U^N$: Find $u^\star \in U$: $\forall i = 1, \ldots, N$

$$u_i^\star = \text{Proj}_{U_{\text{ad}}^i}\left[\mathbb{E}[\vartheta_i^\star \mathbf{G}_i u^\star] + c_I(\vartheta_i^\star)\right]. \tag{4.14}$$

Continuing, since $\mathcal{R}_i$ is assumed to be a coherent risk measure, we have

$$\vartheta_i^\star \in \underset{\vartheta \in \mathfrak{A}_i}{\text{argmax}}\ \mathbb{E}[\vartheta \mathcal{J}_i(y^\star)],$$

where $\mathfrak{A}_i := \text{dom}(\mathcal{R}_i^*)$ is the domain of the Fenchel conjugate $\mathcal{R}_i^*$ of $\mathcal{R}_i$. It is then easy to show that all of the subdifferential inequalities can be joined into a single maximization problem:

$$\max\left\{\sum_{i=1}^{N}\mathbb{E}[\vartheta_i \mathcal{J}_i(\mathbf{A}^{-1}(\mathbf{B}u^\star + \mathbf{b}))],\ \text{over } \vartheta \in \mathfrak{A}\right\}, \tag{4.15}$$

where $\mathfrak{A} := \mathfrak{A}_1 \times \cdots \times \mathfrak{A}_N$. Problem (4.15) always has a solution since the objective is a bounded linear functional and $\mathfrak{A}$ is a weakly-$*$ sequentially compact, closed, and convex set. Inspired by the terminology in [37], we will refer to (4.15) as the "risk trader's problem."

We have thus proven that the risk-averse PDE-constrained NEP can be understood as a type of MOPEC (multiple optimization problems with equilibrium constraints) comprising a single aggregate player, who solves a well-posed variational inequality in u given a fixed risk indicator vector ϑ, and a risk trader who spreads the risk of the decision vector u over the components of ϑ in light of the various objectives $\mathcal{J}_i$ and risk preferences $\mathfrak{A}_i$.

Even in this special case, it is difficult to immediately select an appropriate solution algorithm. Perhaps the main challenge lies in the fact that the risk trader's problem does not have a unique solution. One remedy for this to ensure a unique ϑ for a given u is to replace the objective in (4.15) by

$$\mathbb{E}[\vartheta_i \mathcal{J}_i(\mathbf{A}^{-1}(\mathbf{B}u^\star + \mathbf{b}))] - \frac{\varepsilon}{2}\mathbb{E}[\vartheta_i^2] \quad \varepsilon > 0. \tag{4.16}$$

This was suggested in [24] for treating the nonsmooth risk measure AVaR_β in the context of PDE-constrained optimization under uncertainty. It was later demonstrated that such a regularization is a special case of the deeper theory of epi-regularization of risk measures in [25]. We briefly discuss this notion below.

4.5 Risk-Averse PDE-Constrained Decision Problems: Smooth Approximation

As a means of circumventing the unacceptably slow performance of classical nonsmooth optimization algorithms such as subgradient methods or bundle methods, we proposed smoothing approaches in [24] and [25]. An alternative viewpoint can be found by exploiting the structure of a specific class of coherent risk measures and using an interior-point approach as in [13]. In addition, the analysis in the previous section indicates yet another reason to consider some form of variational smoothing in the context of stochastic PDE-constrained equilibrium problems.

We briefly give the details of epi-regularization as it has proven to be a versatile tool not only for smoothing risk measures but also for analyzing new optimization methods for risk-averse PDE-constrained optimization, cf. [27]. Let $\Psi : \mathcal{X} \to \overline{\mathbb{R}}$ be a proper, closed, and convex functional and $\mathcal{R}$ a regular measure of risk. Then for $\varepsilon > 0$, we define the epi-regularized measure of risk as

$$\mathcal{R}_\varepsilon^\Psi[X] = \inf_{Y \in \mathcal{X}} \left\{ \mathcal{R}[X - Y] + \varepsilon \Psi\left[\varepsilon^{-1}Y\right] \right\} = \inf_{Y \in \mathcal{X}} \left\{ \mathcal{R}[Y] + \varepsilon \Psi\left[\varepsilon^{-1}(X - Y)\right] \right\}.$$

As mentioned above, the regularization in (4.16) is equivalent to using the function $\Psi[X] = \frac{1}{2}\mathbb{E}[X^2]$. Another import example can be seen by setting $\mathcal{X} = L^2(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{R} = \text{AVaR}_\beta$, and $\Psi[X] := \mathbb{E}[X] + \frac{1}{2}\mathbb{E}[X^2]$. This results in

$$\mathcal{R}_\varepsilon^\Psi[X] = \inf_{t \in \mathbb{R}} \left\{ t + \mathbb{E}\left[v_{\beta,\epsilon}(X - t) \right] \right\},$$

which is continuously Fréchet differentiable and in which the scalar function $v_{\beta,\varepsilon}$ is given by

$$v_{\beta,\varepsilon}(x) = \begin{cases} -\frac{\varepsilon}{2}, & \text{if } x \leq -\varepsilon \\ \frac{1}{2\varepsilon}x^2 + x, & \text{if } x \in \left(-\varepsilon, \frac{\varepsilon\beta}{1-\beta} \right) \\ \frac{1}{1-\beta}\left(x - \frac{\varepsilon\beta^2}{2(1-\beta)} \right), & \text{if } x \geq \frac{\varepsilon\beta}{1-\beta}. \end{cases}$$

Epi-regularization has a number of advantageous properties. For example, we can show that the sequence of functionals $\left\{ \mathcal{R}_\varepsilon^\Psi \right\}_{\varepsilon > 0}$ converges in the sense of Mosco to $\mathcal{R}$. Furthermore, under certain assumptions on $\mathcal{J}$ and $\wp$, we can show that weak accumulation points of approximate minimizers $z_\varepsilon^\star$ are optimal for (4.2) and weak accumulation points of approximate stationary points are stationary for (4.2). For more on this topic, we refer to the forthcoming publication [25].

4.6 Risk-Averse PDE-Constrained Optimization: Solution Methods

In this final section, we outline the main components of the recently proposed primal–dual risk minimization algorithm in [27]. This is an all purpose optimization algorithm for minimizing risk measures in the context of PDE-constrained optimization under uncertainty.

In general, the individual problems in our risk-averse setting have the form:

$$\min_{x \in X_{\text{ad}}} \left\{ g(x) + \Phi(G(x)) \right\}, \tag{4.17}$$

where g is a deterministic objective function, G is an uncertain objective function, and Φ is a functional that maps random variables into the real numbers. The functional Φ is typically convex, positively homogeneous, and monotonic with respect to the natural partial order on the space of random variables.

Let $\Phi : \mathcal{Y} \to \mathbb{R}$, where $\mathcal{Y} = L^2(\Omega, \mathcal{F}, \mathbb{P})$. As shown in [41, Th. 6.5], there exists a nonempty, convex, closed, and bounded set $\mathfrak{A} \subseteq \{\theta \in \mathcal{Y}^* \mid \theta \geq 0 \text{ a.s.}\}$ such that a convenient bi-dual representation of Φ is available:

$$\Phi(X) = \sup_{\theta \in \mathfrak{A}} \mathbb{E}[\theta X]. \tag{4.18}$$

Moreover, Φ is continuous and subdifferentiable, cf. [41, Prop. 6.6], and $\mathfrak{A} = \partial\Phi(0)$.

Using these facts, (4.17) exhibits a familiar structure in which, by introducing the Lagrangian-type function $\ell(x, \lambda) := g(x) + \mathbb{E}[\lambda G(x)]$, we can consider the minimax reformulation:

$$\min_{x \in X_{\mathrm{ad}}} \sup_{\lambda \in \mathfrak{A}} \ell(x, \lambda). \tag{4.19}$$

We can then develop a method similar to the classical method of multipliers [16, 36].

To this end, we introduce the (dual) generalized augmented Lagrangian:

$$L(x, \lambda, r) := \max_{\theta \in \mathfrak{A}} \left\{ \ell(x, \theta) - \frac{1}{2r}\mathbb{E}[(\lambda - \theta)^2] \right\}. \tag{4.20}$$

Now, using several techniques from convex analysis, it can be shown that

$$L(x, \lambda, r) = g(x) + \min_{Y \in \mathcal{Y}} \left\{ \Phi(G(x) - Y) + \mathbb{E}[\lambda Y] + \frac{r}{2}\mathbb{E}[Y^2] \right\}. \tag{4.21}$$

In other words, L is the objective in (4.17) with Φ replaced by a multiplier-dependent epi-regularization, where the regularizer is

$$\Psi_{r,\lambda}(Y) = \mathbb{E}[\lambda Y] + \tfrac{r}{2}\mathbb{E}[Y^2].$$

Furthermore, letting

$$\Lambda(x, \lambda, r) := \mathrm{Proj}_{\mathfrak{A}}(rG(x) + \lambda),$$

where $\mathrm{Proj}_{\mathfrak{A}} : \mathcal{Y} \to \mathcal{Y}$ is the projection onto $\mathfrak{A}$, L attains the closed form

$$L(x, \lambda, r) = g(x) + \mathbb{E}[\lambda G(x)] + \frac{r}{2}\mathbb{E}[G(x)^2] - \frac{1}{2r}\mathbb{E}[\{(\mathrm{Id} - \mathrm{Proj}_{\mathfrak{A}})(rG(x) + \lambda)\}^2].$$

For many risk measures of interest, e.g., mean-plus-semideviation or convex combinations of mean and AVaR [27, Sec. 5.1], the optimization problem (4.17) can be rewritten so that $\Phi(Y) = \mathbb{E}[(Y)_+]$. Therefore, the projection operator $\mathrm{Proj}_{\mathfrak{A}}$ can be easily evaluated. For more general coherent risk measures, $\mathfrak{A}$ can be split into box constraints and a simple normalizing constraint that is treatable with a Lagrange multiplier, cf. [27, Sec. 5.2].

The basic algorithm is given in Algorithm 1. A detailed implementable version allowing for inexact subproblem solves, and multiplier-update strategies can be found in [27] (Algorithm 2). A full convergence theory for the primal and dual updates in both convex and nonconvex settings in infinite-dimensional spaces is given in [27, Sec. 4]. Here, the convergence of the primal variables exploits a number

of powerful results arising in the theory of epi-regularization. For the dual variables, a regularity condition that postulates the existence of a saddle point is needed.

Algorithm 1 Primal–dual risk minimization

---
1. **Initialize:** Given $x_0 \in X_{\text{ad}}$, $r_0 > 0$, and $\lambda_0 \in \mathfrak{A}$.
2. **While("Not Converged")**

   (a) Compute $x_{k+1} \in X_{\text{ad}}$ as approximate minimizer of $L(\cdot, \lambda_k, r_k)$.
   (b) Set $\lambda_{k+1} = \Lambda(x_{k+1}, \lambda_k, r_k)$.
   (c) Update r_{k+1}.

3. **End While**

Returning to our game-theoretic setting in Sect. 4.4, we see a clear link to the risk trader's problem (4.15). As mentioned in Sect. 4.4, (4.15) does not admit a unique solution. This makes the numerical solution of the game, in its original form as well as the proposed reduced from, very challenging. The suggestion in (4.16) indicates that we could handle this aspect by applying an epi-regularization technique to the risk measures. Though the suggestion given there is viable, the favorable convergence behavior of Algorithm 1 given in [27, Sec. 4] indicates that the multiplier-dependent epi-regularization update in the primal–dual algorithm is probably better suited (clearly algorithmically motivated). We thus propose a method that successively solves the aggregate player's game using an update formula for ϑ similar to the Λ-operator in the primal–dual algorithm. This avenue of thought will be the focus of future work. Nevertheless, the epi-regularization technique does not rule out the possibility that the associated system of nonlinear and semismooth equations admits distinct solutions. A possible remedy to this issue can be found in the recent publication [10].

5 Outlook

Generalized Nash equilibrium problems with PDE constraints represent a challenging class of infinite-dimensional equilibrium problems. Beyond the deterministic convex setting involving linear elliptic or parabolic PDEs, major theoretical and algorithmic challenges arise. Nevertheless, we have shown that it is still possible to treat some GNEPs involving semilinear, nonsmooth, and even mutlivalued forward problems by appealing to the notions of generalized convexity and isotonic mappings. Due to a lack of convexity, we have chosen to derive stationarity conditions using the versatile limiting variational calculus in the sense of Mordukhovich. In doing so, we have been able to push the boundaries of existence and optimality theory in the deterministic setting beyond linear state systems. Therefore, we may now build upon these advances toward the development of function-space-based numerical methods similar to [20, 21]. The recent results in [23] on augmented

Lagrangian-type methods (also developed within the priority program) may also prove to be useful here.

As outlined above, the stochastic risk-averse setting is now poised to transfer the results from the newly developed theory of risk-averse PDE-constrained optimization [13, 24–28] to the setting of noncooperative strategic games. This will be the focus for the remainder of the project duration. In addition to the algorithmic strategy mentioned above, there are several open theoretical questions relating to variational convergence in the context of strategic games and asymptotic statistical properties of Nash equilibrium in the vein of [41, Chap. 5]. Some progress on related stability issues using probability metrics has been made in the recent Master's thesis [22]. In addition, the results from the deterministic nonlinear case can be folded into the stochastic setting by using the results in [27] for risk-averse control of semilinear equations. Finally, in order to treat even jointly convex state-constrained risk-averse PDE-constrained GNEPs, a sufficient theory of PDE-constrained optimization under uncertainty with state constraints is under development.

References

1. ADAMS, R. A., AND FOURNIER, J. J.-F. *Sobolev Spaces*, second ed. Elsevier, Amsterdam, 2008.
2. APPELL, J., AND ZABREJKO, P. P. *Nonlinear Superposition Operators*. Cambridge Tracts in Mathematics. Cambridge University Press, 2008.
3. ARTZNER, P., DELBAEN, F., EBER, J.-M., AND HEATH, D. Coherent measures of risk. *Math. Finance 9*, 3 (1999), 203–228.
4. BLYTH, T. *Lattices and Ordered Algebraic Structures*. Universitext. Springer London, 2006.
5. BONNANS, J. F., AND SHAPIRO, A. *Perturbation Analysis of Optimization Problems*. Springer Verlag, Berlin, Heidelberg, New York, 2000.
6. BORWEIN, J., AND ZHU, Q. *Techniques of Variational Analysis*. CMS Books in Mathematics. Springer New York, 2006.
7. BREDIES, K., CLASON, C., KUNISCH, K., AND WINCKEL, G. *Control and optimization with PDE constraints*, vol. 164. Springer Science & Business Media, 2013.
8. EILENBERG, S., AND MONTGOMERY, D. Fixed point theorems for multi-valued transformations. *American Journal of Mathematics 68*, 2 (1946), 214–222.
9. FACCHINEI, F., AND KANZOW, C. Generalized Nash equilibrium problems. *4OR 5*, 3 (Sep 2007), 173–210.
10. FARRELL, P. E., CROCI, M., AND SUROWIEC, T. M. Deflation for semismooth equations. *Optimization Methods and Software 0*, 0 (2019), 1–24. DOI: 10.1080/10556788.2019.1613655.
11. FERRIS, M. C., AND PANG, J. S. Engineering and economic applications of complementarity problems. *SIAM Review 39* (1997), 669–713.
12. FÖLLMER, H., AND SCHIED, A. Convex measures of risk and trading constraints. *Finance Stoch. 6*, 4 (2002), 429–447.
13. GARREIS, S., SUROWIEC, T. M., AND ULBRICH, M. An interior-point approach for solving risk-averse PDE-constrained optimization problems with coherent risk measures. *SIAM Journal on Optimization* (2019). submitted.
14. GLICKSBERG, I. L. A further generalization of the Kakutani fixed theorem, with application to Nash equilibrium points. *Proc. Amer. Math. Soc. 3* (1952), 170–174.
15. GOLDBERG, H., KAMPOWSKY, W., AND TRÖLTZSCH, F. On Nemytskij operators in Lp-spaces of abstract functions. *Mathematische Nachrichten 155*, 1 (1992), 127–140.

16. HESTENES, M. R. Multiplier and gradient methods. *Journal of Optimization Theory and Applications 4*, 5 (Nov 1969), 303–320.

17. HILLE, E., AND PHILLIPS, R. S. *Functional analysis and semi-groups*. American Mathematical Society Colloquium Publications, vol. 31. American Mathematical Society, Providence, R. I., 1957. rev. ed.

18. HINTERMÜLLER, M., AND STENGL, S.-M. Generalization of path-following methods for generalized Nash equilibrium problems (tentative title). Preprint, in preparation.

19. HINTERMÜLLER, M., AND STENGL, S.-M. On the convexity of optimal control problems involving nonlinear PDEs or VIs and applications to Nash games (tentative title). Preprint, in preparation.

20. HINTERMÜLLER, M., AND SUROWIEC, T. A PDE-constrained generalized Nash equilibrium problem with pointwise control and state constraints. *Pac. J. Optim. 9*, 2 (2013), 251–273.

21. HINTERMÜLLER, M., SUROWIEC, T., AND KÄMMLER, A. Generalized Nash equilibrium problems in Banach spaces: theory, Nikaido-Isoda-based path-following methods, and applications. *SIAM J. Optim. 25*, 3 (2015), 1826–1856.

22. HOFFHUES, M. Stabilität stochastischer Optimierungsprobleme in Hilberträumen. Master's thesis, Philipps-Universität Marburg, Fachbereich Mathematik und Informatik, Marburg, Germany, 2019.

23. KANZOW, C., KARL, V., STECK, D., AND WACHSMUTH, D. The multiplier-penalty method for generalized Nash equilibrium problems in Banach spaces. *SIAM Journal on Optimization 29*, 1 (2019), 767–793.

24. KOURI, D. P., AND SUROWIEC, T. M. Risk-averse PDE-constrained optimization using the conditional value-at-risk. *SIAM Journal on Optimization 26*, 1 (2016), 365–396.

25. KOURI, D. P., AND SUROWIEC, T. M. Epi-regularization of risk measures for PDE-constrained optimization. *Math. Oper. Res.* (2018). to appear.

26. KOURI, D. P., AND SUROWIEC, T. M. Existence and optimality conditions for risk-averse PDE-constrained optimization. *SIAM/ASA Journal on Uncertainty Quantification 6*, 2 (2018), 787–815.

27. KOURI, D. P., AND SUROWIEC, T. M. A primal-dual algorithm for risk minimization. *Math. Programm. Ser. A.* (2018). submitted.

28. KOURI, D. P., AND SUROWIEC, T. M. Risk-averse optimal control of semilinear elliptic PDEs. *ESAIM COCV* (2018). submitted.

29. KRASNOSEL'SKII, M. A. *Topological methods in the theory of nonlinear integral equations*. Translated by A. H. Armstrong; translation edited by J. Burlak. A Pergamon Press Book. The Macmillan Co., New York, 1964.

30. LEUGERING, G., ENGELL, S., GRIEWANK, A., HINZE, M., RANNACHER, R., SCHULZ, V., ULBRICH, M., AND ULBRICH, S. *Constrained optimization and optimal control for partial differential equations*, vol. 160. Springer Science & Business Media, 2012.

31. LIONS, J. L. *Some aspects of the optimal control of distributed parameter systems*, vol. 6. Siam, 1972.

32. MIGNOT, F. Contrôle dans les inéquations variationelles elliptiques. *Journal of Functional Analysis 22*, 2 (1976), 130–185.

33. MORDUKHOVICH, B. S. *Variational analysis and generalized differentiation I: Basic theory*, vol. 330. Springer Science & Business Media, 2006.

34. MOSER, J. A sharp form of an inequality by N. Trudinger. *Indiana University Mathematics Journal 20*, 11 (1971), 1077–1092.

35. NISAN, N., ROUGHGARDEN, T., TARDOS, E., AND VAZIRANI, V. *Algorithmic Game Theory*. Cambridge University Press, 2007.

36. POWELL, M. J. D. A method for nonlinear constraints in minimization problems. *Optimization* (1969), 283–298.

37. RALPH, D., AND SMEERS, Y. Risk trading and endogenous probabilities in investment equilibria. *SIAM J. Optim. 25*, 4 (2015), 2589–2611.

38. ROCKAFELLAR, R. T., AND URYASEV, S. Optimization of conditional value-at-risk. *Journal of Risk 2* (2000), 21–41.

39. ROCKAFELLAR, R. T., AND URYASEV, S. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science 18*, 12 (2013), 33–53.
40. SHAPIRO, A. On concepts of directional differentiability. *J. Optim. Theory Appl. 66*, 3 (1990), 477–487.
41. SHAPIRO, A., DENTCHEVA, D., AND RUSZCZYNSKI, A. *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia, 2014.
42. TRÖLTZSCH, F. *Optimal control of partial differential equations: theory, methods, and applications*, vol. 112. American Mathematical Soc., 2010.
43. VAINBERG, M. M. *Variational methods for the study of nonlinear operators*. Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam, 1964. With a chapter on Newton's method by L. V. Kantorovich and G. P. Akilov. Translated and supplemented by Amiel Feinstein.
44. WERNER, D. *Funktionalanalysis*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2011.
45. ZOWE, J., AND KURCYUSZ, S. Regularity and stability for the mathematical programming problem in Banach spaces. *Applied Mathematics and Optimization 5*, 1 (1979), 49–62.