ON COMPUTING NASH EQUILIBRIUM IN STOCHASTIC GAMES

VIKAS VIKRAM SINGH∗, N. HEMACHANDRA∗, AND K.S. MALLIKARJUNA RAO∗

INDUSTRIAL ENGINEERING AND OPERATIONS RESEARCH,
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY, MUMBAI 400076, INDIA,
FEBRUARY, 2009.

Abstract. Using the fact that any two player discounted stochastic game with finite state and
action spaces can be recast as a non-convex constrained optimization problem, where each global
minima corresponds to a stationary Nash equilibrium, we present a sequential quadratic program-
ing based algorithm that converges to a KKT point. This KKT point is an \( \epsilon \)-Nash equilibrium for some \( \epsilon > 0 \) and under some suitable conditions we show that this KKT point corresponds to a stationary
Nash equilibrium. The algorithm updates the Hessian matrix of the Lagrangian function in a specific
way. We illustrate various difficulties that can arise while computing stationary Nash equilibrium of
the stochastic game using a variant of pollution tax model. One interesting feature of this model (in
an instance) is that it admits a Nash equilibrium which is independent of the discount factor close
to 1, an extension of Blackwell optimality in Markov decision processes.

Key words. Nonzero-sum stochastic game, Discounted value, Multiple Nash equilibria, \( \epsilon \)-Nash
equilibrium, Non-convex constrained optimization, Sequential quadratic programming, Lagrange
multipliers, MFCQ condition, Regular point, Pollution tax model.

AMS subject classifications. 91A15, 90C26, 91B76.

1. Introduction. Since the seminal work of Shapley [21], stochastic games have
come to constitute an important class of models that can capture game theoretic issues
among the decision makers involved, apart from accounting for random evolution of
the system. Nonzero sum stochastic games under total expected discounted pay-off
always admit a stationary Nash equilibrium [23]. The review article [20] summarizes
various algorithmic aspects of zero-sum stochastic games along with algorithms for
nonzero-sum stochastic games with special structure (like single controller and SER-
SIT). However the literature on algorithmic aspects of general nonzero-sum stochastic
games is rather limited. In [8], a nonlinear optimization formulation of the stochastic
game is formulated such that the global minimum is always zero and the global minima
corresponds to stationary Nash equilibrium. A comprehensive account of stochastic
games is given in the book by Filar and Vrieze [7]. For zero-sum games they also
propose a modified Newton algorithm and observe that certain special cases like SER-
SIT games, etc., can be solved as linear programs.

Recently Prasad, Bhatnagar and Hemachandra [18] gave an algorithm which con-
verges to a Nash equilibrium even in the case of multiple Nash equilibria. Prasad et.
al., first observe that Nash equilibrium can be obtained from a regular Karush-Kuhn-
Tucker (KKT) point of the associated optimization problem. Then they propose a
line search method that uses at each step the direction suggested by Herskovits [12].
To calculate the step length, a separate algorithm is given. The convergence of the
algorithm is based on the assumption that the optimal point is a regular point. A
feasible point is called a regular point if the gradients of constraints that are active
at that point are linearly independent; such a point is also said to satisfy Linear
Independence Constraint Qualification (LICQ) [17, 2].

∗Industrial Engineering and Operations Research, Indian Institute of Technology Bombay, Mum-
bai. E-mails : vikas.singh@iitb.ac.in, nh@iitb.ac.in, mallik.rao@iitb.ac.in.
In this article, we propose a new algorithm to compute the Nash equilibrium of two player nonzero sum stochastic games when the players use total expected discounted criteria. Our algorithm is based on trust region Sequential Quadratic Programming (SQP) methods which are preferred for their faster convergence. First, we identify the Lagrange multipliers for the global minima point of the non-convex optimization problem. Based on these multipliers, we update the Hessian matrix of the Lagrangian function in the intermediate steps. We note that these iterates satisfy the weaker Mangasarian Fromovitz constraint qualification (MFCQ) condition. Thus our algorithm converges to the KKT point of the optimization problem (OP). Such a KKT point can correspond to a Nash equilibrium. In other cases, it corresponds to an $\epsilon$-Nash equilibrium for suitable $\epsilon > 0$. This is because every KKT point need not be the global minima of the optimization problem (OP), i.e., each KKT point may not correspond to a Nash equilibrium of the stochastic game. We illustrate this using a single controller stochastic game and a variant of pollution tax model which is a generalization of a model given in [7] which does not have any special structure. In these examples, in fact, we exhibit a regular KKT point which is a local minima of the optimization problem (OP) (with the objective function has positive value) and corresponds to an $\epsilon$-Nash equilibrium. We give a sufficient condition on the KKT point under which it will correspond to a Nash equilibrium of stochastic game. Apart from having many $\epsilon$-Nash equilibria, a stochastic game can have Nash equilibrium that may be a regular point or a non-regular point. Also, a stochastic game can have multiple Nash equilibria. The pollution tax model exhibits all these phenomenon.

The pollution tax model introduced in our article is a generalization of the model given in [7]. Our model is not a SER-SIT (separable reward and state independent transition) and hence earlier known algorithms to compute Nash equilibrium do not work. We use our algorithm to compute the Nash equilibrium. Our findings imply that at higher discount rates, the firms indeed operate at low pollution levels (as desired by the pollution control board). Next, for a specific example of the pollution model we find a discount factor $\beta_0$ as well as a strategy pair such that for discount factors $\beta \in [\beta_0, 1)$ this strategy pair constitutes Nash equilibrium for both the firms. We call such a pair of strategies Blackwell-Nash equilibrium following the notion of Blackwell optimal strategies in discounted cost Markov decision process that have similar property.

We now briefly introduce some recent literature. Govindan and Wilson [10], Herings and Peeters [11] propose homotopy based methods to compute Nash equilibrium of stochastic games. There is also a lot of interest in learning based techniques to compute the Nash equilibrium. In [13], Hu and Wellman proposed a Q-learning algorithm for general-sum games. But their learning algorithm assures convergence only if the game has unique Nash equilibrium; see also [3]. Another learning algorithm called NashQ algorithm [14] is an extension of the Q-learning algorithm. However the issue of convergence in the case of multiple Nash equilibria is not addressed. Littman [15] has proposed friend-or-foe Q-learning (FFQ) that improves upon NashQ. FFQ algorithm converges to an equilibrium, not necessarily to Nash equilibrium, and assumes either full cooperation among players or the game is a zero sum game. Singh et. al., observed that in a two player iterated general sum game Nash convergence is assured either in strategies or in the very least in average pay-offs.

We now describe the structure of the rest of our paper. In Section 2, we describe the stochastic game and introduce notations that we use throughout this paper.
Section 3 contains the formulation of two player nonzero sum stochastic game as a nonlinear optimization problem. We prepare the ground work for the algorithm in Section 4. The algorithm is presented in Section 5 and convergence analysis is given in Section 6. Section 7 describes the pollution tax model and establishes the fact that this model admits a Blackwell-Nash equilibrium. In Section 8, we study multiple Nash equilibria, regularity of KKT point corresponding to the Nash equilibrium and ε-Nash equilibrium. Section 9 contains conclusions and topics for further research. There are two appendices at the end which contains the optimization problem corresponding to the pollution tax model and a technical lemma.

2. Infinite Horizon Discounted Stochastic game. A two player infinite horizon stochastic game with discounted payoff is a 7-tuple \((S, A^1, A^2, r^1, r^2, p, \beta)\) where  

(i) \(S = \{1, 2, \cdots, N\}\) is state space of the game. The generic element of \(S\) is denoted by \(s\).

(ii) \(A^1\) is the action set of player 1 and \(A^2\) is the action set of player 2. Let \(A^1(s) = \{1, 2, \cdots, m^1(s)\}\) (resp., \(A^2(s) = \{1, 2, \cdots, m^2(s)\}\) denote the set of actions available to player 1 (resp., player 2) when the state is at \(s\). Denote by \(m^1 = \sum_{s=1}^{N} m^1(s)\) and \(m^2 = \sum_{s=1}^{N} m^2(s)\).  

(iii) \(r^1, r^2\) are reward vectors of player 1 and player 2 respectively, i.e.,  
\[
\begin{align*}
r^1 &= (r^1(1, a_1, a_2), r^1(2, a_1, a_2), \cdots, r^1(N, a_1, a_2)) \\
r^2 &= (r^2(1, a_1, a_2), r^2(2, a_1, a_2), \cdots, r^2(N, a_1, a_2))
\end{align*}
\]
where \(r^1(s, a_1, a_2)\) (resp., \(r^2(s, a_1, a_2)\)) is instantaneous reward for player 1 (resp., player 2) if the state is \(s\) and actions chosen by the players are \(a_1\) and \(a_2\) respectively. We can take these rewards as non-negative.  

(iv) \(p(s'|s, a_1, a_2)\) denotes the probability of going to state \(s'\) in next period from the current state \(s\) when the player 1 chooses an action \(a_1 \in A^1(s)\) and player 2 chooses an action \(a_2 \in A^2(s)\).  

(v) \(\beta \in [0, 1)\) is the discount factor.

The game dynamics are as follows. If, at some time \(t\), the game is at state \(s\), player 1 chooses an action \(a_1 \in A^1(s)\) and player 2 chooses an action \(a_2 \in A^2(s)\) independent of each other, then player 1 receives an instantaneous reward of \(r^1(s, a_1, a_2)\) and player 2 receives \(r^2(s, a_1, a_2)\). Further the state of the game switches to new state \(s'\) at \(t+1\) with probability \(p(s'|s, a_1, a_2)\). This leads to the introduction of three stochastic processes \(X_t, \mathbb{A}^1_t, \mathbb{A}^2_t\). \(X_t\) denotes the state of the game, \(\mathbb{A}^1_t\) the action chosen by player 1 and \(\mathbb{A}^2_t\) the action chosen by the player 2 at time \(t\).

At any point of time, the action chosen by the players can be deterministic or randomized. Thus a decision rule \(f_t\) (resp., \(g_t\)) of player 1 (resp.,player 2) at time \(t\) is a function which assigns to each action the probability of taking that action at time \(t\); in general it can depend on all realized states up to and including time \(t\) and on all realized actions up to time \(t-1\). The sequence of decision rule is called the strategy of the players. \((f_1, f_2, \cdots, f_t, \cdots)\) and \((g_1, g_2, \cdots, g_t, \cdots)\) denote the strategies of player 1 and player 2 respectively.

If the initial state of the game is at \(s\) and the players chooses strategies \(f\) and \(g\) respectively, then the total expected \(\beta\)-discounted reward of player \(i, i = 1, 2\), is defined by

\[
v^i_j(s, f, g) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{f, g} \left( r^i_t(X_t, \mathbb{A}^1_t, \mathbb{A}^2_t) \right) .
\]

(2.1)
A pair of strategies \((f^*, g^*)\) of player 1 and player 2 is said to be a Nash equilibrium if

\[
v_1^1(s, f, g^*) \leq v_1^1(s, f^*, g^*)
\]

and

\[
v_2^2(s, f^*, g) \leq v_2^2(s, f^*, g^*)
\]

for all strategies \(f, g\) of player 1 and player 2 respectively and for all \(s \in S\). The salient, and also most appealing, feature of the above definition is that unilateral deviations from \((f^*, g^*)\) either by player 1 and player 2 are not worthwhile. \(v_\beta^i(s, f^*, g^*)\) is called the Nash equilibrium payoff of the \(i\)th player when initial state is \(s, i = 1, 2\).

A pair of strategies \((f_\epsilon, g_\epsilon)\) of player 1 and player 2 is said to be an \(\epsilon\)-Nash equilibrium, if there exist an \(\epsilon > 0\) such that

\[
v_1^1(s, f, g_\epsilon) \leq v_1^1(s, f_\epsilon, g_\epsilon) + \epsilon
\]

and

\[
v_2^2(s, f_\epsilon, g) \leq v_2^2(s, f_\epsilon, g_\epsilon) + \epsilon
\]

for all strategies \(f, g\) of player 1 and player 2 respectively and for all \(s \in S\). This means that unilateral deviations form \(\epsilon\)-Nash equilibrium by a player may give that player at most \(\epsilon\) extra value. Also, the case of \(\epsilon\) is zero corresponds to Nash equilibrium.

It is by now a classical result that an infinite horizon discounted stochastic game with finite state and action spaces admits a Nash equilibrium. Moreover, the Nash equilibrium can be chosen to be stationary (see Theorem 3.8.1, Chapter 3 in [7]). Recall that a strategy is called stationary if it does not depend on time \(t\), i.e., \(f_t = f\) and \(g_t = g\) for all \(t\). We represent the stationary strategy of player 1 (resp., player 2) by \(f\) (resp., \(g\)) instead of \((f(f, f, \cdots))\) (resp., \((g(g, g, \cdots)))\). From now on we restrict our attention to stationary strategies only. Let \(F_S, G_S\) denote the sets of all stationary randomized strategies of player 1 and player 2 respectively and are given by

\[
F_S = \left\{ f = (f(1), f(2), \cdots, f(N)) \mid f(s) = (f(s, 1), f(s, 2), \cdots, f(s, m^1(s))) \right\},
\]

\[
f(s, a^1) \geq 0 \text{ and } \sum_{a^1=1}^{m^1(s)} f(s, a^1) = 1 \forall a^1 \in A^1(s), s \in S \right\\
\]

\[
G_S = \left\{ g = (g(1), g(2), \cdots, g(N)) \mid g(s) = (g(s, 1), g(s, 2), \cdots, g(s, m^2(s))) \right\},
\]

\[
g(s, a^2) \geq 0 \text{ and } \sum_{a^2=1}^{m^2(s)} g(s, a^2) = 1 \forall a^2 \in A^2(s), s \in S \right\\
\]

We now introduce some notation. The state space \(S\) has \(N\) elements, action spaces \(A^1(s)\) has \(m^1(s)\) elements, \(i = 1, 2\) and \(s \in S\). We regard \(f(s)\) as \(m^1(s)\)-dimensional row vector and \(g(s)\) as \(m^2(s)\)-dimensional column vector for all \(s \in S\). For \(i = 1, 2\), \(s, s' \in S\) and \(f \in F_S, g \in G_S\).

- \(R^i(s) = [r^i(s, a^1, a^2)]_{a^1=1, a^2=1}^{m^1(s), m^2(s)}\) where \(R^i(s)\) is reward matrix of player-\(i\) at state \(s\).
A stochastic game is given by optimization problem. The nonlinear optimization problem corresponding to our ways to compute Nash equilibrium is to formulate the stochastic game as a nonlinear optimization problem. There can be non-stationary Nash equilibrium. Our objective in this article is to find a stationary Nash equilibrium by finding one global minima of (OP). It is shown that the optimization problem (OP) has a global minima with value zero and the global minima corresponds to the stationary Nash equilibrium of the stochastic game (see Chapter 3, Section 3.8 in [7] for details). Note that there can be many stationary Nash equilibria and hence there will be multiple global minima. 

Using the above notation, if \( f \in F_S \) and \( g \in G_S \) are the strategies chosen by the players and the initial state of game be \( s \) then \( (2.1) \) can be written as (see [7])

\[
v^k_{s}(s,f,g) = \sum_{t=0}^{\infty} \beta^t [P^t(f,g)]_s v^i(f,g).
\]

### 3. Stochastic Game as a Nonlinear Optimization Problem.

One of the ways to compute Nash equilibrium is to formulate the stochastic game as a nonlinear optimization problem. The nonlinear optimization problem corresponding to our stochastic game is given by

\[
\begin{align*}
\min_{v^1, v^2, f, g} & \sum_{k=1}^{2} \{1_N^{T} [v^k - r^k(f,g) - \beta P(f,g)v^k] \} \\
\text{s.t.} & \quad (i) \quad R^1(s)g(s) + \beta \sum_{s'=1}^{N} P(s'|s)g(s)v^1(s') \leq v^1(s)1_{m^1(s)} \\
& \quad (ii) \quad f(s)R^2(s) + \beta \sum_{s'=1}^{N} f(s)P(s'|s)v^2(s') \leq v^2(s)1_{m^2(s)}^{T} \\
& \quad (iii) \quad \sum_{a^1=1}^{m^1(s)} f(s,a^1) = 1 \\
& \quad (iv) \quad \sum_{a^2=1}^{m^2(s)} g(s,a^2) = 1 \\
& \quad (v) \quad f(s,a^1) \geq 0 \\
& \quad (vi) \quad g(s,a^2) \geq 0
\end{align*}
\]

for all \( a^1 \in A^1(s), a^2 \in A^2(s) \) and \( s \in S \).

It is shown that the optimization problem \([OP]\) has a global minima with value zero and the global minima corresponds to the stationary Nash equilibrium of the stochastic game (see Chapter 3, Section 3.8 in [7] for details). Note that there can be multiple stationary Nash equilibria and hence there will be multiple global minima. There can be non-stationary Nash equilibrium. Our objective in this article is to find a stationary Nash equilibrium by finding one global minima of \([OP]\).
In $\textbf{OP}$, there are total of $2N + 2m^1 + 2m^2$ constraints in $2N + m^1 + m^2$ variables. The inequalities (i) and (ii) corresponds to dynamic programming equations and are nonlinear inequality constraints. We index them by $I_1$. The equalities (iii) and (iv) and inequalities (v) and (vi) capture the fact that $f(s), s \in S$ and $g(s), s \in S$ are probability distributions. We index the constraints (iii) and (iv) by $E$ and (v) and (vi) by $I_2$. Note the constraints in $E$ are linear equality constraints and the constraints in $I_2$ are linear inequality constraints.

We introduce a vector $x \in \mathbb{R}^{2N+m^1+m^2}$ defined by:

- $x_{jN+i} = \nu^{j+1}(i)$ for $i = 1, 2, \ldots, N$ and $j = 0, 1$
- $x_{2N+m^1(0)+\ldots+m^1(j)+i} = f(j+1, i)$ for $j = 0, 1, \ldots, N-1$ and $i = 1, 2, \ldots, m^1(j+1)$
- $x_{2N+m^1+m^2(0)+\ldots+m^2(j)+i} = g(j+1, i)$ for $j = 0, 1, \ldots, N-1$ and $i = 1, 2, \ldots, m^2(j+1)$.

Using this notation, we can now recast the optimization problem $\textbf{OP}$ as follows:

\[
\begin{aligned}
\min_x & \quad h(x) \\
\text{s.t.} & \quad c_i(x) \leq 0 \text{ for all } i \in I_1 \\
& \quad c_i(x) = 0 \text{ for all } i \in E \\
& \quad c_i(x) \leq 0 \text{ for all } i \in I_2
\end{aligned}
\]

for appropriately chosen functions $h(\cdot), c_i(\cdot), i \in I_1 \cup E \cup I_2$. In the sequel, we use the problems $\textbf{OP}$ and $\textbf{OP}'$ interchangeably as they are equivalent.

If $x$ is a feasible point of the $\textbf{OP}'$ with an objective function value $h(x) > 0$ then the strategy part of $x$ forms an $\epsilon$-Nash equilibrium $[8, 7]$ with

\[
\epsilon = \frac{h(x)}{1-\beta}.
\]

Remark 3.1. One can easily see that the diagonal elements of the Hessian matrix of the objective function are zeros and hence the trace is zero. Due to this the matrix will have both positive as well as negative eigenvalues. Consequently the objective function is non-convex. Further, the nonlinear inequality constraints are non-convex. Hence the optimization problem is non-convex optimization problem. So, computing a global minima is difficult.

4. Initial Feasible Point and Some Auxiliary Results. In this section, we present some results which are useful in designing the algorithm. Our first task is to find an initial feasible point in such a way that all the intermediate iterates of the algorithm are also feasible. In fact, our method finds an relative interior point of the feasible space of $\textbf{OP}$ and the algorithm ensures that the subsequent points are also interior points. Lemma 4.1 shows that any feasible point of the optimization problem for discount factor higher than $\beta$ is also feasible for $\beta$. This is very helpful when the algorithm gets stuck at a KKT point which is not global minima (and hence not Nash equilibrium). Lemma 4.3 establishes the fact that at the global minima, Lagrange multipliers exist and satisfy the KKT condition. This provides us a suitable candidate for Lagrange multipliers while updating the Hessian matrix of the Lagrangian function of the optimization problem.

We now indicate the procedure to find a relative interior initial feasible point of the optimization problem $\textbf{OP}$ for the execution of algorithm 5.3. To achieve this, we
choose the initial feasible point such that the inequality constraints in $I_1$ are strictly feasible.

Let $x = ((v^1)^T, (v^2)^T, f, g)^T$ be any $2N + m_1 + m_2$ dimensional vector. Let

$$f^0(s, a^1) = \frac{1}{m^1(s)} \text{ and } g^0(s, a^1) = \frac{1}{m^2(s)}$$

for $a^1 \in A^1(s), a^2 \in A^2(s)$ and $s \in S$.

If we use these initial strategy vectors in the optimization problem (OP), as observed in [18], the optimization problem decomposes into two linear programs given by

$$\begin{align*}
\min_{v^1} & \{1_N^T(v^1 - r^1(f^0, g^0) - \beta P(f^0, g^0)v^1)\} \\
\text{s.t} & \\
R^1(s)g^0(s) + \beta \sum_{s'=1}^N P(s'|s)g^0(s')v^1(s') \leq v^1(s)1_{m^1(s)} \text{ for all } s \in S
\end{align*}$$

(4.1)

$$\begin{align*}
\min_{v^2} & \{1_N^T(v^2 - r^2(f^0, g^0) - \beta P(f^0, g^0)v^2)\} \\
\text{s.t} & \\
f^0(s)R^2(s) + \beta \sum_{s'=1}^N f^0(s')P(s'|s)v^2(s') \leq v^2(s)1_{m^2(s)}^T \text{ for all } s \in S
\end{align*}$$

(4.2)

where the first LP is in variables $(x_1, x_2, \cdots, x_N)$ and the other LP is in variables $(x_{N+1}, x_{N+2}, \cdots, x_{2N})$. Observe that these vectors are the value vectors of player 1 and player 2 respectively and hence they are bounded. Thus the feasible regions of these linear programs are bounded polyhedrons. In [13] the initial feasible point was obtained by solving the (4.1) and (4.2) by phase one of simplex algorithm which gives the required corner points of the these polyhedrons. But we choose interior points of these polyhedrons in our algorithm. Indeed, we first find the center of the largest ball which can be inscribed in these polyhedrons. Finding the center and the radius of this largest ball inside the bounded polyhedron is equivalent to solving a certain linear program (as discussed in Section 8.5 in [19]). The first $N$ components of initial feasible point $x^0$ is the center of the largest ball in first polyhedron. The next $N$ components are given by the center of the largest ball in the second polyhedron. The remaining components are the above initial feasible strategies. From the construction it is obvious that the initial feasible point $x^0$ is an interior point.

We now show that initial feasible point for higher discount rate than $\beta$ is also feasible for $\beta$.

**Lemma 4.1.** Let $F_\beta$ and $F_{\beta'}$ denote the feasible space of optimization problem (OP) for fixed discount factor $\beta$ and $\beta'$ respectively. If $\beta \leq \beta'$ then $F_{\beta'} \subseteq F_\beta$.

**Proof.** The proof is a simple consequence of the fact that the discount factor $\beta$ appears only in the constraints (i) and (ii) of (OP), where the terms multiplying $\beta$ are nonnegative. \qed

**Remark 4.2.** In view of above, the initial point of the algorithm for solving optimization problem (OP) at the discount factor $\beta'$ can be used as an initial feasible point for all discount factors $\beta$ such that $\beta \leq \beta'$. This fact is helpful in avoiding local minima; see Section 8 for details.
In the following lemma, we show that Lagrange multipliers exist and satisfy the KKT conditions at the global minima.

**Lemma 4.3.** *At the global minima of the optimization problem the corresponding Lagrange multipliers exist.*

**Proof.**

The objective function can be written as

\[
\sum_{s=1}^{N} \sum_{a^1=1}^{m^1(s)} f(s, a^1) \left[ v^1(s) - [R^1(s)g(s)]_{a^1} - \beta \sum_{s'=1}^{N} [P(s'|s)g(s)]_{a^1} v^1(s') \right] + \sum_{s=1}^{N} \sum_{a^2=1}^{m^2(s)} g(s, a^2) \left[ v^2(s) - [f(s)R^2(s)]_{a^2} - \beta \sum_{s'=1}^{N} [f(s')P(s'|s)]_{a^2} v^2(s') \right].
\]

(4.3)

Let \((v^1*, v^2*, f^*, g^*)\) be a global minima of the optimization problem.

Let us define the Lagrange multipliers as follows:

\[
\begin{align*}
\lambda_s^{1, a^1*} &= f^*(s, a^1) \\
\lambda_s^{2, a^2*} &= g^*(s, a^2) \\
\gamma^*_s &= 0 \\
\delta^*_s &= 0 \\
\alpha_s^{1, a^1*} &= v^1*(s) - [R^1(s)g^*(s)]_{a^1} - \beta \sum_{s'=1}^{N} [P(s'|s)g^*(s)]_{a^1} v^1*(s') \\
\alpha_s^{2, a^2*} &= v^2*(s) - [f^*(s)R^2(s)]_{a^2} - \beta \sum_{s'=1}^{N} [f^*(s')P(s'|s)]_{a^2} v^2*(s')
\end{align*}
\]

(4.4)

for all \(a^1 \in A^1(s), a^2 \in A^2(s)\) and \(s \in S\). Here \(\lambda_s^{1, a^1*}\) is the Lagrange multiplier of the constraint (i) in (OP) when state is \(s \in S\) and action \(a^1 \in A^1(s)\) is used by player 1. Similarly \(\lambda_s^{2, a^2*}\) is the Lagrange multiplier of the constraint (ii) in (OP) when state is \(s \in S\) and action \(a^2 \in A^2(s)\) is used by player 2. The Lagrange multipliers for equality constraints (OP) (iii) and (OP) (iv) at state \(s\) are \(\gamma^*_s\) and \(\delta^*_s\) respectively. \(\alpha_s^{1, a^1*}\) and \(\alpha_s^{2, a^2*}\) denote the Lagrange multipliers for the constraints (OP) (v) and (vi) respectively when state is \(s \in S\), action \(a^1 \in A^1(s)\) and \(a^2 \in A^2(s)\). Let \(\lambda^*\) denote the vector of Lagrange multipliers defined as in (4.4). Then we can see that the following condition holds,

\[
\nabla L(v^1*, v^2*, f^*, g^*, \lambda^*) = 0
\]

(4.5)

where \(L(\cdot)\) is the Lagrangian function of the optimization problem (OP). As \(((v^1*)^T, (v^2*)^T, f^*, (g^*)^T)\) is a global minima of the optimization problem (OP), we note that each term of the objective function (4.3) is zero. Hence, we can write

\[
\begin{align*}
\lambda_s^{1, a^1*} \left( [R^1(s)g^*(s)]_{a^1} + \beta \sum_{s'=1}^{N} [P(s'|s)g^*(s)]_{a^1} v^1*(s') - v^1*(s) \right) &= 0 \\
\lambda_s^{2, a^2*} \left( [f^*(s)R^2(s)]_{a^2} + \beta \sum_{s'=1}^{N} [f^*(s')P(s'|s)]_{a^2} v^2*(s') - v^2*(s) \right) &= 0
\end{align*}
\]

(4.6)
\begin{equation}
\begin{aligned}
\alpha^{1,a^1_*} f^*(s,a^1) = 0 \\
\alpha^{2,a^2_*} g^*(s,a^2) = 0
\end{aligned}
\end{equation}

(4.7)

\begin{align}
\lambda^{1,a^1_*} \geq 0, \alpha^{1,a^1_*} \geq 0, \alpha^{2,a^2_*} \geq 0, \gamma^*_s \in \mathbb{R}, \delta^*_s \in \mathbb{R}
\end{align}

(4.8)

for all $a^1 \in A^1(s), a^2 \in A^2(s)$ and $s \in S$. Then, the equations (4.6)-(4.8) constitute the complementary slackness conditions so that (4.5)-(4.8) are the KKT conditions of (OP) at $(v^1, v^2)^T, (f^*, g^*)^T, (\lambda^*)^T)$. Thus, at the global minima of the optimization problem (OP) the KKT conditions hold with the above Lagrange multipliers.

Remark 4.4. The global minima need not be a regular point. In fact, we show such examples in Section 8.

5. The Algorithm. In this section, we present the algorithm to compute the Nash equilibrium. Our approach is based on trust region SQP [17, 6]. The SQP methods are iterative methods where we start from some iterate and solve a quadratic program to find its global minima. This global minima is used to generate the next iterate. This gives a descent direction $d$ along which the feasible point $x$ remains feasible. The quadratic program related to (OP) is given in Subsection 5.1. The procedure to choose initial feasible point is already discussed in previous section. The construction of Hessian matrix which is needed in the quadratic program is given in Subsection 5.2. Finally the algorithm is presented in Subsection 5.3. The algorithm generates a sequence of points that converge to a point with zero descent direction. So, the limit point of algorithm will be a KKT point of (OP) by Lemma 5.1. Later we give a sufficient condition on the KKT point under which it is indeed a global minima of (OP) and hence a Nash equilibrium of the given stochastic game.

5.1. Quadratic program. Let $x$ be a feasible point of (OP) and $\rho$ be the trust region radius. The global minima $d$ of the quadratic program $\text{QP}(x, \rho)$ given by

\begin{equation}
\begin{aligned}
\min_d q(d) &= d^T \nabla h(x) + \frac{1}{2} d^T H(x) d \\
\text{s.t.} & \\
& c_i(x) + a^T_i(x)d \leq 0, \ i \in I_1 \\
& c_i(x) + a^T_i(x)d = 0, \ i \in E \\
& c_i(x) + a^T_i(x)d \leq 0, \ i \in I_2 \\
& \|d\|_\infty \leq \rho
\end{aligned}
\end{equation}

(\text{QP}(x, \rho))

gives the required descent direction. Here $a_i(x) = \text{grad}(c_i(x))$ and $H(x)$ is the Hessian matrix of the Lagrangian function of (OP) whose construction is given in Subsection 5.2. There are many methods available for solving the quadratic program $\text{QP}(x, \rho)$. The active set method is very popular and it converges in finite steps when the quadratic program is convex. But the main drawback of the active set method is that it is slow if the number of variables in quadratic program are large. The gradient projection method for solving quadratic program is designed to make the rapid changes in the active set method. The interior point method is another one which is more efficient for the convex quadratic program. For detailed information about these methods, see Chapter-18 of [17]. For every feasible point $x$ of (OP) we solve the $\text{QP}(x, \rho)$. Since $\text{QP}(x, \rho)$ is feasible and Hessian matrix $H(x)$, whose construction
given in Subsection 5.2 ensures positive definiteness, we can always find a global minima \( d \) of \( QP(x, \rho) \). When the global minima of \( QP(x, \rho) \) is zero, then \( x \) will be KKT point of \( OP' \). This is content of the following lemma whose proof can be found in [9, 17]. For the sake of completeness, we provide the proof in Appendix B in a more general setting.

**Lemma 5.1.** If \( d = 0 \) solves the \( QP(x, \rho) \) then \( x \) is a KKT point of the optimization problem \( OP' \).

### 5.2. Construction of Hessian matrix \( H \).

To calculate the Hessian matrix \( H \) at some feasible point \( x \), we need to know the Lagrange multipliers of the constraints in \( OP' \). Note that the linear constraints in \( OP' \) do not contribute to the Hessian matrix \( H \). So we need to consider the Lagrange multipliers corresponding to constraints which belong to index set \( I_1 \) only, i.e., the constraints \((i)\) and \((ii)\) in \( OP \). Thus the number of constraints that contribute to Hessian is \( m_1 + m_2 \).

For each \( s \in S \), let

\[
\lambda^1_s = \left( \lambda^{1,1}_s, \lambda^{1,2}_s, \ldots, \lambda^{1,m(s)}_s \right)
\]
\[
\lambda^2_s = \left( \lambda^{2,1}_s, \lambda^{2,2}_s, \ldots, \lambda^{2,m(s)}_s \right)
\]

be vectors with positive components and define

\[
\hat{\lambda} = \left( \lambda^1_1, \lambda^1_2, \ldots, \lambda^1_N, \lambda^2_1, \lambda^2_2, \ldots, \lambda^2_N \right)^T.
\]

Here, \( \lambda^1_s \) (resp., \( \lambda^2_s \)) denotes the candidate Lagrange multipliers at point \( x \) associated with inequality constraints corresponding to the state \( s \) in the set of inequality constraints \( OP'(i) \) (resp., \( OP'(ii) \)).

Set \( c(x) = (c_1(x), c_2(x), \ldots, c_{m_1+m_2}(x))^T \) where the components of the vector \( c(x) \) are in the same order as in the definition of the vector \( \lambda \). Now the Lagrangian function is given by

\[
L(x, \lambda) = h(x) + \lambda^T c(x).
\]

The Lemma 4.3 suggests the following candidate Lagrange multipliers:

\[
\begin{align*}
\lambda^{1,a^1}_s = x_{2N+m^1(1)+m^1(2)+\ldots+m^1(s-1)+a^1} \\
\lambda^{2,a^2}_s = x_{2N+m^1+m^2(1)+m^2(2)+\ldots+m^2(s-1)+a^2}
\end{align*}
\]

for \( a^1 \in A^1(s) \), \( a^2 \in A^2(s) \) and \( s \in S \). Define the Hessian matrix \( H(x) \) of \( L(x, \lambda) \) by

\[
H(x) = \nabla^2_{xx} L(x, \lambda)|_{\lambda = \hat{\lambda}}.
\]

As the Hessian matrix \( H(x) \) need not be positive definite, we use Levenberg-Marquardt modification to make it positive definite [5, 11].

### 5.3. An SQP Based Algorithm for Nash equilibrium.

We now present the algorithm.

**Step-1** Choose the input parameters \( \rho^0 > 0 \), \( \sigma \in (0, 1) \). Set \( k = 0 \)

**Step-2** Find an interior point \( x^0 \) as an initial feasible point by the procedure of Section 4 such that

\[
c_i(x^0) < 0, \; i \in I_1
\]
Step-3 Choose $\rho \geq \rho^0$ and go to Step-4

Step-4 Calculate $H(x^k)$ using (5.2), make it positive definite if necessary. Solve QP($x^k, \rho$) and obtain $d$, the global minima point of QP($x^k, \rho$).
If $d = 0$, then go to Step-6.

Step-5 If $c_i(x^k + d) < 0$ for $i \in I_1$ and

$$h(x^k) - h(x^k + d) \geq \sigma (q(0) - q(d))$$ (5.3)

Then

$$\rho^{(k)} = \rho, d^k = d, x^{k+1} = x^k + d^k, k \leftarrow k + 1$$

Go to Step-3
Else
Reduce the trust region radius $\rho \leftarrow \rho^2$.
Go to Step-4

Step-6 Stop

6. Convergence of Algorithm. In this section we first show that our algorithm converges to a KKT point of (OP). Later, we give simple conditions when such KKT point corresponds to Nash equilibrium of the stochastic game.

Lemma 6.1. The Mangasarian Fromovitz constraint qualification (MFCQ) condition for the optimization problem (OP) holds at every iterate of the algorithm.

Proof. A feasible point $x$ of the problem (OP) is said to satisfy the MFCQ if the following conditions hold [16], [2]:

1. The vectors $a_i := \nabla c_i(x)$ where $i \in E$ are linearly independent.

2. There exist a vector $y$ that satisfies $y^T a_i = 0, i \in E$ and $y^T a_i < 0, i \in A^0$ where $A^0$ denotes the set of active inequality constraints.

For optimization problem (OP), there are $2N$ number of equality constraints and we can easily see that gradients of these constraints are linearly independent for any feasible point. So the first condition for MFCQ holds.

At every iterate of the algorithm all the nonlinear constraints of the optimization problem (OP) are strictly feasible. So the active set at an iterate of the algorithm will contain some of the nonnegative constraints. For a state $s$ there will be $m^1(s)$ number of nonnegative constraints and one equality constraint w.r.t player 1. Let $l = m^1(1) + m^1(2) + \cdots + m^1(s - 1)$. Thus these $m^1(s)$ constraints are:

$$\begin{align*}
-x_{2N+l+1} \leq 0 \\
-x_{2N+l+2} \leq 0 \\
\vdots \\
-x_{2N+l+m^1(s)} \leq 0 
\end{align*}$$

$$x_{2N+l+1} + x_{2N+l+2} + \cdots + x_{2N+l+m^1(s)} = 1.$$

Suppose some subset of the above inequality constraints are active. We form a vector $y \in \mathbb{R}^{2N+m^1(s)}$ as below. We take its components corresponding to this subset as some positive numbers. We take components corresponding to inactive constraints such that their sum is negative sum of already chosen components. For example,
suppose in (6.1) the first $m^1(s) - 2$ are active. We form $y \in \mathbb{R}^{2N + m^1 + m^2}$ by first choosing $m^1(s) - 2$ components as some positive reals as:

\[
\begin{align*}
y_{2N+l+1} &> 0 \\
y_{2N+l+2} &> 0 \\
&\vdots \\
y_{2N+l+m^1(s)-2} &> 0
\end{align*}
\]

and then taking other 2 components such that

\[
y_{2N+l+m^1(s)-1} + y_{2N+l+m^1(s)} = -\left(y_{2N+l+1} + y_{2N+l+2} + \cdots + y_{2N+l+m^1(s)-2}\right).
\]

We can see that with this choice of components of $y$ condition (2) of MFCQ holds for the constraints corresponding to state $s$ and w.r.t. player 1. Similarly, we can choose other components of $y$ and the second condition of MFCQ holds for the constraints w.r.t. player 2 at state $s$. We repeat this for all states $s \in S$. We complete formation of $y$ by taking its first 2 components as some reals so that conditions (1) and (2) above are satisfied. Thus, MFCQ conditions hold at every iterate of the algorithm.

**Remark 6.2.** We can prove the above lemma in a different way also. Using the fact that the active set at every iterate contains a subset of the nonnegative constraints, it is easy to show that the gradients of all the equality constraints together with those of the active nonnegative constraints are linearly independent. It means that every iterate of our algorithm in fact satisfies the LICQ, i.e., these points are regular points. It is known that LICQ implies MFCQ.

**Lemma 6.3.** At every iterate of our algorithm the optimization problem (OP') is feasible.

**Proof.** We start from a feasible point $x^0$ of (OP') at which

\[c_i(x^0) < 0 \text{ for all } i \in I_1\]

Each $c_i(\cdot)$, $i \in I_1$, is a continuous function and therefore there exist a small ball centered at $x^0$ in which all the constraints are strictly feasible, i.e,

\[c_i(x^0 + d) < 0 \text{ for all } i \in I_1.\]

By using the constraints corresponding to the set $E$ and $I_2$ of the QP($x^k, \rho$) we can easily see that equality and non-negative constraints of (OP') will be satisfied at next point $x^1$. Similarly we can show that at each iteration the optimization problem (OP') is feasible.

From the Lemma 6.3, we can say that the quadratic program QP($x^k, \rho$) will have a solution because the QP($x^k, \rho$) will be feasible and $H(x^k)$ is positive definite. Thus Step-4 and Step-5 of the algorithm give a descent direction $d$ where there is a sufficient reduction in the objective function value of (OP') from Lemma 6.4.

**Lemma 6.4.** The sufficient reduction condition (5.3) of the algorithm holds and the inner iteration of the algorithm terminates in a finite number of steps.

**Proof.** Since, the Mangasarian Fromovitz constraint qualification holds at every iteration of the algorithm and the standard assumption in Section 3 of [9] holds for the optimization problem (OP'), the proof follows directly from the Lemma-5 and Lemma-6 of [9].

**Theorem 6.5.** The sequence of points generated by the algorithm (along a subsequence, if needed) converges to a KKT point of the optimization problem (OP').
Proof. Let \( \{x^k\} \) be a sequence of points generated by the algorithm. Since the sequence is bounded it has a convergent subsequence. With an abuse of notation, let \( x^k \to x^* \). Since \( h \) and \( d \) both are continuous functions of \( x \), as \( k \to \infty \)

\[
h(x^k) \to h(x^*), \quad d^k \to d^*.\]

At \( x^k \) the condition \[[5.3]\] holds, i.e.,

\[
h(x^k) - h(x^{k+1}) \geq \sigma(q(0) - q(d^k)). \tag{6.3}
\]

Taking limit as \( k \to \infty \) in the above we get

\[
h(x^*) - h(x^*) \geq \sigma(0 - q(d^*)) \]

which implies

\[
q(d^*) \geq 0. \tag{6.4}
\]

For every \( k \), \( d^k \) is an unique global minima of the quadratic program \( \text{QP}(x^k, \rho) \). Since \( d^k \) is feasible solution of \( \text{QP}(x^k, \rho) \) this implies

\[
c_i(x^k) + \alpha_i^T(x^k)d^k \leq 0 \text{ for all } i \in I_1
\]

\[
c_i(x^k) + \alpha_i^T(x^k)d^k = 0 \text{ for all } i \in E
\]

\[
c_i(x^k) + \alpha_i^T(x^k)d^k \leq 0 \text{ for all } i \in I_2
\]

\[
\|d^k\|_{\infty} \leq \rho^k
\]

so that as \( k \to \infty \)

\[
c_i(x^*) + \alpha_i^T x^* d^* \leq 0 \text{ for all } i \in I_1
\]

\[
c_i(x^*) + \alpha_i^T x^* d^* = 0 \text{ for all } i \in E
\]

\[
c_i(x^*) + \alpha_i^T x^* d^* \leq 0 \text{ for all } i \in I_2
\]

\[
\|d^*\|_{\infty} \leq \bar{\rho}.
\]

This implies that \( d^* \) is also feasible for \( \text{QP}(x^*, \bar{\rho}) \). Hence \( d^* \) is also an unique global minima of \( \text{QP}(x^*, \bar{\rho}) \). \( d = 0 \) is always a feasible solution of the \( \text{QP}(x^k, \rho) \) for some \( \rho \geq \rho^0 \) and hence \( q(d^k) < 0 \) for all \( k \). Letting \( k \to \infty \), we have \( q(d^*) \leq 0 \). Combining this with \[[6.4],\] we obtain \( q(d^*) = 0 \). By the uniqueness of the global minima of the quadratic program \( \text{QP}(x^*, \rho) \), we deduce that \( d^* = 0 \). Now Lemma \[5.1\] completes the proof of the theorem. \( \square \)

Remark 6.6. The KKT point \( x^* \) will be an \( \epsilon \)-Nash equilibrium with \( \epsilon \) as given in \[[3.1] \].

The theorem below gives a sufficient condition when \( x^* \) can be a Nash equilibrium.

Theorem 6.7. Every KKT point of the optimization problem (OP) will be Nash equilibrium of the stochastic game if the corresponding Lagrange multipliers are as proposed in Lemma \[4.3\].

Proof. Recall that at the KKT point \( x^{*T} = (v^{1*T}, v^{2*T}, f^*, g^{*T}) \) the Lagrange
Singh et al. multipliers proposed in Lemma 4.3 are as:

\[
\begin{align*}
\lambda_1^{s,a^1} &= f^*(s,a^1) \\
\lambda_2^{s,a^2} &= g^*(s,a^2) \\
\gamma_s^* &= 0 \\
\delta_s^* &= 0 \\
\alpha_1^{s,a^1} &= v_1^*(s) - [R^1(s)g^*(s)]a^1 - \beta \sum_{s' = 1}^N [P(s'|s)g^*(s)]a^1 v_1^*(s') \\
\alpha_2^{s,a^2} &= v_2^*(s) - [f^*(s)R^2(s)]a^2 - \beta \sum_{s' = 1}^N [f^*(s)P(s'|s)]a^2 v_2^*(s')
\end{align*}
\]

(6.5)

Then from complementary slackness part of KKT condition we can see that every term of the optimization problem is zero. Since optimization problem (OP) is feasible at \( x^* \) and objective function value is zero, \( x^* \) is global minima of the optimization problem (OP). So, \((f^*,g^*)\) will be Nash equilibrium of the stochastic game. \( \square \)

**Remark 6.8.** The condition in Theorem 6.7 is only a sufficient condition and not necessary. We will see an example where the KKT point is a Nash equilibrium but the corresponding Lagrange multipliers are different from those given in (6.5). Such a Nash equilibrium point is not a regular point.

**Remark 6.9.** An obvious condition for the KKT point to be Nash equilibrium is that the objective function is zero at this point.

7. Pollution Tax Model and Blackwell-Nash equilibrium. Stochastic games have several real life applications. One of the familiar examples is fishery game \([7]\) which belongs to a class of stochastic games called AR-AT games. Pollution tax model which is also given in \([7]\) is an example of stochastic game belonging to the class SER-SIT. We give a short description of this model.

There are two firms which contribute to the emission of a certain pollutant and suppose that the government can detect only the combined emissions. For the sake of simplicity, let us suppose that they both have the opportunity to produce only two types of emissions: Dirty(D) or Clean(C). When at least one of the firm produces D the government will impose a tax (tax=3) on both the firms. Now suppose that for the different combinations of the production levels C and D the profits for the firms in a certain period can be expressed with the following bimatrix game payoff: This situation can be modeled as a two state stochastic game in the following way:

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>(4,5)</td>
<td>(3,8)</td>
</tr>
<tr>
<td>D</td>
<td>(7,4)</td>
<td>(6,7)</td>
</tr>
</tbody>
</table>

There are two states of the game, state 1 and state 2, and each state is associated with a particular tax value. The payoff matrix for each state is shown in the following tables:

- **State 1 (tax = 0):**
  - C: (4,5)  D: (3,8)
  - C: (1,0)  D: (0,1)
  - C: (0,1)  D: (6,7)

- **State 2 (tax = 3):**
  - C: (1,2)  D: (0,5)
  - C: (1,0)  D: (0,1)
  - C: (4,1)  D: (3,4)
The data of the problem are of a special type. Namely, the rewards in corresponding cells of both states differ by a fixed amount (tax=3), that is, the reward is the sum of a term that only depends on the action and a term that only depends on the state. This property is called SER, or separable rewards. Furthermore, the transitions belonging to a pair of actions in one state are the same as in the other state. This property is called SIT, or state independent transitions. This model has a special structure, i.e, SER-SIT and to solve such game algorithms are available in the literature [20].

7.1. A Variant of Pollution tax model. We propose a more realistic pollution tax model than the model discussed above. We model this as a nonzero sum stochastic game and hence finding Nash equilibrium is much more difficult. We use the algorithm given in Subsection 5.3 to study the pollution tax model. We model this as a two player (two firms) 3-state stochastic game.

Consider the model as above but now assume that the policy of the government is to impose no tax (tax = 0) if both the firms do not pollute. It imposes less but the same tax (tax = 2) if the pollution level is at intermediate level, i.e, in (0, 4]. If the pollution level is high, i.e., more than 4, then the firms pay higher tax (tax = 4).

Note that more production implies higher level of pollution. Thus, we can say that immediate reward (profit) of one firm increases with its pollution level and decreases with the pollution level of other firm. We assume that tax at time \( t - 1 \) is paid at time \( t \) by both the firms.

State space. State of the game is the tax imposed on the firms by the government. Thus \( S = \{0, 2, 4\} \).

Action Spaces. The actions of the firms are the level of pollution that they cause. It is different in different states.

- Action space for Firm 1 at state 0, i.e., \( A^1(0) = \{0, 3, 5\} \).
- Action space for Firm 1 at state 2, i.e., \( A^1(2) = \{0, 2, 3\} \).
- Action space for Firm 1 at state 4, i.e., \( A^1(4) = \{0, 2\} \).
- Action space for Firm 2 at state 0, i.e., \( A^2(0) = \{0, 3, 5\} \).
- Action space for Firm 2 at state 2, i.e., \( A^2(2) = \{0, 2, 4\} \).
- Action space for Firm 2 at state 4, i.e., \( A^2(4) = \{0, 3\} \).

Transition probabilities. The transition probabilities are stationary and are given by

\[
\begin{align*}
p(s' = 0|s, a^1, a^2) &= 1 \text{ if } a^1 + a^2 = 0 \\
p(s' = 2|s, a^1, a^2) &= 1 \text{ if } a^1 + a^2 \in (0, 4] \\
p(s' = 4|s, a^1, a^2) &= 1 \text{ if } a^1 + a^2 > 4
\end{align*}
\]

for all \( a^1 \in A^1(s) \), \( a^2 \in A^2(s) \) and \( s \in S \).

Reward. We assume that the rewards of both the players are stationary. Recall that tax at time \( t - 1 \) is the state of the game at time \( t \). So, the rewards of both the players are given by

\[
\begin{align*}
    r^1(s, a^1, a^2) &= 5 + \frac{1}{4} a^1 - \frac{1}{5} a^2 - s \\
    r^2(s, a^1, a^2) &= 6 + \frac{1}{3} a^2 - \frac{1}{5} a^1 - s
\end{align*}
\]

for all \( a^1 \in A^1(s) \), \( a^2 \in A^2(s) \) and \( s \in S \). From (7.2), we note that the reward functions are separable in state and action variables. However, for state independent
transition property to hold, we need action sets to be same in each state. But in
above, the action sets are different for different states. Thus this stochastic
game does not satisfy SER-SIT property and hence can not be solved by the methods
used in [7].

We compute the Nash equilibrium of the stochastic game for various discount
factors. We found the KKT point of the optimization problem at which the objective
function value is zero, i.e, the algorithm gives the Nash equilibrium of the Pollution
tax model. The Table 7.1 summarizes the computational results.

<table>
<thead>
<tr>
<th>Discount factor</th>
<th>s = 0</th>
<th>s = 2</th>
<th>s = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>β = 0</td>
<td>Firm 1</td>
<td>(0, 0, 1)</td>
<td>(0, 0, 1)</td>
</tr>
<tr>
<td></td>
<td>Firm 2</td>
<td>(0, 0, 1)</td>
<td>(0, 0, 1)</td>
</tr>
<tr>
<td>β = 0.1</td>
<td>Firm 1</td>
<td>(0, 0, 1)</td>
<td>(0, 0, 1)</td>
</tr>
<tr>
<td></td>
<td>Firm 2</td>
<td>(0, 0, 1)</td>
<td>(0, 0, 1)</td>
</tr>
<tr>
<td>β = 0.67</td>
<td>Firm 1</td>
<td>(1, 0, 0)</td>
<td>(1, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>Firm 2</td>
<td>(1, 0, 0)</td>
<td>(1, 0, 0)</td>
</tr>
<tr>
<td>β = 0.7</td>
<td>Firm 1</td>
<td>(1, 0, 0)</td>
<td>(1, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>Firm 2</td>
<td>(1, 0, 0)</td>
<td>(1, 0, 0)</td>
</tr>
<tr>
<td>β = 0.8</td>
<td>Firm 1</td>
<td>(1, 0, 0)</td>
<td>(1, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>Firm 2</td>
<td>(1, 0, 0)</td>
<td>(1, 0, 0)</td>
</tr>
<tr>
<td>β = 0.9</td>
<td>Firm 1</td>
<td>(1, 0, 0)</td>
<td>(1, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>Firm 2</td>
<td>(1, 0, 0)</td>
<td>(1, 0, 0)</td>
</tr>
<tr>
<td>β = 0.98</td>
<td>Firm 1</td>
<td>(1, 0, 0)</td>
<td>(1, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>Firm 2</td>
<td>(1, 0, 0)</td>
<td>(1, 0, 0)</td>
</tr>
</tbody>
</table>

7.2. Blackwell-Nash equilibrium. For finite state-action Markov decision
models with discounted criteria, there is a discount factor β0 such that for all β ∈ [β0, 1)
the same policy, say π, is optimal [19]. Such a policy is called Blackwell optimal
policy. MDPs can be considered as a special case of stochastic games. So, we say
(f∗, g∗) Blackwell-Nash equilibrium of a stochastic game if there exists a β0 such that
for β ∈ [β0, 1) (f∗, g∗) is Nash equilibrium of the stochastic game. In this section
we prove the existence of Blackwell-Nash equilibrium for this particular example of
pollution tax model. The optimization problem corresponding to this stochastic game
is given in Appendix A.

From Table 7.1 we can see that for a high discount factor, say β = 0.9 the Nash
equilibrium is

\[ f^* = ((1, 0, 0), (1, 0, 0), (1, 0)) , \quad g^* = ((1, 0, 0), (1, 0, 0), (1, 0)). \]
Substituting these in (2.2), we obtain, after a straightforward computation,

\[ v^1 = v^1(f^*, g^*) = \left( \frac{5 \beta + 3}{1 - \beta}, \frac{4 \beta + 1}{1 - \beta}, \frac{1 - \beta}{1 - \beta} \right)^T, \]

\[ v^2 = v^2(f^*, g^*) = \left( \frac{6 \beta + 4}{1 - \beta}, \frac{4 \beta + 2}{1 - \beta}, \frac{1 - \beta}{1 - \beta} \right)^T. \]

In terms of vector \( x^* \), (see Appendix A)

\[
(x_1^*, x_2^*, x_3^*) = \left( \frac{5 \beta + 3}{1 - \beta}, \frac{4 \beta + 1}{1 - \beta}, \frac{1 - \beta}{1 - \beta} \right)
\]

\[
(x_4^*, x_5^*, x_6^*) = \left( \frac{6 \beta + 4}{1 - \beta}, \frac{4 \beta + 2}{1 - \beta}, \frac{1 - \beta}{1 - \beta} \right)
\]

\[
(x_7^*, x_8^*, x_9^*) = (1, 0, 0) \quad (x_{10}^*, x_{11}^*, x_{12}^*) = (1, 0, 0) \quad (x_{13}^*, x_{14}^*) = (1, 0)
\]

\[
(x_{15}^*, x_{16}^*, x_{17}^*) = (1, 0, 0) \quad (x_{18}^*, x_{19}^*, x_{20}^*) = (1, 0, 0) \quad (x_{21}^*, x_{22}^*) = (1, 0)
\]

We can see that at this point \( x^* = ((v^1)^T, (v^2)^T, f^*, g^*) \) the objective function value is zero for all the discount factor \( \beta \in [0, 1] \). For this \( x^* \) we consider all the nonlinear inequality constraints as a function of \( \beta \). We find the common interval of \( \beta \in [\beta_0, \beta_1] \subseteq [0, 1] \) for which all the inequality constraints (A.1)-(A.16) are feasible. The equality constraints (A.17)-(A.22) and nonnegative constraints (A.23) are feasible at \( x^* \) and these constraints are independent of \( \beta \). So, for all the discount factor \( \beta \), which belongs to the interval \([\beta_0, \beta_1]\), \((f^*, g^*)\) will be the Nash equilibrium of stochastic game as defined above. If \([\beta_0, \beta_1] = [\beta_0, 1]\) then \((f^*, g^*)\) is Blackwell-Nash strategy. So, we find \( \beta_0 \) such that all inequality constraints are simultaneously feasible at \( x^* \).

For this \( x^* \), the constraint (A.1) \( c_1(\beta) = 0 \) for all \( \beta \). From constraint (A.2) we have \( c_2(\beta) = (0.75 - 2\beta) \), \( c_2(\beta) \leq 0 \) for all \( \beta \geq 0.375 \) which gives

\[ c_2(\beta) \leq 0 \quad \forall \beta \in [0.375, 1). \]  

(7.3)

From constraint (A.3) \( c_3(\beta) = (1.25 - 4\beta) \), so

\[ c_3(\beta) \leq 0 \quad \forall \beta \in [0.3125, 1). \]  

(7.4)

From the constraint (A.4) \( c_4(\beta) = 0 \) for all \( \beta \). From constraint (A.5) \( c_5(\beta) = (0.5 - 2\beta) \), so

\[ c_5(\beta) \leq 0 \quad \forall \beta \in [0.25, 1). \]  

(7.5)

From (A.6) \( c_6(\beta) = (0.75 - 2\beta) \). This implies

\[ c_6(\beta) \leq 0 \quad \forall \beta \in [0.375, 1). \]  

(7.6)

The constraint (A.7) \( c_7(\beta) = 0 \) for all \( \beta \). From constraint (A.8) \( c_8(\beta) = (0.5 - 2\beta) \). This implies

\[ c_8(\beta) \leq 0 \quad \forall \beta \in [0.25, 1). \]  

(7.7)

The constraint (A.9) \( c_9(\beta) = 0 \) for all \( \beta \). From constraint (A.10) we have \( c_{10}(\beta) = (1 - 2\beta) \). This implies

\[ c_{10}(\beta) \leq 0 \quad \forall \beta \in [0.5, 1). \]  

(7.8)
From (A.11) we have $c_{11}(\beta) = (1.67 - 4\beta)$. This implies
\[ c_{11}(\beta) \leq 0 \forall \beta \in [0.4175, 1). \quad (7.9) \]

The constraint (A.12) $c_{12}(\beta) = 0$ for all $\beta$.
The constraint (A.13) becomes $c_{13}(\beta) = (0.67 - 2\beta)$. This implies
\[ c_{13}(\beta) \leq 0 \forall \beta \in [0.335, 1). \quad (7.10) \]

From constraint (A.14) $c_{14}(\beta) = (1.34 - 2\beta)$. This implies
\[ c_{14}(\beta) \leq 0 \forall \beta \in [0.67, 1). \quad (7.11) \]

The constraint (A.15) $c_{15}(\beta) = 0$ for all $\beta$.
Finally, from constraint (A.16) $c_{16}(\beta) = (1 - 2\beta)$. So
\[ c_{16}(\beta) \leq 0 \forall \beta \in [0.5, 1). \quad (7.12) \]

From equation (7.3)-(7.12) we can see that the common interval of discount factor is $[0.67, 1)$ for which all the constraints are feasible at $x^*$. So, $(f^*, g^*)$ is a Nash equilibrium for all $\beta \in [0.67, 1)$ and hence a Blackwell-Nash equilibrium. We summarize the above in the following theorem.

**Theorem 7.1.** The instance of the pollution tax model described in Subsection 7.1 admits a Blackwell-Nash equilibrium.

### 8. Multiple Nash Equilibria, Regularity of Nash equilibrium point and \( \epsilon \)-Nash equilibrium of stochastic game

In this section we will discuss about the multiple Nash equilibria and local minima of the pollution tax model. We give multiple Nash equilibria at the same discount factor. We also discuss the role of regularity condition at the Nash equilibrium and non-Nash equilibrium.

**8.1. Multiple Nash equilibria.** We know that for a given discount factor $\beta$ stochastic game can have multiple Nash equilibria. If we start algorithm from different initial point then we can get multiple Nash equilibria at the same discount factor $\beta$. We know by Lemma 4.1 that initial point of algorithm for higher discount factor will be the initial point for lower discount factor value. We found some multiple Nash equilibria in pollution tax model. We compute three Nash equilibria at $\beta = 0.67$ by starting from different initial points. These are summarized in Table 8.1.

<table>
<thead>
<tr>
<th>Discount factor as in Lemma 4.1</th>
<th>$s = 0$</th>
<th>$s = 2$</th>
<th>$s = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta' = 0.98$</td>
<td>Firm 1 (0, 0, 1)</td>
<td>(0, 0, 1)</td>
<td>(1, 0)</td>
</tr>
<tr>
<td></td>
<td>Firm 2 (0, 0, 1)</td>
<td>(0, 0, 1)</td>
<td>(1, 0)</td>
</tr>
<tr>
<td>$\beta' = 0.7$</td>
<td>Firm 1 (1, 0, 0)</td>
<td>(1, 0, 0)</td>
<td>(1, 0)</td>
</tr>
<tr>
<td></td>
<td>Firm 2 (1, 0, 0)</td>
<td>(0.483, 0.517)</td>
<td>(1, 0)</td>
</tr>
<tr>
<td>$\beta' = 0.67$</td>
<td>Firm 1 (1, 0, 0)</td>
<td>(1, 0, 0)</td>
<td>(1, 0)</td>
</tr>
<tr>
<td></td>
<td>Firm 2 (1, 0, 0)</td>
<td>(1, 0, 0)</td>
<td>(1, 0)</td>
</tr>
</tbody>
</table>
8.2. Regularity of Nash equilibrium. We know that if the KKT point of any optimization problem is regular then the corresponding Lagrange multipliers are unique. We compute three Nash equilibria at $\beta = 0.67$. The first Nash equilibrium strategy of Table 8.1 together with corresponding value vector is given below

$$x^* = (10.7422, 8.4422, 8.1972, 14.5346, 12.6046, 11.7382, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0).$$

At this point the total number of active constraints together with equality constraints are 22 and gradient of all the constraints are linearly independent, i.e, $x^*$ is regular point of $(\Omega)$. So, the corresponding Lagrange multipliers are unique and same as proposed in Lemma 4.3. Let $\lambda_{\text{ineq}}$ denote the Lagrange multipliers corresponding to the nonlinear inequality constraints, $\lambda_{\text{eq}}$ denote the Lagrange multipliers for the equality constraints and $\lambda_+$ correspond to nonnegative constraints. These Lagrange multipliers are:

$$\begin{align*}
\lambda_{\text{ineq}}^* &= (0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0) \\
\lambda_{\text{eq}}^* &= (0, 0, 0, 0, 0, 0) \\
\lambda_+^* &= (1.25, 0.5, 0, 0.5859, 0.25, 0, 0, 1.041, 1.67, 0.67, 0, 0.7595, 0.67, 0, 0.2931) 
\end{align*}$$

(8.1)

Every Nash equilibrium need not correspond to a regular point. The second and third Nash equilibria in Table 8.1 are not regular and we get two set of Lagrange multipliers for these points. These points and Lagrange multipliers are as follows:

The second point is

$$x^* = (15.1515, 11.4207, 11.1515, 18.1818, 16.1818, 14.1818, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0.483, 0, 0.517, 1, 0).$$

At this point total number of active constraints together with equality constraints are 22 and number of variable are also 22 but the gradients of active constraints are linearly dependent. We have two sets of Lagrange multipliers corresponding to this point as below.

$$\begin{align*}
\lambda_{1\text{ineq}}^{1*} &= (1, 0, 0, 1, 0, 0, 1, 0, 1.3464, 0, 0, 0.6536, 0, 0, 1, 0) \\
\lambda_{1\text{eq}}^{1*} &= (-0.2984, 5.6056, 0, 0, 0, 0) \\
\lambda_+^{1*} &= (0, 1.0522, 0.1552, 0, 1.2234, 0, 1.9742, 0, 0.34, 1.01, 0, 0.67, 0, 0, 0.34) 
\end{align*}$$

(8.2)

and

$$\begin{align*}
\lambda_{2\text{ineq}}^{2*} &= (1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0.483, 0, 0.517, 1, 0) \\
\lambda_{2\text{eq}}^{2*} &= (0, 0, 0, 0, 0, 0) \\
\lambda_+^{2*} &= (0, 1.7496, 1.43, 0, 0.783, 0.533, 0, 1.9966, 0, 0.34, 1.01, 0, 0.67, 0, 0.34) 
\end{align*}$$

(8.3)

The third Nash equilibrium correspond to the point

$$x^* = (15.1515, 13.1515, 11.1515, 18.1818, 16.1818, 14.1818, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0).$$
KKT point and corresponding Lagrange multipliers are as follows: 

$$\lambda^{1*}_{ineq} = (1, 0, 0, 1, 0, 1, 0, 0, 0.6959, 0, 0, 0.8502, 0, 0.4539, 1, 0)$$
$$\lambda^{1*}_{eq} = (5.5295, -4.9213, 0, 0, 0)$$
$$\lambda^{1*}_+ = (0, 1.18, 2.5492, 0, 0.3108, 0, 0, 0.84, 0, 0.34, 1.01, 0, 0.67, 0, 0, 0.34)$$

(8.4)

$$\lambda^{2*}_{ineq} = (1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0)$$
$$\lambda^{2*}_{eq} = (0, 0, 0, 0, 0, 0)$$
$$\lambda^{2*}_+ = (0, 0.59, 1.43, 0, 0.84, 0.59, 0, 0.84, 0, 0.34, 1.01, 0, 0.67, 0, 0, 0.34)$$

(8.5)

In both the cases first set of Lagrange multipliers are computed from the algorithm and second set are as in Lemma [4.3]

### 8.3. Existence of Local Minima and ϵ-Nash equilibrium

The algorithm converges to KKT point of the optimization problem [OP]. We found KKT points at $\beta = 0.2$ and $\beta = 0.3$ and these points are regular also but the objective function value is nonzero, i.e, these points are the local minima of the optimization problem. This means our algorithm may find only $\epsilon$-Nash equilibrium with $\epsilon$ as in (3.1). We may avoid this local minima by starting from different initial point from Lemma [4.1]

For this pollution tax model at $\beta = 0.2$, and $\beta = 0.3$ if we start from initial point calculated at $\beta = 0.98$ then algorithm gives the Nash equilibrium. For $\beta = 0.2$ the KKT point and corresponding Lagrange multipliers are as follows:

$$x = (6.475, 3.175, 1.125, 8.32, 5.39, 3.25, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$
$$\lambda_{ineq} = (0, 0, 0.8, 0, 0, 1, 0, 1.2, 0, 0, 0.8, 0, 0, 1, 0, 1.2)$$
$$\lambda_{eq} = (1.348, 0, -0.65, 0.639, 0, -0.225)$$
$$\lambda_+ = (0, 0.616, 1.034, 0.34, 0.25, 0, 0.2556, 0, 0, 0.366, 0.614, 0.912, 0.67, 0, 0.774, 0)$$

(8.6)

The order of Lagrange multipliers are same as the order of constraints in the optimization problem defined above. This pair $(x, \lambda)$ satisfy the KKT conditions of the optimization problem [OP]. We find that the point $x$ is regular point but at this point objective function value is $h(x) = 0.836$, i.e, $x$ is a local minima of the optimization problem. The $\epsilon$-Nash equilibrium is

$$f_\epsilon = ((1, 0, 0), (0, 0, 1), (0, 1)), \quad g_\epsilon = ((1, 0, 0), (0, 0, 1), (0, 1)).$$

with $\epsilon = 1.045$. 

For $\beta = 0.3$ the KKT point and Lagrange multipliers are

\[
\begin{align*}
x &= (6.4972, 3.3736, 1.4121, 8.7843, 5.8543, 3.7143, 1, 0, 0, 0, 0, 1, 0, 1, 0.706, 0.294, 0, 0, 0, 1, 0, 1) \\
\lambda_{ineq} &= (0.8863, 0, 0.0896, 0, 0, 1.2286, 0.7955, 0, 0, 0, 0.7882, 0, 0, 0.9118, 0, 1.3) \\
\lambda_{eq} &= (2.6253, 0.5164, -1.2027, 0.081, -0.7712, 0.2004) \\
\lambda_+ &= (0, 0.1, 0.21, 1.4743, .052, .2324, 0, .2241, 0, 0, .0485, 1.0154, .7614, 0, 0.3922, 0)
\end{align*}
\]

(8.7)

The pair $(x, \lambda)$ satisfies the KKT conditions of the optimization problem. We find that the point $x$ is a regular point but at this point objective function value is $h(x) = 0.2019$, i.e., $x$ is a local minima of the optimization problem. The $\epsilon$-Nash equilibrium is

\[
f_\epsilon = ((1, 0, 0), (0, 0, 1), (0, 1)), \quad g_\epsilon = ((0.706, 0.294, 0), (0, 0, 1), (0, 1)),
\]

with $\epsilon = 0.2884$.

The Table 8.2 gives Nash equilibrium for some discount factors where we were able to bypass $\epsilon$-Nash equilibrium by starting with an initial point that is feasible for $\beta = 0.98$.

<table>
<thead>
<tr>
<th>Discount factor used for initial point</th>
<th>Discount factor for N.E. of stochastic game</th>
<th>$s = 0$</th>
<th>$s = 2$</th>
<th>$s = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta' = 0.98$</td>
<td>$\beta = 0.2$</td>
<td>Firm 1</td>
<td>(0, 0, 1)</td>
<td>(0, 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Firm 2</td>
<td>(0, 0, 1)</td>
<td>(0, 1)</td>
</tr>
<tr>
<td>$\beta' = 0.98$</td>
<td>$\beta = 0.3$</td>
<td>Firm 1</td>
<td>(0, 0, 1)</td>
<td>(1, 0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Firm 2</td>
<td>(0, 0, 1)</td>
<td>(0, 1)</td>
</tr>
<tr>
<td>$\beta' = 0.98$</td>
<td>$\beta = 0.4$</td>
<td>Firm 1</td>
<td>(0, 0, 1)</td>
<td>(1, 0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Firm 2</td>
<td>(0, 0, 1)</td>
<td>(0, 1)</td>
</tr>
<tr>
<td>$\beta' = 0.98$</td>
<td>$\beta = 0.5$</td>
<td>Firm 1</td>
<td>(1, 0, 0)</td>
<td>(0, 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Firm 2</td>
<td>(1, 0, 0)</td>
<td>(1, 0)</td>
</tr>
<tr>
<td>$\beta' = 0.98$</td>
<td>$\beta = 0.6$</td>
<td>Firm 1</td>
<td>(0, 0, 1)</td>
<td>(1, 0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Firm 2</td>
<td>(0, 0, 1)</td>
<td>(1, 0)</td>
</tr>
</tbody>
</table>

We now illustrate $\epsilon$-Nash equilibrium in a particular stochastic game. This is the single controller game given by Raghavan and Filar [20]. The components of the stochastic game are follows:

**State Space.** $S = \{1, 2, 3\}$

**Action Spaces.** $A^i(s) = \{1, 2\}$ $\forall$ $s \in S$ and $i = 1, 2$

The rewards and transition probabilities are given by the following tables.

For such games, an algorithm to compute Nash equilibrium is available [20]. We find the KKT point of the optimization problem by using our algorithm. This KKT
The objective function value is 1.024. Thus, for $\epsilon = 5.12$, which is very large, the $\epsilon$-Nash equilibrium is

$$f_\epsilon = ((0.7946, 0.2054), (0, 1), (0, 1)), \quad g_\epsilon = ((0.5385, 0.4615), (0.5625, 0.4375), (0, 1)).$$

We can avoid this local minima by starting from different initial point which is calculated at higher discount factor (see Lemma 4.1) or by changing suitable trust region radius. We next compute the global minima at $\beta = 0.8$ by using our algorithm starting with the same initial feasible point but with different trust region radius. The
global minima and corresponding Lagrange multipliers are as follows:

\[
\begin{align*}
  x^* &= (28.3784, 35.210811, 22.6471, 25.184941, 0.2353, 0.7647, 0.1, 0.0753, 0.9247, 0.5385, 0.4615, 1, 0, 0.5714, 0.4286) \\
  \lambda_{ineq}^* &= (0.2353, 0.7647, 0, 1, 0.0753, 0.9247, 0.5385, 0.4615, 1, 0, 0.5714, 0.4286) \\
  \lambda_q^* &= (0, 0, 0, 0, 0) \\
  \lambda_+^* &= (0, 0, 7, 0, 0, 0, 0, 0, 0, 2.2047, 0, 0)
\end{align*}
\] (8.9)

The point \( x^* \) is a regular point of the optimization problem with Lagrange multipliers same as those proposed in Lemma 4.3.

9. Conclusions and Further Work. In this article we gave an algorithm, along with its convergence analysis, to compute stationary Nash equilibrium of a two player discounted stochastic game. We also introduced the notion of Blackwell-Nash equilibrium and showed that an instance of pollution tax model admits existence of such strategies. We discussed the various difficulties that arise while computing Nash equilibrium of stochastic games using this pollution tax model. Finding Nash equilibrium for two player stochastic games with large state and action sets can be attempted with our algorithm as SQP methods that we primarily use are scalable [17].

We were able to overcome the difficulty of getting trapped at a local minima by using an initial feasible point at a higher discount factor. We do not have any theoretical justification for this at this point. We hope to come back to this issue at a later point of time. Other topics for further research include, apart from devising algorithms to compute the Nash equilibrium of multi-player stochastic games, studying the existence of Blackwell-Nash equilibrium for stochastic games, etc.

Appendix A. Optimization problem for Pollution tax model.

We give here details of the optimization problem for the pollution tax model of Section 7.1. We denote the decision variables of the optimization problem in terms of the general vector as follows. The total number of decision variables for this optimization problem are 22 and \( m^1(0) = 0, m^1(1) = 3, m^1(2) = 3, m^1(3) = 2, m^2(0) = 0, m^2(1) = 3, m^2(2) = 3, m^2(3) = 2 \). Let \( x \in \mathbb{R}^{22} \) where

- \( x_{3j+i} = v_{j+1}(i) \) for \( i = 1, 2, 3 \) and \( j = 0, 1 \)
- \( x_{6+ m^1(0)+ \ldots + m^1(j)+i} = f(j + 1, i) \) for \( j = 0, 1, 2 \) and \( i = 1, 2, \ldots, m^1(j+1) \)
- \( x_{6+ m^1+m^2(0)+ \ldots + m^2(j)+i} = g(j + 1, i) \) for \( j = 0, 1, 2 \) and \( i = 1, 2, \ldots, m^2(j+1) \).

The objective function and constraints of optimization problem corresponding to
our pollution tax model are as follows:

\[
\begin{align*}
\min_{x \in \mathbb{R}^{22}} & \quad h(x) = -x_7 [(\beta x_{15} - 1)x_1 + \beta x_{16}x_2 + \beta x_{17}x_3 + 5x_{15} + 4.4x_{16} + 4x_{17}] \\
- & \quad x_8 [-x_1 + \beta x_{15}x_2 + \beta(x_{16} + x_{17})x_3 + 5.75x_{15} + 5.15x_{16} + 4.75x_{17}] \\
- & \quad x_9 [-x_1 + \beta x_3 + 6.25x_{15} + 5.65x_{16} + 5.25x_{17}] \\
- & \quad x_{10} [\beta x_{18}x_1 + (\beta(x_{19} + x_{20}) - 1)x_2 + 3x_{18} + 2.6x_{19} + 2.2x_{20}] \\
- & \quad x_{11} [(\beta(x_{18} + x_{19}) - 1)x_2 + \beta x_{20}x_3 + 3.5x_{18} + 3.1x_{19} + 2.7x_{20}] \\
- & \quad x_{12} [(\beta x_{18} - 1)x_2 + \beta(x_{19} + x_{20})x_3 + 3.75x_{18} + 3.35x_{19} + 2.95x_{20}] \\
- & \quad x_{13} [\beta x_{21}x_1 + \beta x_{22}x_2 - x_3 + x_{21} + 0.4x_{22}] \\
- & \quad x_{14} [\beta x_{21}x_2 + (\beta x_{22} - 1)x_3 + 1.5x_{21} + 0.9x_{22}] \\
- & \quad x_{15} [(\beta x_7 - 1)x_4 + \beta x_5x_7 + \beta x_9x_6 + 6x_7 + 5.4x_8 + 5x_9] \\
- & \quad x_{16} [-x_4 + \beta x_5x_7 + \beta(x_8 + x_9)x_6 + 7x_7 + 6.4x_8 + 6x_9] \\
- & \quad x_{17} [-x_4 + \beta x_6 + 7.67x_7 + 7.07x_8 + 6.67x_9] \\
- & \quad x_{18} [\beta x_{10}x_4 + (\beta(x_{11} + x_{12}) - 1)x_5 + 4x_{10} + 3.6x_{11} + 3.4x_{12}] \\
- & \quad x_{19} [(\beta(x_{10} + x_{11}) - 1)x_5 + \beta x_{12}x_6 + 4.67x_{10} + 4.27x_{11} + 4.07x_{12}] \\
- & \quad x_{20} [(\beta x_{10} - 1)x_5 + \beta(x_{11} + x_{12})x_6 + 5.34x_{10} + 4.94x_{11} + 4.74x_{12}] \\
- & \quad x_{21} [\beta x_{13}x_4 + \beta x_{14}x_5 - x_6 + 2x_{13} + 1.6x_{14}] \\
- & \quad x_{22} [\beta x_{13}x_5 + (\beta x_{14} - 1)x_6 + 3x_{13} + 2.6x_{14}] \\
\end{align*}
\]

\[
\begin{align*}
(\beta x_{15} - 1)x_1 + \beta x_{16}x_2 + \beta x_{17}x_3 + 5x_{15} + 4.4x_{16} + 4x_{17} & \leq 0 \quad (A.1) \\
- x_1 + \beta x_{15}x_2 + \beta(x_{16} + x_{17})x_3 + 5.75x_{15} + 5.15x_{16} + 4.75x_{17} & \leq 0 \quad (A.2) \\
- x_1 + \beta x_3 + 6.25x_{15} + 5.65x_{16} + 5.25x_{17} & \leq 0 \quad (A.3) \\
\beta x_{18}x_1 + (\beta(x_{19} + x_{20}) - 1)x_2 + 3x_{18} + 2.6x_{19} + 2.2x_{20} & \leq 0 \quad (A.4) \\
(\beta x_{18} + x_{19}) - 1)x_2 + \beta x_{20}x_3 + 3.5x_{18} + 3.1x_{19} + 2.7x_{20} & \leq 0 \quad (A.5) \\
(\beta x_{18} - 1)x_2 + \beta(x_{19} + x_{20})x_3 + 3.75x_{18} + 3.35x_{19} + 2.95x_{20} & \leq 0 \quad (A.6) \\
\beta x_{21}x_1 + \beta x_{22}x_2 - x_3 + x_{21} + 0.4x_{22} & \leq 0 \quad (A.7) \\
\beta x_{21}x_2 + (\beta x_{22} - 1)x_3 + 1.5x_{21} + 0.9x_{22} & \leq 0 \quad (A.8) \\
(\beta x_7 - 1)x_4 + \beta x_5x_7 + \beta x_9x_6 + 6x_7 + 5.4x_8 + 5x_9 & \leq 0 \quad (A.9) \\
- x_4 + \beta x_5x_7 + \beta(x_8 + x_9)x_6 + 7x_7 + 6.4x_8 + 6x_9 & \leq 0 \quad (A.10) \\
- x_4 + \beta x_6 + 7.67x_7 + 7.07x_8 + 6.67x_9 & \leq 0 \quad (A.11) \\
\beta x_{10}x_4 + (\beta(x_{11} + x_{12}) - 1)x_5 + 4x_{10} + 3.6x_{11} + 3.4x_{12} & \leq 0 \quad (A.12) \\
(\beta(x_{10} + x_{11}) - 1)x_5 + \beta x_{12}x_6 + 4.67x_{10} + 4.27x_{11} + 4.07x_{12} & \leq 0 \quad (A.13) \\
(\beta x_{10} - 1)x_5 + \beta(x_{11} + x_{12})x_6 + 5.34x_{10} + 4.94x_{11} + 4.74x_{12} & \leq 0 \quad (A.14) \\
\beta x_{13}x_4 + \beta x_{14}x_5 - x_6 + 2x_{13} + 1.6x_{14} & \leq 0 \quad (A.15) \\
\beta x_{13}x_5 + (\beta x_{14} - 1)x_6 + 3x_{13} + 2.6x_{14} & \leq 0 \quad (A.16)
\end{align*}
\]
\begin{align*}
\text{(A.17)} & \quad x_7 + x_8 + x_9 = 1 \\
\text{(A.18)} & \quad x_{10} + x_{11} + x_{12} = 1 \\
\text{(A.19)} & \quad x_{13} + x_{14} = 1 \\
\text{(A.20)} & \quad x_{15} + x_{16} + x_{17} = 1 \\
\text{(A.21)} & \quad x_{18} + x_{19} + x_{20} = 1 \\
\text{(A.22)} & \quad x_{21} + x_{22} = 1
\end{align*}
\begin{align*}
x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}, x_{22} \geq 0
\end{align*}

Appendix B. Proof of the Lemma 5.1. Let us consider the general nonlinear optimization problem

\[
\begin{aligned}
\min_{x} & \quad h(x) \\
\text{s.t} & \quad c_i(x) \leq 0 \text{ for all } i \in I \\
& \quad c_i(x) = 0 \text{ for all } i \in E
\end{aligned}
\]

where \(I\) denote the index set of inequality constraints and \(E\) denote the set of equality constraints. Let the cardinality of set \(I\) is \(m_1\) and cardinality of set \(E\) is \(m_2\). We solve the optimization problem defined above by trust region sequential quadratic programming method. The trust region \(QP(x, \rho)\) for the above optimization problem at some feasible point \(x \in \mathbb{R}^n\) and trust region radius \(\rho > 0\) is given below.

\[
\begin{aligned}
\min_{d} & \quad q(d) = d^T \nabla h(x) + \frac{1}{2} d^T H(x)d \\
\text{s.t} & \quad c_i(x) + a_i^T(x)d \leq 0, \quad i \in I \\
& \quad c_i(x) + a_i^T(x)d = 0, \quad i \in E \\
& \quad \|d\|_{\infty} \leq \rho
\end{aligned}
\]

where \(a_i(x) = \nabla c_i(x)\) and let \(A(x)^T = [a_1, a_2, \ldots, a_{m_1+m_2}]\).

It is already known that if \(d = 0\) solves (B.2) then \(x\) is a KKT point of (B.1) \cite{9}. We give this result in detail.

Let \(\lambda\) be a \(m_1 + m_2\) dimensional vector and

\[
\mu^T = (\mu_1^+, \mu_1^-, \mu_2^+, \mu_2^-, \cdots, \mu_n^+, \mu_n^-),
\]

be a \(2n\) dimensional vector and define

\[
\nu^T = (\mu_1^+ - \mu_1^-, \mu_2^+ - \mu_2^-, \cdots, \mu_n^+ - \mu_n^-).
\]

Let \(d\) be the unique global minima of the \(QP(x, \rho)\). Then \(d\) satisfies the KKT conditions.
conditions

\[ H(x)d + \nabla h(x) + \lambda^T A(x) + \nu = 0 \]

\[ \lambda_i(c_i(x) + a_i^T d) = 0, \; i \in I \]

\[ \mu_i^+(d_i - \rho) = 0, \; \mu_i^-(d_i - \rho) = 0, \; i = 1, 2, \ldots, n \]

\[ \lambda_i \geq 0, \; i \in I \]

\[ \mu_i^+ \geq 0, \; \mu_i^- \geq 0, \; i = 1, 2, \ldots, n \]

\[ c_i(x) + a_i^T(x)d \leq 0, \; i \in I \]

\[ c_i(x) + a_i^T(x)d = 0, \; i \in E \]

\[ \|d\|_\infty \leq \rho. \]

(B.3)

If \( d = 0 \) is a solution of \( \text{QP}(x, \rho) \) then from the above KKT system we can say \( \mu = 0 \) because \( \rho > 0 \). At \( d = 0 \) with \( \mu = 0 \) the above KKT system becomes

\[ \nabla h(x) + \lambda^T A(x) = 0 \]

\[ \lambda_i c_i(x) = 0, \; i \in I \]

\[ \lambda_i \geq 0, \; i \in I \]

\[ c_i(x) \leq 0 \; i \in I \]

\[ c_i(x) = 0 \; i \in E. \]

(B.4)

But, the above are KKT conditions of the original optimization problem and this implies that \( x \), when \( d = 0 \), is KKT point of the optimization problem.

REFERENCES


