AN INTERIOR POINT MINIMAX ALGORITHM WITH APPLICATIONS IN ENGINEERING

Stanislav Žaković * **Berc Rustem** *,[1]
Stratos Pistikopoulos **

* *Department of Computing, Imperial College London, London SW7 2AZ*
** *CPSE, Imperial College London, London SW7 2AZ*

Abstract: In this paper an algorithm for the continuous minimax problem with decoupled constraints is considered. The initial problem is transformed to an equivalent equality constrained problem, where the logarithmic barrier function is used to ensure feasibility. Computational results are included which illustrate the efficient performance of the algorithm. *Copyright ©2005 IFAC*

Keywords: Worst–Case Design, Minimax Problems

1. FORMULATION OF THE PROBLEM

Consider the following problem:

$$\begin{aligned} &\min_x \max_{y\in Y} f(x,y), \\ &s.t.\ g(x)=0, \\ &\quad x\geq 0, \end{aligned} \tag{1}$$

where Y is a compact subset of R^m, $x \in R^n$, $f(x,y)$ is continuous in x and y, twice continuously differentiable in x, and $g : R^n \to R^k$ is continuous and twice differentiable in x.

Such problems arise in numerous disciplines, including n–person games (Rosen, 1965), finance (Rustem and Howe, 2001), economics (Zakovic *et al.*, 2002) and others. In general, they are used by the decision maker to assess the worst–case strategy of the opponent and compute the optimal response. The opponent can also be interpreted as nature choosing the worst–case value of the uncertainty, and the solution would be the strategy which ensures the optimal response to the worst–case. Neither the decision maker nor the opponent would benefit by deviating unilaterally from this strategy. The solution of (1) can be characterized as a saddle point when $f(x,\cdot)$ is convex in x and $f(\cdot,y)$ is concave in y.

In this paper we consider an algorithm in which $f(x,y)$ is not necessarily convex in x and concave in y. Previous attempts to solve this problem have mostly focused on unconstrained problems (Kiwiel (Kiwiel, 1987), Demyanov and Pevnyi (Demyanov and Pevnyi, 1972), Rustem and Howe (Rustem and Howe, 2001)), with notable exceptions such as Breton and El Achem who use projected subgradient and bundle methods (Breton and Hachem, 1995). In the present paper we consider an algorithm for solving the constrained problem by extending the approach of Rustem and Howe (Rustem and Howe, 2001). An earlier attempt to solve a similar problem, with decoupled constraints, was made by Sasai (Sasai, 1974), but our approach differs in the choice of the barrier function, the merit function and the interior point search direction. A survey of algorithms for computing saddle points can be found in (Demyanov and Pevnyi, 1972), (Rustem and Howe, 2001).

[1] Corresponding author.

The penalty formulation to ensure feasibility regarding the inequality constraints on the slack variable is realized using a barrier function such as $-\log(x_i)$. The framework for solving the problem is closely related to the interior point literature (e.g. (Akrotirianakis and Rustem, 2005), (El-Bakry *et al.*, 1996)). The transformed min–max problem is given by:

$$\begin{aligned} \min_x \max_{y \in Y} \; & f(x,y) - \mu \sum_{i=1}^{n} \log(x_i), \\ s.t. \; & g(x) = 0, \\ & x \geq 0. \end{aligned} \tag{2}$$

We define the merit function $\Psi(x; c, \mu)$ as:

$$\Psi(x; c, \mu) = \max_{y \in Y(x_k)} \psi(x, y; c, \mu) \tag{3}$$

$$= \max_{y \in Y} \{ f(x,y) + \frac{c}{2} \|g(x)\|_2^2 - \mu \sum_{i=1}^{n} \log(x_i) \}.$$

Now, (1) can be approximated with:

$$\begin{aligned} \min_x \; & \Psi(x; c, \mu) \\ s.t. \; & g(x) = 0. \end{aligned} \tag{4}$$

The algorithm discussed below solves problem (1) and is based on a sequence of optimization problems characterized by a penalty $c \geq 0$ and a barrier $\mu \geq 0$ parameters. To establish first order necessary conditions for the solution of (1), we define the following set of maximizers at x:

$$Y(x) = \{ \quad y \quad | \quad y = \arg\max_{\eta \in Y} \psi(x, \eta; c, \mu) \}.$$

Given the maximizers $y \in Y(x)$ and using the fact that the maximum over a set of scalars is equal to the maximum over their convex hull, (3) can be equivalently expressed as:

$$\begin{aligned} \min_x \max_{\zeta^y \in B} \; & \sum_{y \in Y(x)} \zeta^y \psi(x, y; c, \mu), \\ s.t. \; & g(x) = 0, \end{aligned}$$

where

$$B = \{ \zeta^y | \sum_y \zeta^y = 1, \quad \zeta^y \geq 0 \}.$$

In this paper, we discuss an interior point algorithm to solve (4). Interior point methods allow for a direction of descent Δw, which can be used to determine the multipliers ζ in terms of the subgradient of $\psi(x, \cdot; c, \mu)$. The main features of the algorithm are the generation of a descent direction based on a subgradient of $\psi(x, \cdot; c, \mu)$, an approximate Hessian, in the presence of possible multiple maximizers of (3) and a stepsize strategy to ensure sufficient progress at each iteration.

2. AUGMENTED LAGRANGIAN AND BASIC ITERATION

The Lagrangian associated with (3) is given by:

$$\begin{aligned} L(x, y, \lambda; c, \mu) = \Phi(x) + \frac{c}{2} \|g(x)\|_2^2 \\ - \mu \sum_{i=1}^{n} \log(x_i) - \lambda^T g(x) \}, \end{aligned}$$

where $\Phi(x) = \max_{y \in Y} f(x, y)$.

The first order optimality condition for the problem (3) and the associated Lagrangian, after invoking the nonlinear transformation $z = \mu X^{-1} e$ (we use X to denote the diagonal matrix with diagonal x and employ analogous notation for other quantities) are the following system of nonlinear equations

$$F = \begin{bmatrix} \nabla_x \Phi(x) - z - \nabla_x g^T(x) \lambda \\ g(x) \\ XZe - \mu e \end{bmatrix} = 0, \tag{5}$$

and $F = F(x, \lambda, z; c, \mu)$, $e \in R^n$ is a vector of all ones.

For μ fixed, system (5) is solved using the quasi–Newton method. Therefore, the first order change to the system needs to be found. The $k-th$ Newton iteration for solving (5) can be written as:

$$\nabla F_k \Delta w_k = -F_k,$$

where $F_k = F(x_k, \lambda_k, z_k; c, \mu)$, $w_k = (x_k, \lambda_k, z_k)$ and $\Delta w_k = (\Delta x_k, \Delta \lambda_k, \Delta z_k)$.

At the $k-th$ iteration the system can also be written as:

$$\begin{bmatrix} H_k & -\nabla_x g_k^T & -I \\ \nabla_x g_k & 0 & 0 \\ Z_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta x_k \\ \Delta \lambda_k \\ \Delta z_k \end{bmatrix} = -F_k, \tag{6}$$

where H_k is a positive definite approximation of the Hessian of the augmented Lagrangian of the equality constrained problem:

$$L_{ec}(x, y, \lambda; c) = \Phi(x) - \lambda^T g(x).$$

The Hessian H_k is computed using the updating formula suggested by Powell (Powell, 1978).

Starting from an initial point w_0, the algorithm generates a sequence $\{w_k\}$

$$w_{k+1} = w_k + \alpha_k \Delta w_k, \quad k = 0, 1, 2, ...$$

In order to keep w_{k+1} feasible, we need to ensure that the new iterates x_{k+1} and z_{k+1} are strictly greater than 0. Different strategies have been proposed starting from (Zhang *et al.*, 1990), where

a single step size for different variables is used, to (Yamashita and Yabe, 1996) where a number of different strategies are suggested. The algorithm presented in this paper uses different step–sizes for the primal and dual variables. Hence, $\alpha_k = (\alpha_{x_k}, \alpha_{z_k}, \alpha_{z_k})$ where α_{x_k} and α_{z_k} are the step–lengths for the primal variables x and the pair of dual variables λ and z respectively.

The algorithm moves from one inner iteration to another (with μ fixed) by minimizing the max of merit function $\Psi(x; c, \mu)$ which is achieved by appropriately selecting the penalty parameter c at each inner iteration. The monotonic decrease of (3) and the rules for determining the primal and dual step–sizes ensure that the inner iterates converge to the solution of (3) for fixed value of μ. By reducing μ and $\{\mu\} \to 0$ the optimum of the initial problem (4) is obtained. Full convergence analysis of the proposed method can be found in (Rustem and Zakovic, 2005).

3. COMPUTING THE GRADIENT OF $\Phi(X)$

The subdifferential of the merit function can be expressed as:

$$\partial\Psi(x; c, \mu) = conv\{\nabla_x \psi(x, y; c, \mu) \quad | \quad y \in Y(x)\}.$$

For non–unique maximizers $y^i \in Y(x), i = 1, ..., n_{max}$ by Caratheodory's theorem (Rockafellar, 1970) a vector $\nabla_x \Psi(x; c, \mu) \in \partial\Psi(x; c, \mu)$ can be characterized by at most $(n + 1)$ vectors $\nabla_x \psi(x, y; c, \mu) \in \partial\Psi(x; c, \mu)$ so that

$$\nabla_x \Psi(x; c, \mu) = \sum_{y \in Y(x)} \zeta^y \nabla_x \psi(x, y; c, \mu)$$
$$= \sum_{y \in Y(x)} \beta^y \nabla_x f(x, y) + c \nabla_x g^T g - \mu X^{-1} e \quad (7)$$

In this paper the quasi–Newton descent direction is based on an approximation of $\Psi(x)$ and y_{k+1} which ensures descent of the merit function. The direction Δx_k is a descent direction for Ψ, at the current point x_k, if

$$\max_{y \in Y(x_k)} (\Delta x_k^T \nabla_x \Psi(x_k, y; c_k, \mu) + \frac{1}{2} \|\Delta x_k\|^2_{H_k} \leq 0. \quad (8)$$

The motivation for this choice is the selection of worst–case descent direction among possible maximizers. Hence, for such y_{k+1} we have:

$$y_{k+1} = \arg\max_{y \in Y} \{\Psi_k(\Delta x_k)\} \quad (9)$$
$$= \arg \max_{y \in Y(x_k)} \{\Delta x_k^T (\nabla f(x_k, y) - \mu X_k^{-1} e)$$
$$- c_k \|g_k\|_2^2 + \frac{1}{2} \|\Delta x_k\|^2_{H_k}\}.$$

When there is more than one maximizer, then the new direction $\Delta w(\beta)$ is computed:

$$\Delta w(\beta) = -\nabla F_k^{-1} F_k^{\beta},$$

where F_k^{β} is the following non–linear system:

$$\sum_{y \in Y(x_k)} \beta_k^y \nabla_x f(x_k, y) - \nabla_x g_k^T \lambda$$
$$g_k$$
$$X_k Z_k e - \mu e$$

and

$$\beta_k^y = \arg\max_{\beta \in B} \{\sum_{\beta} \beta \Psi_k(\Delta x_k)\}$$
$$= \arg\max_{\beta \in B} \{\Delta x_k^T (\sum_{y \in Y(x_k)} \beta \nabla f(x_k, y) - \mu X_k^{-1} e)$$
$$+ \frac{1}{2} \|\Delta x_k\|^2_{H_k}\}, \quad (10)$$

because $g_k(x)$ is not a function of β_k^y.

When the unique maximizer y_{k+1} is chosen, β can be set to 1, corresponding to the given y_{k+1} so that

$$F_k^{\beta=1} = \begin{bmatrix} \nabla_x f(x_k, y_{k+1}) - z_k - \nabla_x g_k^T \lambda \\ g_k \\ X_k Z_k e - \mu e \end{bmatrix}.$$

Therefore, we define two possible directions $\Delta w(y_{k+1})$ and $\Delta w(\beta)$, depending on the gradient $\nabla_x \Phi(x)$ used. The direction $\Delta w(y_{k+1})$ is easier to compute, as it does not entail the solution of the quadratic programming problem (10). The two possible derivatives of $\nabla_x \Phi(x_k)$, together with the corresponding direction Δw_k are shown in Table 1 below.

Table 1. Two choices for $\nabla_x \Phi(x_k)$.

$\nabla_x \Phi(x_k)$	Δw_k
$\nabla_x f(x_k, y_{k+1})$	$\Delta w_k(y_{k+1}) = -\nabla F_k^{-1} F_k^{\beta=1}$
$(\beta^y)^T \nabla_x f(x_k, y)$	$\Delta w_k(\beta) = -\nabla F_k^{-1} F_k^{\beta}$

Let $\epsilon_g > 0$ be a finite precision to which the equality constraints are satisfied. Thus, for merit function (3), we have a worst case feasibility precision

$$\|g(x)\|^2 > \epsilon_g. \quad (11)$$

It can be shown that descent is always guaranteed if (11) holds or $g(x) = 0$ and the penalty parameter $c_k = c_k(\epsilon_g)$ remains finite.

Assume that at, some inner iteration k, $0 < \|g(x_k)\|^2 \leq \epsilon_g$ and the descent condition (8) is not satisfied. In such circumstances a switch to a different merit function:

$$\|F(x, y, \lambda, z; c, \mu)\|_2^2 \tag{12}$$

is made for all consecutive inner iterations. Once the convergence of the inner iteration is achieved, the algorithm returns to minimizing merit function (3). This is a variation of the so called "watchdog" technique, which was first suggested by Chamberlain *et al.* in (Chamberlain *et al.*, 1982). In the context of interior point methods it was also used by Gay *et al.* in (Gay *et al.*, 1997). The convergence criteria for (12) has been well established (El-Bakry *et al.*, 1996), (Zakovic *et al.*, 2000).

Interior Point Minimax Algorithm

(1) choose x^0, y^0, λ^0, z^0, $\mu^0 > 0$;
(2) set $l = 0, k = 0$;
(3) **WHILE** $\|F(x^l, y^l, \lambda^l, z^l; c, \mu)\|_2/(1+\|x_l, y_l, \lambda_l, z_l\|_2^2) \geq \epsilon_0$
(4) $(x_k, \lambda_k, z_k) = (x^l, \lambda^l, z^l)$
(5) **WHILE** $\|F(x_k, y_k, \lambda_k, z_k; c, \mu)\|_2 \geq \eta\mu$
(6) compute Δw_k
(7) compute $\alpha_k = (\alpha_{x_k}, \alpha_{z_k}, \alpha_{z_k})$
(8) $\omega_{k+1} = \omega_k + \alpha_k \Delta\omega_k$
(9) $k = k + 1$
(10) **END WHILE**
(11) compute μ^{l+1}
(12) $(x^{l+1}, y^{l+1}, \lambda^{l+1}, z^{l+1}) = (x_k, y_k, \lambda_k, z_k)$
(13) $l = l + 1$
(14) **END WHILE**

4. NUMERICAL RESULTS AND APPLICATIONS

The algorithm defined in this paper has been implemented and the experience with numerical examples has been reported in (Rustem and Zakovic, 2005). A number of examples, both with convex–concave functions (i.e. $f(x, y)$ convex in x and concave in y) and convex–convex functions, where multiple maximizers are present, and we examine the behavior of the algorithm in presence of such points. A comparison with the semi–infinite algorithm has been presented in (Zakovic and Rustem, 2003).

4.1 Robust Optimal Design of Dynamic Experiments

To obtain predictive models of processing systems, we are faced with the problem of having to estimate several freely-varying parameters within the model from collected experimental data in order to validate the model. Due to the nonlinearity of these models, the ease at which this can be accomplished not only depends on the parameterization of the model (*i.e.*, the mathematical form of the model), but also on "where" in the experiment space the data have been collected. The main questions that arise when designing optimally informative experiments are – how should we adjust time–varying controls, initial conditions or the duration of the experiment to generate the maximum amount of information for parameter identification?

The predicted amount of information contained within a set of experimental data can be used to design future experiments that are optimally rich in information for parameter estimation purposes. To quantify this, an information matrix for dynamic experiment design is defined as:

$$M_I(\theta, \phi) \equiv \sum_{r=1}^{M} \sum_{s=1}^{M} \tilde{\sigma}_1^{rs} Q_r^T Q_s, \tag{13}$$

where ϕ is a vector of experiment decision variables (*i.e.*, sampling times of response variables to be predicted by the model, time–varying controls to be applied to the process, etc.) and θ is the vector of model parameters to be estimated from data. The $(n_{sp} \times p)$ matrix Q_r is the matrix of first–order dynamic sensitivity coefficients of the r^{th} response variable in the model computed at each of n_{sp} sampling points (the number of which is chosen *a priori*):

$$Q_r \equiv \begin{bmatrix} \frac{\partial y_r(\theta, \phi, t_1)}{\partial \theta_1} & \cdots & \frac{\partial y_r(\theta, \phi, t_1)}{\partial \theta_p} \\ \cdots & \cdots & \cdots \\ \frac{\partial y_r(\theta, \phi, t_{n_{sp}})}{\partial \theta_1} & \cdots & \frac{\partial y_r(\theta, \phi, t_{n_{sp}})}{\partial \theta_p} \end{bmatrix} \tag{14}$$

To design future experiments in the face of uncertainty in the parameters, θ, we solve the following max-min optimization problem:

$$\phi_R = \arg\max_{\phi \in \Phi} \{\min_{\theta \in \Theta} \{\det(M_I(\theta, \phi))\}\}, \tag{15}$$

where, $\Phi \subset \mathcal{R}^n$ and $\Theta \subset \mathcal{R}^m$ represent feasible regions (upper and lower bounds on each element of ϕ and θ, respectively), and thus there are no functional constraints on the variables.

A typical example is presented to illustrate the application of the robust optimal experiment design. Consider a fed-batch reactor in which the fermentation of baker's yeast is carried out. To model this process the following model is proposed:

$\underline{M_I(\theta, \mathbf{u}, t)}$

$$\begin{aligned} \frac{dy_1}{dt} &= (r - u_1 - \theta_4) y_1 \\ \frac{dy_2}{dt} &= -\frac{r y_1}{\theta_3} + u_1(u_2 - y_2) \end{aligned} \tag{16}$$

$$r = \frac{\theta_1 y_2}{\theta_2 + y_2}.$$

Within this process, there are two time-varying controls (u_1, u_2) and two measured concentrations (y_1, y_2).

Here, we have $\Phi \subset \mathcal{R}^{29}$ – one varying initial condition, ten sampling times, and two time–varying inputs parameterized by piecewise constant trajectories, each delineated by five time intervals (representing 18 variables); and $\Theta \subset \mathcal{R}^4$ as there are four model parameters to be estimated from experimental data. In table 2 we show the solution to (15) for the model describing the semi-batch fermentation of baker's yeast (Asprey and Macchietto, 2000), using two algorithms - the min–max algorithm (**MMX**)presented in this paper and the semi–infinite algorithm (**SIP**) presented in (Zakovic and Rustem, 2003).As can be seen, algorithm **SIP** required two more iterations than the algorithm **MMX** to converge. However, solving the *max* problem is much more expensive, reflected in the CPU times needed to solve the problem with the two different algorithms.

Table 2. Iterations of **MMX** and **SIP** when solving problem (15).

		MMX	**SIP**
iteration		results	
I	**max**	39.38954	39.38954
II	**max**	4.57472	11.362926
III	**max**	4.48416	11.325922
IV	**max**	4.47273	10.034939
V	**max**	–	10.033279
VI	**max**	–	4.4723
time		7622.21s	13318.36s

5. CONCLUSIONS

In this paper, we have extended the algorithm for unconstrained continuous minimax problems presented in Rustem & Howe (Rustem and Howe, 2001) to include constraints on the minimizing variable x. The interior point approach is used to ensure feasibility of the variables. The algorithm uses quasi–Newton search direction, conditional on approximate maximizers.

To illustrate the applicability of the algorithm, a problem from engineering was presented, namely an example of robust optimal design of dynamic experiments. The problem was solved successfully, and the complexity of the optimal design example was illustrated in the computational time used to solve it – approximately two hours.

REFERENCES

Akrotirianakis, I. and B. Rustem (2005). A globally convergent interior point algorithm for general non–linear programming problems. *J. of Opt.Theory and Appl.*

Asprey, S.P. and S. Macchietto (2000). *Comp. Chem. Eng.* pp. 1261–1267.

Breton, M. and S. El Hachem (1995). Algorithms for the solution of stochastic dynamic minimax problems. *Comp. Opt. and Appl.* **4**, 317–345.

Chamberlain, R.M., C. Lemarechal, H.C. Pedersen and M.J.D. Powell (1982). The watchdog technique for forcing convergence in algorithms for constrained optimization. *Mathematical Programming Study* pp. 1–17.

Demyanov, V.F. and A.B. Pevnyi (1972). Numerical methods for finding saddle points. *USSR Comp. Math. and Math. Phys.* **12**, 1099–1127.

El-Bakry, A.S., R.A. Tapia, T. Tsuchiya and Y. Zhang (1996). On the formulation and theory of the newton interior point method for nonlinear programming. *J. of Opt.Theory and Appl.* **89**, 507–541.

Gay, D.M., M.L. Overon and M.H. Wright (1997). A primal–dual interior method for nonconvex nonlinear programming. *Technical Report 97–4–08, Bell Laboratories.*

Kiwiel, K.C. (1987). A direct method of linearization for continuous minimax problems. *J. of Opt.Theory and Appl.* **55**, 271–287.

Powell, M.J.D. (1978). A fast algorithm for nonlinearly constrained optimization calculations. In: *Numerical Analysis Proceedings, Biennial Conference, Dundee* (G.A. Watson, Ed.). pp. 144–157. Springer Verlag. Berlin.

Rockafellar, R.T. (1970). *Convex Analysis.* Princeton University Press. Princeton, NJ.

Rosen, J.B. (1965). Existence and uniqueness of equilibrium points for concave n–person games. *Econometrica* **33**, 520–534.

Rustem, B. and M.A. Howe (2001). *Algorithms for Worst-case Design with Applications to Risk Management.* Princeton University Press.

Rustem, B. and S. Zakovic (2005). An interior point algorithm for continuous minimax problems. *J. of Opt.Theory and Appl.* submitted.

Sasai, H. (1974). An interior penalty method for minimax problems with constraints'. *SIAM J. Control* **12**, 643–649.

Yamashita, H. and H. Yabe (1996). Superlinear and quadratic convergence of some primal–dual interior point methods for constrained optimization. *Mathematical Programming* **75**, 377–397.

Zakovic, S. and B. Rustem (2003). Semi–infinite programming and applications to minimax problems. *Annals of OR* **124**, 81–110.

Zakovic, S., B. Rustem and C.C. Pantelides (2000). An interior point algorithm for com-

puting saddle points of constrained continuous minimax. *Annals of Operations Research* pp. 59–78.

Zakovic, S., B. Rustem and V. Wieland (2002). A continuous minimax problem and its application to inflation targeting. In: *Decision and Control in Management Science* (G. Zaccour, Ed.). Kluwer Academic Publishers.

Zhang, Y., R.A. Tapia and F. Potra (1990). On the superlinear convergence of interior point algorithms for a general class of problems. *Technical Report TR90–9, Department of Mathematical Sciences, Rice University, USA*.