Adaptive Optimization with Constraints: Convergence and Oscillatory Behaviour

Fernando J. Coito[1] and João M. Lemos[2]

[1] Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa
fjvc@fct.unl.pt
[2] INESC ID-Lisboa
jlml@inesc-id.pt

Abstract. The problem of adaptive minimization of globally unknown functionals under constraints on the independent variable is considered in a stochastic framework. The CAM algorithm for vector problems is proposed. By resorting to the ODE analysis for analysing stochastic algorithms and singular perturbation methods, it is shown that the only possible convergence points are the constrained local minima. Simulation results in 2 dimensions illustrate this result.

1 Introduction

There are engineering optimization problems in which the global form of both the cost functional and the constraints are unknown. In these problems, when the independent variable is settled to a specific value, the corresponding value of the functional can be read and the decision whether the constraints are or are not being violated can be made. The solution amounts to a number of values applied to the plant according to a functional and constraint models, which are adapted from incoming data. Although these extremum seeking methods have already been the subject of early literature in Adaptive Systems – see [3] for a review – they are receiving increasing interest in recent literature.

This kind of problems are solved in [1,6] by using a self-tuning extremum seeker in which the cost functional is locally approximated by a quadratic function and no constraints are assumed in the independent variable. In this work, the above algorithm is extended for incorporating constraints and the use of vector independent variable.

2 Problem Formulation

Let $y(\cdot)$ be a differentiable function of $\Re^2$ in $\Re$. Consider the following problem

Problem 1 Find $\mathbf{x}^* = \begin{bmatrix} x_1^* & x_2^* \end{bmatrix}^T$ such that $y\left(\mathbf{x}^*\right)$ is minimum, subject to the set of constraints $\mathbf{g}\left(\mathbf{x}^*\right) \leq \mathbf{0}$ where $\mathbf{g} \in \Re^n$ and $\mathbf{0}$ is the null vector. □

According to the Kuhn-Tucker theorem, Problem 2 is equivalent to the following

J.S. Marques et al. (Eds.): IbPRIA 2005, LNCS 3523, pp. 19–26, 2005.

Problem 2 Define the Lagrangean function $\pounds(\mathbf{x},\boldsymbol{\rho}) \stackrel{\Delta}{=} y(\mathbf{x}) + \boldsymbol{\rho}^T \mathbf{g}(\mathbf{x})$.

Find the $\mathbf{x}^*$ minimizing $\pounds(\mathbf{x},\boldsymbol{\rho}^*)$, in which $\boldsymbol{\rho}^*$ is a vector of Lagrange multipliers, satisfying the Kuhn-Tucker complementary condition where $\stackrel{\bullet}{\times}$ is the *term-by-term* multiplication:

$$\boldsymbol{\rho}^* \stackrel{\bullet}{\times} \mathbf{g}(\mathbf{x}^*) = \mathbf{0} \tag{1}$$ □

Hereafter, the following assumption is supposed to hold:

H0. The global form of functions $y(\cdot)$ and $\mathbf{g}(\cdot)$ is unknown and may be possibly time varying. However, for each $\mathbf{x}$, $y(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ may be observed, possibly corrupted by observation noise. □

3 The CAM algorithm

The algorithm that solves Problem 2 must accomplish two tasks: the adjustment of the Lagrange multipliers $\boldsymbol{\rho}$ in order to fulfill the Kuhn-Tucker complementary condition (1) and, once $\boldsymbol{\rho}$ is settled, to adjust $\mathbf{x}(t)$.

3.1 Adjustment of the Lagrange multiplier

Following the development in [4] $\boldsymbol{\rho}$ is adjusted according to a gradient minimization scheme:

$$\boldsymbol{\rho}(t) = \boldsymbol{\rho}(t-1) + \varepsilon\gamma(t-1)\boldsymbol{\rho}(t-1) \stackrel{\bullet}{\times} \mathbf{g}(\mathbf{x}(t)) \tag{2}$$

where ε is a vanishing small parameter and $\{\gamma(t)\}$ is a sequence of positive gains [5].

3.2 Adaptive optimization

H1. It is assumed that, close to $\mathbf{x}^*$, the Lagrangean function $\pounds(\mathbf{x},\boldsymbol{\rho}^*)$ may be approximated by a quadratic function:

$$L(t) \stackrel{\Delta}{=} \pounds(\mathbf{x}(t),\boldsymbol{\rho}) = \pounds^* + \left[\mathbf{x}(t) - \mathbf{x}^*\right]^T \mathbf{A}\left[\mathbf{x}(t) - \mathbf{x}^*\right] + \bar{e}(t) \tag{3}$$

in the sequel it will be assumed $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$ to be symmetric, which does not affect the problem generality. $\mathbf{A}$, $\pounds^*$ and $\mathbf{x}^*$ are unknown parameters, which depend on the value of $\boldsymbol{\rho}$; $\bar{e}$ is a residue.

Define the increments:

$$\begin{aligned} &\Delta L(t) \stackrel{\Delta}{=} L(t) - L(t-1) && \Delta x_i(t) \stackrel{\Delta}{=} x_i(t) - x_i(t-1) \quad ; \quad i = 1,2 \\ &\Delta x_i^2(t) \stackrel{\Delta}{=} x_i^2(t) - x_i^2(t-1) \quad ; \quad i = 1,2 && \Delta[x_1 x_2](t) \stackrel{\Delta}{=} x_1(t)\cdot x_2(t) - x_1(t-1)\cdot x_2(t-1) \end{aligned} \tag{4}$$

Then equation (3) may be written as

$$\Delta L(t) = \begin{bmatrix}\theta_1 & \cdots & \theta_5\end{bmatrix}\begin{bmatrix}\Delta x_1(t) & \Delta x_2(t) & \Delta x_1^2(t) & \Delta x_2^2(t) & \Delta[x_1 x_2](t)\end{bmatrix}^T + e(t) \tag{5}$$

where

$$\begin{aligned} &\theta_1 = -2a_{11}x_1^* - 2a_{12}x_2^* && \theta_3 = a_{11} && \theta_3 = 2a_{12} \\ &\theta_2 = -2a_{22}x_2^* - 2a_{12}x_1^* && \theta_4 = a_{22} \end{aligned}$$

and $e(t) \stackrel{\Delta}{=} \bar{e}(t) - \bar{e}(t-1)$ is assumed to be an uncorrelated zero mean stochastic sequence such that all moments exist.

Defining $\boldsymbol{\theta}^* \stackrel{\Delta}{=} [\theta_1 \quad \cdots \quad \theta_5]$ and $\boldsymbol{\varphi}(t) \stackrel{\Delta}{=} [\Delta x_1(t) \quad \Delta x_2(t) \quad \Delta x_1^2(t) \quad \Delta x_2^2(t) \quad \Delta[x_1 x_2](t)]^T$ expression (5) yields

$$\Delta L(t) = \boldsymbol{\theta}^* \boldsymbol{\varphi}(t) + e(t) \tag{6}$$

which constitutes a linear regression model in which $\boldsymbol{\theta}^*$ is the vector of coefficients to estimate and $\boldsymbol{\varphi}$ is the data vector.

The vector $\boldsymbol{\theta}^*$ may be estimated using a recursive least-squares algorithm, and the value of **x** that minimizes $L(\mathbf{x})$ is given by:

$$\begin{bmatrix} x_1^* & x_2^* \end{bmatrix}^T = \begin{bmatrix} \dfrac{2\theta_1\theta_4 - \theta_2\theta_5}{\theta_5^2 - 4\theta_3\theta_4} & \dfrac{2\theta_2\theta_3 - \theta_1\theta_5}{\theta_5^2 - 4\theta_3\theta_4} \end{bmatrix}^T \tag{7}$$

3.3 The CAM algorithm

Combining both the above procedures results in the following *Constrained Adaptive Minimization* (CAM) algorithm:

1. Apply $\mathbf{x}(t)$ to the system and measure $y(t)$ and $\mathbf{g}(\mathbf{x}(t))$
2. Adjust the Lagrange multiplier vector according to equation (2).
3. Build the Lagrangean function associated with the current Lagrange multiplier vector and the current value $y(t)$.
4. Compute the increments (4).
5. Using a RLS algorithm update the estimates of $\boldsymbol{\theta}$ in the model (6).
6. Update the estimates according to

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix}^T = \begin{bmatrix} \dfrac{2\theta_1\theta_4 - \theta_2\theta_5}{\theta_5^2 - 4\theta_3\theta_4} & \dfrac{2\theta_2\theta_3 - \theta_1\theta_5}{\theta_5^2 - 4\theta_3\theta_4} \end{bmatrix}^T + \boldsymbol{\eta}(t) \tag{8}$$

7. Increment the time and go back to step 1.

4 ODE analysis

The CAM algorithm is now analyzed using the ODE method for analyzing stochastic algorithms [5] and singular perturbation theory for ordinary differential equations [2].

The algorithm is associated with the following set of differential equations:

$$\frac{d\,\boldsymbol{\rho}(t)}{d\,t} = \varepsilon \boldsymbol{\rho}(t) \overset{\bullet}{\times} \mathbf{g}(\mathbf{x}(t)) \qquad \frac{d\,\boldsymbol{\theta}(t)}{d\,t} = R^{-1}\mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\rho}) \tag{9}$$

where $R \stackrel{\Delta}{=} E[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t)]$ and $\mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\rho}) \stackrel{\Delta}{=} E\{\boldsymbol{\varphi}(t)[\Delta L(t) - \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}]\}$.

Define the functions $\mathbf{G}(\boldsymbol{\theta}, \boldsymbol{\rho})$ and $\mathbf{H}(\boldsymbol{\theta}, \boldsymbol{\rho})$

$$\mathbf{G}(\boldsymbol{\theta}, \boldsymbol{\rho}) \stackrel{\Delta}{=} R^{-1}\mathbf{f}(\boldsymbol{\theta}, \boldsymbol{\rho}) \qquad \mathbf{H}(\boldsymbol{\theta}, \boldsymbol{\rho}) \stackrel{\Delta}{=} \boldsymbol{\rho} \overset{\bullet}{\times} \mathbf{g}(\mathbf{x}(t)) \tag{10}$$

Making use of (10) and changing the time scale by $\tau = \varepsilon t$, equations (9) may then be written in the standard form for singular perturbation analysis:

$$\frac{d\,\boldsymbol{\rho}(\tau)}{d\,\tau} = \mathbf{H}(\boldsymbol{\theta}, \boldsymbol{\rho}) \qquad \varepsilon \frac{d\,\boldsymbol{\theta}(\tau)}{d\,\tau} = \mathbf{G}(\boldsymbol{\theta}, \boldsymbol{\rho}) \tag{11}$$

According to the ODE theory exposed in [5], the only possible convergence points of the CAM algorithm are the equilibrium points of (11), such that the Jacobian matrix has all its eigenvalues in the left complex half-plane:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{H}}{\partial \boldsymbol{\rho}} & \frac{\partial \mathbf{H}}{\partial \boldsymbol{\theta}} \\ \frac{\partial \mathbf{G}}{\partial \boldsymbol{\rho}} & \frac{\partial \mathbf{G}}{\partial \boldsymbol{\theta}} \end{bmatrix} \quad (12)$$

H2. The disturbance signal $\boldsymbol{\eta}$ in (8) ensures the persistent excitation requirement, i.e. $E\left[\boldsymbol{\varphi}(\tau)\boldsymbol{\varphi}^T(\tau)\right]$ is full rank. □

H3. The function $\mathbf{G}\left(\bar{\boldsymbol{\theta}}, \boldsymbol{\rho}^*\right)$ has isolated real roots □

The equilibrium points of (11) are characterized by one of the following conditions:

A-equilibria

$$\boldsymbol{\rho} = \mathbf{0} \qquad f(\boldsymbol{\theta}, \mathbf{0}) = \mathbf{0} \quad (13)$$

B-equilibria

$$\mathbf{g}(\mathbf{x}) = \mathbf{0} \quad \text{and thus } \boldsymbol{\rho} = \boldsymbol{\rho}^* \qquad f\left(\boldsymbol{\theta}, \boldsymbol{\rho}^*\right) = \mathbf{0} \quad (14)$$

4.1 Analysis of the A-equilibria

If (13) holds the constrained minimum equals the unconstrained minimum. The constrained minimum is therefore interior to the region defined by the set of constraints.

If the persistent excitation requirement holds, as $\frac{\partial \mathbf{H}}{\partial \boldsymbol{\theta}} = \mathbf{0}$, the Jacobian matrix (12) becomes lower triangular and its eigenvalues are the ones of $\frac{\partial \mathbf{H}}{\partial \boldsymbol{\rho}} = [\mathbf{g}(\mathbf{x})]_D$ and $\frac{\partial \mathbf{G}}{\partial \boldsymbol{\theta}} = -\mathbf{I}$, where $\mathbf{I}$ is the diagonal unit matrix and $[\mathbf{g}(\mathbf{x})]_D$ is a diagonal matrix whose elements are the $g_i(\mathbf{x})$. As ρ_i=0 which implies $g_i(\mathbf{x})$<0, all the Jacobian eigenvalues have negative real parts. Thus the only possible convergence points are solutions of Problem 1.

4.2 Analysis of the B-equilibria

If (14) holds the constrained minimum is different from the unconstrained minimum, being located on the boundary of the region defined by $\mathbf{g}(\mathbf{x}) \le \mathbf{0}$. In this case $\frac{\partial \mathbf{H}}{\partial \boldsymbol{\theta}}$ is no longer null. Thus, the Jacobian matrix is not lower triangular, and the analysis from the previous section does not hold.

Making use of the singular perturbation theory (Kokotovic, *et al.*, 1986), assuming that the parameter ε in (2) is vanishing small the two equations in (11) may be seen as the slow and fast subsystems, respectively.

Assume that H3 holds and consider the boundary layer correction $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \bar{\boldsymbol{\theta}}$ whose dynamics is

$$\frac{d\tilde{\boldsymbol{\theta}}}{d\tau} = \frac{1}{\varepsilon}\mathbf{G}\left(\bar{\boldsymbol{\theta}}, \boldsymbol{\rho}^*\right) \quad (15)$$

H4. Assume that $\tilde{\boldsymbol{\theta}}(\tau)=\mathbf{0}$ is an equilibrium point of (15), asymptotically stable, uniformly in $\boldsymbol{\rho}^*$, and that $\boldsymbol{\theta}(0)-\bar{\boldsymbol{\theta}}(0)$ belongs to its domain of attraction. □

Proof of H4: It follows from $\frac{d\tilde{\boldsymbol{\theta}}}{d\tau}=\frac{1}{\varepsilon}R^{-1}E\{\boldsymbol{\varphi}(t)[\boldsymbol{\varphi}^T(t)(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta})+e(t)]\}=-\frac{1}{\varepsilon}R^{-1}R\tilde{\boldsymbol{\theta}}=-\frac{1}{\varepsilon}\tilde{\boldsymbol{\theta}}$ □

H5. The eigenvalues of $\frac{\partial \mathbf{G}}{\partial \boldsymbol{\theta}}$, calculated for ε=0, have strictly negative real part. □

Proof of H5: It results from $\frac{\partial \mathbf{G}}{\partial \boldsymbol{\theta}}=-\frac{\partial}{\partial \boldsymbol{\theta}}R^{-1}E\{\boldsymbol{\varphi}(\tau)\boldsymbol{\varphi}^T(\tau)\}\tilde{\boldsymbol{\theta}}=-\frac{\partial}{\partial \boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}=-\mathbf{I}$ □

Since these assumptions hold, Tikhonov's theorem (Kokotovic, *et al.*, 1986) allows to conclude that the only piossible convergence points of the CAM algorithm are the constrained minima of the optimization problem 1.

5 Simulation results

The ODE analysis characterizes the possible convergence points of the CAM algorithm. Yet, it does not prove that the algorithm will actually converge. In order to exhibit the algorithm convergence features, a number of simulations are presented.

5.1 Example 1

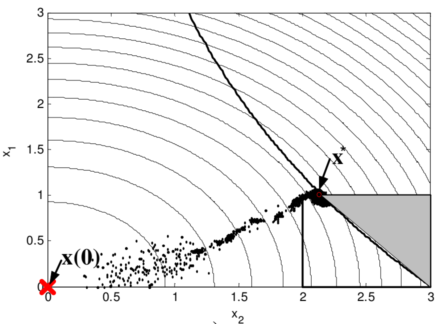
In this example Problem 1 is considered, in which

$$y(\mathbf{x})=(\mathbf{x}-\mathbf{x_0})^T(\mathbf{x}-\mathbf{x_0}) \qquad \text{where } \mathbf{x_0}=[0.6 \quad 0.8]^T \tag{16}$$

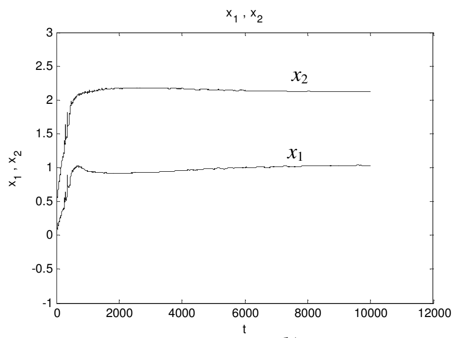
$$\begin{array}{lll} g_1(\mathbf{x})=3-x_2e^{x_1/3}\le 0 & g_3(\mathbf{x})=x_1-1 \le 0 & g_5(\mathbf{x})=x_2-3 \le 0 \\ g_2(\mathbf{x})=-x_1 \le 0 & g_4(\mathbf{x})=2-x_2 \le 0 & \end{array} \tag{17}$$

The identification is performed using RLS with exponential forgetting factor.

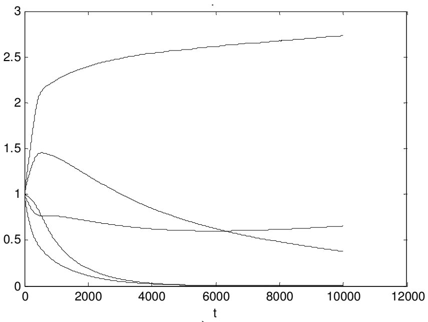
Figure 1 presents the evolution of the optimum estimate towards the feasibility region. The constrained minimum is on the frontier of the region. Thus while the Lagrange multipliers related to the inactive constraints go to zero $(\rho_i \to 0)$, those related to active constraints converge to the optimum $\rho_j \to \rho^*$ (figure 1.*c*).



a)



b)



c)

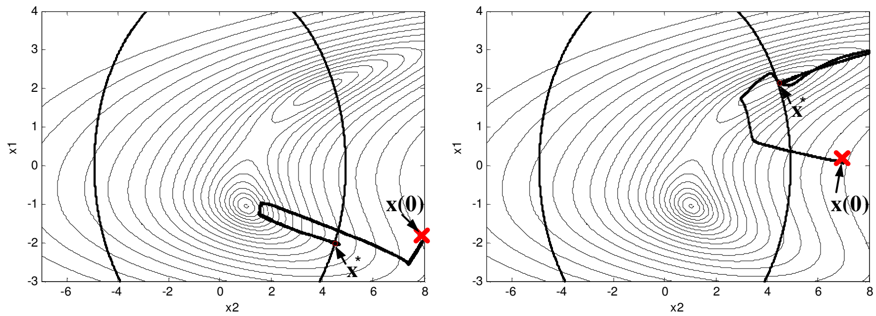
Fig. 1. Adaptive optimum search from example 1. *a*) The gray area is the feasibility region. *b*) x_1 and x_2 time evolution. *c*) Evolution of the Lagrange multipliers

5.2 Example 2: multiple local minima

The ODE analysis presented states that the convergence points are local minima from the constrained optimization problem. Thus, it is interesting to see what occurs when more than one minimum exists. In this example the function to be minimized is given by $y(\mathbf{x}) = 9 + \frac{9}{2}x_1 - 4x_2 + x_1^2 + 2x_2^2 - 2x_1x_2 + x_1^4 - 2x_1^2x_2$ and it is subject to the constraint $g(\mathbf{x}) = 24.25 - \left(x_1^2 + x_2^2\right) \leq 0$.

Experiments using the updating scheme from equation (8) have shown that with this scheme assumption H1 and equation (6) would not hold. Thus in the experiment presented the updating scheme (8) was replaced by the following gradient scheme:

$$\mathbf{x}(t+1) = \mathbf{x}(t) - \delta \frac{\partial L}{\partial \mathbf{x}} \Big/ \left\| \frac{\partial L}{\partial \mathbf{x}} \right\| \qquad (18)$$



a) *b*)

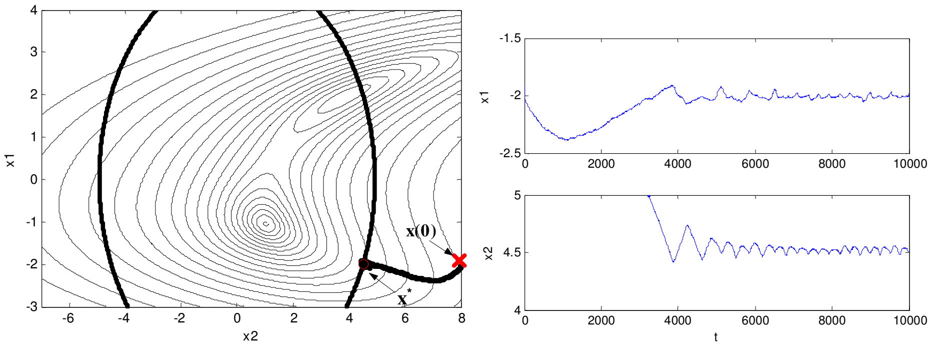
Fig. 2. Adaptive optimum search for Example 2. The feasibility region lies outside the bold line. *a*) $\mathbf{x}(0)$=[-1.9 7.95]^T *b*) $\mathbf{x}(0)$=[0.198 6.95]^T.

Figure 2.*a* presents the algorithm evolution when it starts from the initial point $\mathbf{x}(0)$=[-1.9 7.95]^T. It converges towards a local minimum, located at $\mathbf{x}^* = [-2 \quad 4.51]^T$, with a value of the objective function of 19.6.

In figure 2.*b* the algorithm is started from a different initial point, $\mathbf{x}(0)$=[0.198 6.95]^T. In this case it converges to another local minimum located at $\mathbf{x}^*$=[2.14 4.51]^T, which corresponds to a value of the objective function of 1.21 (the absolute constrained minimum).

The minimum to which the algorithm converges depends on the initial point $\mathbf{x}(0)$, and in which domain of attraction it lies.

5.3 Example 3: improved performance



a) *b*)

Fig. 3. Adaptive optimum search for Example 3. *a*) The transient does not violate the feasibility region. *b*) A detail of the algorithm steady state.

In example 2, the adaptation presents a strong transient that strongly violates the feasibility region. This results from the dynamics (2) of the Lagrange multiplier ρ. In order to improve the algorithm performance the gain $\gamma(t)$ is computed according to the following scheme:

$$\gamma(t) = \text{sat}\left\{\min, \max, \frac{\left[\lambda_\gamma \gamma(t-1) + k_\gamma \, \mathrm{g}(\mathbf{x}(t))\right]}{\text{sat}\{f_\gamma, \infty, \rho(t)\}}\right\} \tag{19}$$

where sat{*min*, *max*, z} corresponds to the function that saturates z between the values *min* and *max*.

Results of the algorithm performance with this modification are presented in figure 3. The adaptation converges towards a local minimum and does not strongly violate the feasibility region.

The values of λ_γ and f_γ are chosen so that $\gamma(t)$ changes rapidly between its *max* and *min* values $(\lambda_\gamma / f_\gamma > 1)$. In the example λ_γ=0.95 and f_γ=0.1 were used. Figure 3.*b* shows the steady state behaviour. The algorithm doesn't actually converge to a value. Instead it oscillates between the two sides of the constraint border.

6 Conclusion

The problem of adaptive minimization of globally unknown functionals under constraints on the independent variable was addressed in a stochastic framework. The CAM algorithm for vector problems was proposed. By resorting to the ODE analysis for analysing stochastic algorithms and singular perturbation methods, it was shown that the only possible convergence points are the constrained local minima. A number of simulation results in 2 dimensions were presented to illustrate this result. Modifications to the original algorithm were introduced to improve performance.

REFERENCES

[1] Bozin, A. S. and M. B. Zarrop (1991). Self-tuning extremum optimizer – Convergence and robustness properties". *Prep. ECC 91*, 672-677.

[2] Kokotovic, P., H. K. Khalil and J. O'Reilly (1986). *Singular Perturbation Methods: Analysis and Design*. Academic Press. 1986.

[3] Krstič, M. (2000) Performance improvement and limitations in extremum seeking control. *Systems & Control Letters*, **39**, 313-326.

[4] Lemos, J. M (1992). Adaptive Optimization with Constraints. *Prep. 4th IFAC Symp. Adaptive Systems in Control and Signal Processeing*, 53-58.

[5] Ljung, L. (1977). Analysis of Recursive Stochastic Algorithms. *IEEE Transactions on Automatic Control*, **AC-22**, Nº4, 551-575.

[6] Wellstead, P. E. and P. G. Scotson (1990). Selftuning extremum control. *IEE Proceedings*, **137**, Pt D. Nº 3, 165-175.