Optim Lett
DOI 10.1007/s11590-008-0090-9

ORIGINAL PAPER

A smoothing algorithm for finite min–max–min problems

Angelos Tsoukalas · Panos Parpas · Berç Rustem

Received: 19 February 2008 / Accepted: 27 May 2008

Abstract We generalize a smoothing algorithm for finite min–max to finite min–max–min problems. We apply a smoothing technique twice, once to eliminate the inner min operator and once to eliminate the max operator. In mini–max problems, where only the max operator is eliminated, the approximation function is decreasing with respect to the smoothing parameter. Such a property is convenient to establish algorithm convergence, but it does not hold when both operators are eliminated. To maintain the desired property, an additional term is added to the approximation. We establish convergence of a steepest descent algorithm and provide a numerical example.

1 Introduction

We propose an algorithm for the finite min–max–min problem

$$\min_{x\in\mathbb{R}^n} \Phi(x) = \max_{i\in I} \min_{j\in J} f_{i,j}(x)$$

with finite sets I, J and continuously differentiable functions $f_{i,j}$ with bounded first derivatives for all $i \in I$, $j \in J$.

The min–max–min problem has applications in multiple fields including computer aided design [1], facility location [2] and civil engineering [3].

A. Tsoukalas (✉) · P. Parpas · B. Rustem
Department of Computing, Imperial College, London, UK
e-mail: at102@doc.ic.ac.uk

P. Parpas
e-mail: pp500@doc.ic.ac.uk

B. Rustem
e-mail: br@doc.ic.ac.uk

To tackle the problem, we generalize the smoothing technique described in [3–7], which approximates a max function

$$\max_i g_i(x)$$

with the smooth approximation

$$\log\left(\sum_i \exp(\epsilon g_i(x))\right)/\epsilon.$$

The same idea, where in this case the maximization is over an interval and an integral appears in place of the summation, appears in the laplace method, that studies the asymptotic behavior of integrals [8]. In the context of optimization, it has been used in convergence proofs of simulated annealing methods [9] to describe a distribution over the state space of solutions that converges to the global optimum as the parameter ϵ increases [10].

This smoothing technique has been used to solve the minimax problem in [5–7] by transforming it into a smooth minimization problem. In [7] an adaptive method for the update of the parameter ϵ is used which provides robustness in view of potential ill-conditioning problems. Furthermore, the same technique has been used to transform the min–max–min problem to minimax problems [3], which are solved with a different method.

In Sect. 2 we extend this methodology to eliminate both the inner minimization and the maximization operands and reduce the min–max–min problem to the minimization of a smooth function and in Sect. 3 we provide a steepest descent algorithm of the approximation function and we show that it converges to a point that satisfies a first order optimality condition for the original min–max–min problem. In Sect. 4 we provide a computational example.

2 Smoothing of max–min functions

Proposition 2.1 provides a smooth approximation of $\max_{i\in I} \min_{j\in J} f_{i,j}(x)$.

Proposition 2.1 *Let*

$$\Phi(x) = \max_{i\in I} \min_{j\in J} f_{i,j}(x) \tag{1}$$

Let $M_I = |I|,\ M_J = |J|$ *and* $\epsilon_I > 0$, $\epsilon_J < 0$. *Then*

$$\Phi(x) + \frac{ln(M_J)}{\epsilon_J} \le \frac{1}{\epsilon_I} ln\left[\sum_{i\in I}\left[\sum_{j\in J}\exp(\epsilon_J f_{i,j}(x))\right]^{\frac{\epsilon_I}{\epsilon_J}}\right] \le \Phi(x) + \frac{ln(M_I)}{\epsilon_I}.$$

Proof Let

$$\Phi_i(x) = \min_j f_{i,j}(x).$$

Then

$$\Phi_i \leq f_{i,j}, \quad \forall i \in I, \quad j \in J,$$

$$\exp(\epsilon_J \Phi_i) \leq \sum_{j\in J} \exp(\epsilon_J f_{i,j}) \leq M_J \exp(\epsilon_J \Phi_i), \quad \forall i \in I,$$

$$\epsilon_J \Phi_i \leq \ln\left(\sum_{j\in J} \exp(\epsilon_J f_{i,j})\right) \leq \ln(M_J) + \epsilon_J \Phi_i, \quad \forall i \in I,$$

$$\Phi_i + \frac{1}{\epsilon_J} \ln(M_J) \leq \frac{1}{\epsilon_J} \ln\left(\sum_{j\in J} \exp(\epsilon_J f_{i,j})\right) \leq \Phi_i, \quad \forall i \in I,$$

$$\epsilon_I \Phi_i + \frac{\epsilon_I}{\epsilon_J} \ln(M_J) \leq \ln\left(\left[\sum_{j\in J} \exp\left(\epsilon_J f_{i,j}\right)\right]^{\frac{\epsilon_I}{\epsilon_J}}\right) \leq \epsilon_I \Phi_i, \quad \forall i \in I,$$

$$M_J^{\frac{\epsilon_I}{\epsilon_J}} \cdot \exp(\epsilon_I \Phi_i) \leq \left[\sum_{j\in J} \exp(\epsilon_J f)\right]^{\frac{\epsilon_I}{\epsilon_J}} \leq \exp(\epsilon_I \Phi_i), \quad \forall i \in I,$$

$$\begin{aligned} \ln\left(\sum_{i\in I} \exp(\epsilon_I \Phi_i)\right) + \frac{\epsilon_I}{\epsilon_J} ln(M_J) &\leq \ln\left(\sum_{i\in I}\left[\sum_{j\in J} \exp(\epsilon_J f_{i,j})\right]^{\frac{\epsilon_I}{\epsilon_J}}\right) \\ &\leq \ln\left(\sum_{i\in I} \exp(\epsilon_I \Phi_i)\right). \end{aligned} \tag{2}$$

We have

$$\Phi = \max_i \Phi_i$$

$$\exp(\epsilon_I \Phi) \leq \sum_{i\in I} \exp(\epsilon_I \Phi_i) \leq M_I \exp(\epsilon_I \Phi)$$

$$\epsilon_I \Phi \leq \ln\left(\sum_{i\in I} \exp(\epsilon_I \Phi_i)\right) \leq \ln(M_I) + \epsilon_I \Phi. \tag{3}$$

From (2,3), we have

$$\ln(M_I) + \epsilon_I \Phi \geq \ln\left(\sum_{i\in I} \exp(\epsilon_I \Phi_i)\right)$$

$$\geq \ln\left(\sum_{i\in I}\left[\sum_{j\in J} \exp(\epsilon_J f)\right]^{\frac{\epsilon_I}{\epsilon_J}}\right)$$

$$\geq \ln\left(\sum_{i\in I} \exp(\epsilon_I \Phi_i)\right) + \frac{\epsilon_1}{\epsilon_J}\ln(M_J)$$

$$\geq \epsilon_I \Phi + \frac{\epsilon_1}{\epsilon_J}\ln(M_J)$$

$$\Phi + \frac{1}{\epsilon_J}\ln(M_J) \leq \frac{1}{\epsilon_I}\ln\left(\sum_{i\in I}\left[\sum_{j\in J} \exp(\epsilon_J f)\right]^{\frac{\epsilon_I}{\epsilon_J}}\right) \leq \frac{1}{\epsilon_I}\ln(M_I) + \Phi$$

□

Set

$$\Phi_{\epsilon_I,\epsilon_J}(x) = \frac{1}{\epsilon_I}\ln\left(\sum_{i\in I}\left[\sum_{j\in J} \exp(\epsilon_J f(x))\right]^{\frac{\epsilon_I}{\epsilon_J}}\right)$$

It is shown in [5] that the convergence of the approximation method described in the Introduction is monotonic. Proposition 2.2 shows that this is not the case when we eliminate both operands.

Proposition 2.2 *For* $\epsilon_{J_2} < \epsilon_{J_1} < 0$*,*

$$\Phi_{\epsilon_I,\epsilon_{J_1}}(x) < \Phi_{\epsilon_I,\epsilon_{J_2}}(x),$$

and for $\epsilon_{I_2} > \epsilon_{I_1} > 0$*,*

$$\Phi_\alpha(x, \epsilon_{I_1}, \epsilon_J) > \Phi_\alpha(x, \epsilon_{I_2}, \epsilon_J).$$

Proof Let $g_i(x, \epsilon_J) = \left[\sum_{j\in J} \exp(\epsilon_J f_{i,j}(x))\right]^{\frac{1}{\epsilon_J}} > 0.$

$$
\begin{aligned}
\frac{g_i(x,\epsilon_{J_2})}{g_i(x,\epsilon_{J_1})} &= \frac{\left[\sum_{j\in J}\exp(\epsilon_{J_2} f_{i,j}(x))\right]^{\frac{1}{\epsilon_{J_2}}}}{\left[\sum_{j\in J}\exp(\epsilon_{J_1} f_{i,j}(x))\right]^{\frac{1}{\epsilon_{J_1}}}} \\
&= \left[\frac{\sum_{j\in J}\exp(\epsilon_{J_2} f_{i,j}(x))}{\left[\sum_{j\in J}\exp(\epsilon_{J_1} f_{i,j}(x))\right]^{\frac{\epsilon_{J_2}}{\epsilon_{J_1}}}}\right]^{\frac{1}{\epsilon_{J_2}}} \\
&= \left[\sum_{j\in J}\left[\frac{\exp(\epsilon_{J_1} f_{i,j}(x))}{\sum_{k\in J}\exp(\epsilon_{J_1} f_{i,k}(x))}\right]^{\frac{\epsilon_{J_2}}{\epsilon_{J_1}}}\right]^{\frac{1}{\epsilon_{J_2}}}.
\end{aligned}
$$

Since

$$
\frac{\epsilon_{J_2}}{\epsilon_{J_1}} > 1, \quad \frac{\exp(\epsilon_{J_1} f_{i,j}(x))}{\sum_{k\in J}\exp(\epsilon_{J_1} f_{i,k}(x))} < 1 \quad \text{and} \quad \epsilon_{J_2} < 0,
$$

we have

$$
\frac{g_i(x,\epsilon_{J_2})}{g_i(x,\epsilon_{J_1})} \geq \left[\sum_{j\in J}\frac{\exp(\epsilon_{J_1} f_{i,j}(x))}{\sum_{k\in J}\exp(\epsilon_{J_1} f_{i,k}(x))}\right]^{\frac{1}{\epsilon_{J_2}}} = 1.
$$

Therefore

$$
\begin{gathered}
g_i(x,\epsilon_{J_2}) \geq g_i(x,\epsilon_{J_1}) \\
\frac{1}{\epsilon_I}\ln\left(\sum_{i\in I} g_i(x,\epsilon_{J_2})^{\epsilon_I}\right) \geq \frac{1}{\epsilon_I}\ln\left(\sum_{i\in I} g_i(x,\epsilon_{J_1})^{\epsilon_I}\right) \\
\Phi_{\epsilon_I,\epsilon_{J_2}}(x) \geq \Phi_{\epsilon_I,\epsilon_{J_1}}(x).
\end{gathered}
$$

For the second part we have

$$
\begin{aligned}
\frac{\left[\sum_{i\in I} g_i(x,\epsilon_J)^{\epsilon_{I_2}}\right]^{\frac{1}{\epsilon_{I_2}}}}{\left[\sum_{i\in I} g_i(x,\epsilon_J)^{\epsilon_{I_1}}\right]^{\frac{1}{\epsilon_{I_1}}}} &= \left[\frac{\sum_{i\in I} g_i(x,\epsilon_J)^{\epsilon_{I_2}}}{\left[\sum_{i\in I} g_i(x,\epsilon_J)^{\epsilon_{I_1}}\right]^{\frac{\epsilon_{I_2}}{\epsilon_{I_1}}}}\right]^{\frac{1}{\epsilon_{I_2}}} \\
&= \left[\frac{\sum_{i\in I} g_i(x,\epsilon_J)^{\epsilon_{I_1}\frac{\epsilon_{I_2}}{\epsilon_{I_1}}}}{\left[\sum_{i\in I} g_i(x,\epsilon_J)^{\epsilon_{I_1}}\right]^{\frac{\epsilon_{I_2}}{\epsilon_{I_1}}}}\right]^{\frac{1}{\epsilon_{I_2}}} \leq 1.
\end{aligned}
$$

Therefore

$$
\begin{gathered}
\left[\sum_{i\in I} g_i(x,\epsilon_J)^{\epsilon_{I_2}}\right]^{\frac{1}{\epsilon_{I_2}}} \le \left[\sum_{i\in I} g_i(x,\epsilon_J)^{\epsilon_{I_1}}\right]^{\frac{1}{\epsilon_{I_1}}} \\
\ln\left(\left[\sum_{i\in I} g_i(x,\epsilon_J)^{\epsilon_{I_2}}\right]^{\frac{1}{\epsilon_{I_2}}}\right) \le \ln\left(\left[\sum_{i\in I} g_i(x,\epsilon_J)^{\epsilon_{I_1}}\right]^{\frac{1}{\epsilon_{I_1}}}\right) \\
\Phi_{\epsilon_{I_2},\epsilon_J}(x) \le \Phi_{\epsilon_{I_1},\epsilon_J}(x).
\end{gathered}
$$

□

We see that when we increase $|\epsilon_I|$, $\Phi_{\epsilon_I,\epsilon_J}$ decreases and when we increase $|\epsilon_J|$, $\Phi_{\epsilon_I,\epsilon_J}$ increases. In order to prove convergence of the local optimization algorithm it would be convenient if the approximation function was decreasing with both $|\epsilon_I|$ and $|\epsilon_J|$. Consider the alternative approximation function

$$\Phi^{'}_{\epsilon_I,\epsilon_J}(x) = \Phi_{\epsilon_I,\epsilon_J}(x) - \frac{\ln(M_J)}{\epsilon_J}.$$

It is obvious that $\Phi^{'}_{\epsilon_I,\epsilon_J}$ decreases as $|\epsilon_I|$ increases. Next we examine what happens when we increase $|\epsilon_J|$.

Lemma 2.3 *For all $b > 1, 0 \le a_i \le 1$, we have*

$$\left(\frac{1}{1+\sum_{i=1}^{m-1} a_i}\right)^b + \sum_{i=1}^{m-1}\left(\frac{a_i}{1+\sum_{k=1}^{m-1} a_k}\right)^b \ge m^{1-b}.$$

Proof Let

$$f(a_1, a_2, \ldots, a_{m-1}) = 1 + \sum_{i=1}^{m-1} a_i^b - m^{1-b}\left(1+\sum_{i=1}^{m-1} a_i\right)^b, \quad 0 \le a_i \le 1.$$

The set $\{(a_1, \ldots, a_{m-1}) | 0 \le a_i \le 1\}$ is closed and bounded and therefore, f has a global minimum a^*.

The partial derivatives of f are

$$\frac{\partial f}{\partial a_j} = b a_j^{b-1} - m^{1-b} b\left(1+\sum_{i=1}^{m-1} a_i\right)^{b-1}.$$

Since

$$\frac{\partial f(a_1, \ldots, a_j = 0, \ldots, a_{m-1})}{\partial a_j} < 0$$

a_i^* is strictly positive for all i. Solving the system

$$\frac{\partial f}{\partial a_j} = 0, \quad \forall j = 1, \ldots, m-1,$$

we obtain the unique solution $a_j = 1$ for all j. Since this boundary point is the only stationary point and there is no other boundary point eligible for a global minimum, we have $a_j^* = 1$ for all j.

As a result we obtain

$$f(a_1, \ldots, a_{m-1}) \geq f(1, 1, \ldots, 1) = 1 + m - 1 - m^{1-b} m^b = 0.$$

Dividing by $\left(1 + \sum_{i=1}^{m-1} a_i\right)^b$ and re-arranging, we have

$$\left(\frac{1}{1 + \sum_{i=1}^{m-1} a_i}\right)^b + \sum_{i=1}^{m-1} \left(\frac{a_i}{1 + \sum_{k=1}^{m-1} a_k}\right)^b \geq m^{1-b}.$$

□

Proposition 2.4

$$\Phi'_{\epsilon_I, \epsilon_{J_1}}(x) \geq \Phi'_{\epsilon_I, \epsilon_{J_2}}(x)$$

for $\epsilon_{J_2} < \epsilon_{J_1} < 0$.

Proof Let $g_i'(x, \epsilon_J) = \left[\sum_{j \in J} \exp(\epsilon_J f_{i,j}(x))\right]^{\frac{1}{\epsilon_J}} M_J^{\frac{-1}{e_j}} > 0$.

$$\begin{aligned}
\frac{g_i'(x, \epsilon_{J_2})}{g_i'(x, \epsilon_{J_1})} &= \frac{\left[\sum_{j \in J} \exp(\epsilon_{J_2} f_{i,j}(x))\right]^{\frac{1}{\epsilon_{J_2}}}}{\left[\sum_{j \in J} \exp(\epsilon_{J_1} f_{i,j}(x))\right]^{\frac{1}{\epsilon_{J_1}}}} \frac{M_J^{\frac{1}{\epsilon_{J_1}}}}{M_J^{\frac{1}{\epsilon_{J_2}}}} \\
&= \left[\frac{\sum_{j \in J} \exp(\epsilon_{J_2} f_{i,j}(x))}{\left[\sum_{j \in J} \exp(\epsilon_{J_1} f_{i,j}(x))\right]^{\frac{\epsilon_{J_2}}{\epsilon_{J_1}}}}\right]^{\frac{1}{\epsilon_{J_2}}} \frac{M_J^{\frac{1}{\epsilon_{J_1}}}}{M_J^{\frac{1}{\epsilon_{J_2}}}} \\
&= \left[\sum_{j \in J} \left[\frac{\exp(\epsilon_{J_1} f_{i,j}(x))}{\sum_{k \in J} \exp(\epsilon_{J_1} f_{i,k}(x))}\right]^{\frac{\epsilon_{J_2}}{\epsilon_{J_1}}}\right]^{\frac{1}{\epsilon_{J_2}}} \frac{M_J^{\frac{1}{\epsilon_{J_1}}}}{M_J^{\frac{1}{\epsilon_{J_2}}}}
\end{aligned}$$

Let

$$f_i^o(x) = \max_j f_{i,j}(x) \quad \text{and} \quad a_{i,j} = \frac{\exp(f_{i,j}(x))}{\exp(f_i^o(x))}.$$

We have $0 \le a_{i,j} \le 1 \quad \forall i, j \quad$ and $\forall i, \quad \exists j$ with $a_{i,j} = 1$. Therefore

$$\frac{g_i'(x, \epsilon_{J_2})}{g_i'(x, \epsilon_{J_1})} = \left[\sum_{j\in J}\left[\frac{a_{i,j}\exp(\epsilon_{J_1} f_i^o(x))}{\sum_{k\in J} a_{i,k}\exp(\epsilon_{J_1} f_i^o(x))}\right]^{\frac{\epsilon_{J_2}}{\epsilon_{J_1}}}\right]^{\frac{1}{\epsilon_{J_2}}} \frac{M_J^{\frac{1}{\epsilon_{J_1}}}}{M_J^{\frac{1}{\epsilon_{J_2}}}}$$

$$= \left[\sum_{j\in J}\left[\frac{a_{i,j}}{\sum_{k\in J} a_{i,k}}\right]^{\frac{\epsilon_{J_2}}{\epsilon_{J_1}}}\right]^{\frac{1}{\epsilon_{J_2}}} \frac{M_J^{\frac{1}{\epsilon_{J_1}}}}{M_J^{\frac{1}{\epsilon_{J_2}}}}$$

$$\le \left(M_J^{1-\frac{\epsilon_{J_2}}{\epsilon_{J_1}}}\right)^{\frac{1}{\epsilon_{J_2}}} \frac{M_J^{\frac{1}{\epsilon_{J_1}}}}{M_J^{\frac{1}{\epsilon_{J_2}}}} = 1$$

where the inequality holds due to Lemma 2.3 and the fact that $\frac{\epsilon_{J_2}}{\epsilon_{J_1}} \ge 1$, $\frac{1}{\epsilon_{J_2}} < 0$. Therefore,

$$g_i'(x, \epsilon_{J_2}) \le g_i'(x, \epsilon_{J_1})$$
$$\Phi'_{\epsilon_I, \epsilon_{J_2}}(x) \le \Phi'_{\epsilon_I, \epsilon_{J_1}}(x).$$

□

Choosing

$$\epsilon_I = -\epsilon_J = \epsilon > 0$$

we obtain the approximation function

$$\Phi_\epsilon(x) = \frac{1}{\epsilon}\ln\left(\sum_{i\in I}\frac{1}{\sum_{j\in J}\exp(-\epsilon f_{i,j}(x))}\right) + \frac{\ln(M_J)}{\epsilon}$$

which is a decreasing function of ϵ.

3 A steepest-descent algorithm

The following theorem from [11], gives a first order optimality condition for min–max–min problems.

Theorem 3.1 *If $f_{i,j}$ are continuously differentiable, and $\hat{x}$ is a local minimizer of $\Phi(x)$ then $0 \in \bar{G}\Phi(x)$ where*

$$\bar{G}\Phi(x) = conv_i conv_j \left\{\begin{bmatrix} f_{i,j}(x) - \Phi_i(x) \\ \Phi(x) - \Phi_i(x) \\ \nabla f_{i,j}(x) \end{bmatrix}\right\}.$$

We have

$$\nabla \Phi_\epsilon(x) = \sum_{i \in I, j \in J} \mu_{i,j}^\epsilon \nabla f_{i,j}(x) \tag{4}$$

where

$$\mu_{i,j}^\epsilon = \frac{\frac{\exp(-\epsilon f_{i,j}(x))}{(\sum_{k \in J} \exp(-\epsilon f_{i,k}(x)))^2}}{\sum_{\rho \in I} \frac{1}{\sum_{k \in J} \exp(-\epsilon f_{\rho,k}(x))}}. \tag{5}$$

Note that

$$\sum_{i \in I, j \in J} \mu_{i,j}^\epsilon = 1, \quad \forall \epsilon,$$

and

$$\lim_{\epsilon \to \infty} \mu_{i,j}^\epsilon = 0 \quad \forall (i, j) \notin (\hat{I}, \hat{J}_i),$$

where

$$\hat{I} = \{i \in I : \Phi(x) = \min_j f_{i,j}(x)\},$$
$$\hat{J}_i = \{j \in J : \min_j f_{i,j}(x) = f_{i,j}(x)\}.$$

Corollary 3.2 *If*

$$\lim_{\epsilon \to \infty} \nabla \Phi_\epsilon(\hat{x}) = 0,$$

then $\hat{x}$ satisfies the stationarity condition of Theorem 3.1.

We present the following Armijo Gradient algorithm.

Concerning step 7 we remark that any increasing updating rule $u(\epsilon_k)$ can be used as long as

$$\lim_{k \to \infty} \epsilon_k = \infty. \tag{6}$$

For example we can choose $u(\epsilon_k) = \epsilon_k + \gamma$. Condition 6 is also satisfied, by the adaptive parameter update proposed in [3,7], which has the advantage of avoiding ill-conditioning due to a quick increase or ϵ.

The proof of the next proposition is similar to the proof of stationarity of limit points for gradient methods in [12]. We repeat the arguments for completeness.

Proposition 3.3 *Let $\{x_k\}$ be a bounded sequence generated by Algorithm 1. Then the sequence $\{x_k\}$ converges to a stationary point.*

Algorithm 1 Armijo Gradient Algorithm

1: Set $\sigma, \beta \in (0, 1)$, $s > 0$
2: Choose $\epsilon_0 > 0$
3: Set $k = 0$
4: Compute the steepest descent direction

$$h_k = -\nabla \Phi_{\epsilon_k}(x_k).$$

5: Compute the step-size

$$\alpha_k = \max_{l \in \mathbb{N}} \{s\beta^l : \Phi_{\epsilon_k}(x_k + s\beta^l h(k)) - \Phi_{\epsilon_k}(x_k) \leq -\sigma\beta^l s||h_k||^2\}$$

6: Set $x_{k+1} = x_k + \alpha_k h_k$
7: Set $\epsilon_{k+1} = u(\epsilon_k), \quad k = k + 1$
8: go to 4

Proof To arrive at a contradiction assume that $\hat{x}$ is a limit point of $\{x_k\}$ with

$$\lim_{\epsilon \to \infty} ||\nabla \Phi_\epsilon(\hat{x})|| \geq \rho,$$

with $\rho > 0$.

Since Φ_ϵ is continuous and monotonically non-increasing with respect to both ϵ and x, and given

$$|\Phi_{\epsilon_1} - \Phi_{\epsilon_2}| \leq \frac{M_I + M_J}{\min\{\epsilon_1, \epsilon_2\}}, \quad \forall \epsilon_1, \epsilon_2 > 0,$$

it follows that $\Phi_{\epsilon_k}(x_k)$ converges to the finite value

$$\Phi(\hat{x}) = \lim_{\epsilon \to \infty} \Phi_\epsilon(\hat{x}).$$

We have

$$\begin{aligned} \Phi_{\epsilon_k}(x_k) - \Phi_{\epsilon_{k+1}}(x_{k+1}) &\geq \Phi_{\epsilon_k}(x_k) - \Phi_{\epsilon_k}(x_{k+1}) \\ &\geq \sigma\alpha_k||\nabla \Phi_{\epsilon_k}(x_k)||^2, \end{aligned}$$

where the first inequality follows from Propositions 1.2, 1.4 and the second inequality by the definition of Armijo's rule.

Hence, since $\sigma > 0$,

$$\alpha_k||\nabla \Phi_{\epsilon_k}(x_k)||^2 \to 0,$$

and therefore, by the hypothesis,

$$\{\alpha_k\} \to 0.$$

From the definition of Armijo's rule, there is an index $\hat{k} \geq 0$ such that

$$\Phi_{\epsilon_k}(x_k) - \Phi_{\epsilon_k}\left(x_k - \frac{\alpha_k}{\beta}\nabla\Phi_{\epsilon}(x_k)\right) < \frac{\sigma\alpha_k}{\beta}||\nabla\Phi_{\epsilon}(x_k)||^2, \quad \forall k \geq \hat{k},$$

that is, the initial step-size will be reduced at least once for all $k \geq \hat{k}$.

This can be written as

$$\frac{\Phi_{\epsilon_k}(x_k) - \Phi_{\epsilon_k}\left(x_k - \frac{\alpha_k||\nabla\Phi_{\epsilon}(x_k)||}{\beta}\frac{\nabla\Phi_{\epsilon}(x_k)}{||\nabla\Phi_{\epsilon}(x_k)||}\right)}{\frac{\alpha_k||\nabla\Phi_{\epsilon}(x_k)||}{\beta}} < \sigma||\nabla\Phi_{\epsilon}(x_k)||, \quad \forall k \geq \hat{k}.$$

By the Mean Value Theorem, we obtain

$$\nabla\Phi_{\epsilon_k}\left(x_k - \tilde{\alpha}_k\frac{\nabla\Phi_{\epsilon}(x_k)}{||\nabla\Phi_{\epsilon}(x_k)||}\right)^T\frac{\nabla\Phi_{\epsilon}(x_k)}{||\nabla\Phi_{\epsilon}(x_k)||} < \sigma||\nabla\Phi_{\epsilon}(x_k)||, \quad \forall k \geq \hat{k},$$

where $\tilde{\alpha}_k$ is a scalar in the interval $\left[0, \frac{\alpha_k||\nabla\Phi_{\epsilon}(x_k)||}{\beta}\right]$.

Since $||\nabla\Phi_{\epsilon}||$ is bounded, taking limits we obtain

$$\lim_{\epsilon\to\infty}||\nabla\Phi_{\epsilon}(\hat{x})||(1-\sigma) \leq 0,$$
$$\lim_{\epsilon\to\infty}||\nabla\Phi_{\epsilon}(\hat{x})|| \leq 0,$$

which contradicts the initial assumption.

Therefore,

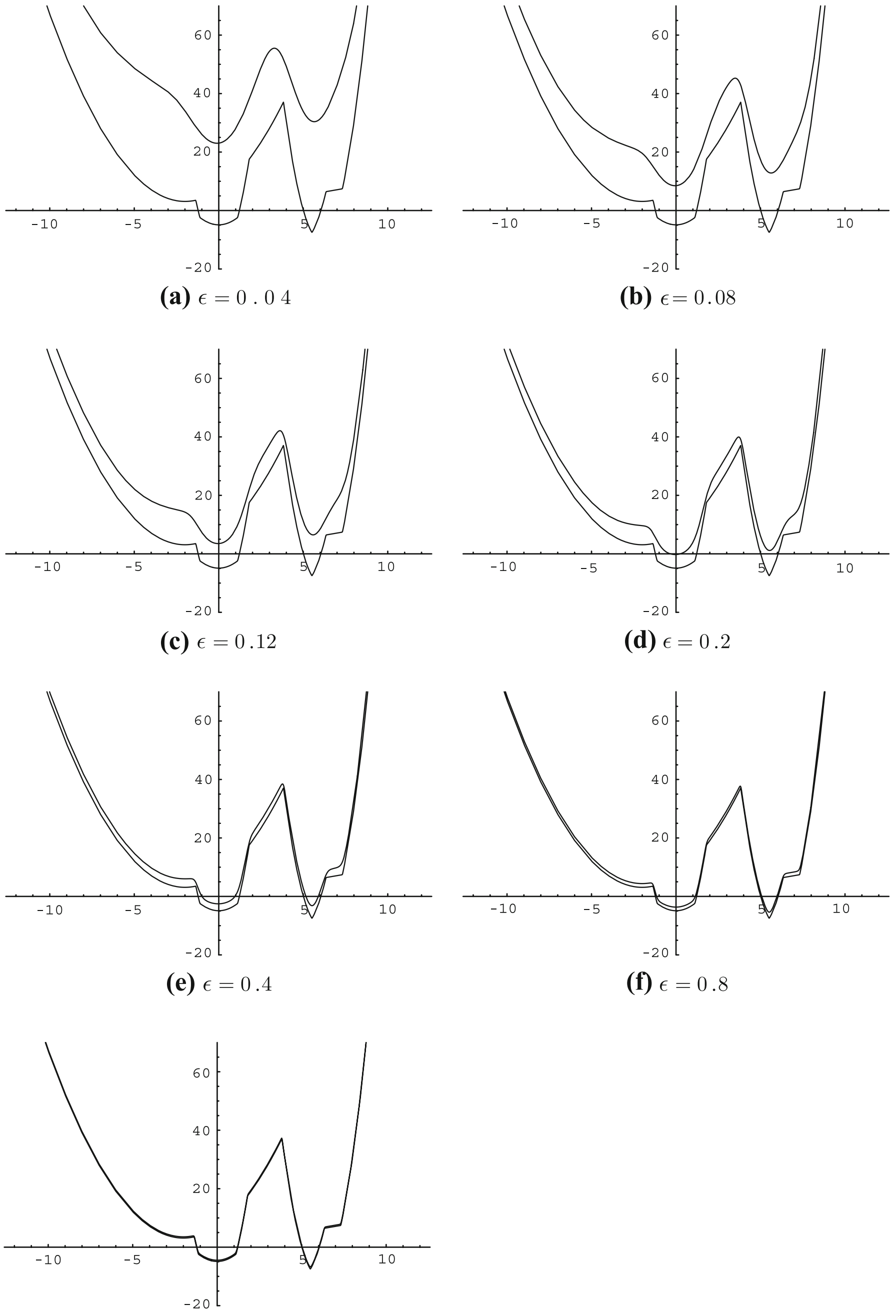
$$\lim_{\epsilon\to\infty}||\nabla\Phi_{\epsilon}(\hat{x})|| = 0$$

and $\hat{x}$ is a stationary point of $\Phi(x)$ by Corollary 3.2. □

4 Numerical example

To prevent overflows, the values of $\Phi_{\epsilon}(x)$ and $\mu_{i,j}^{\epsilon}$ have to be computed carefully, in a similar way to the approximations of mini–max problems in [4,6].

$$\begin{aligned}\Phi_{\epsilon}(x) &= \frac{1}{\epsilon}\ln\left(\sum_{i\in I}\frac{1}{\sum_{j\in J}\exp(-\epsilon f_{i,j}(x))}\right) + \frac{\ln(M_J)}{\epsilon}\\ &= \Phi(x) + \frac{1}{\epsilon}\ln\left(\sum_{i\in I}\frac{1}{\sum_{j\in J}\exp(-\epsilon(f_{i,j}(x) - \Phi(x)))}\right) + \frac{\ln(M_J)}{\epsilon},\end{aligned}$$



(a) $\epsilon = 0.04$ **(b)** $\epsilon = 0.08$

(c) $\epsilon = 0.12$ **(d)** $\epsilon = 0.2$

(e) $\epsilon = 0.4$ **(f)** $\epsilon = 0.8$

(g) $\epsilon = 2$

Fig. 1 Approximations for increasing values of ϵ

and

$$\mu_{i,j}^{\epsilon} = \frac{\frac{exp(-\epsilon f_{i,j}(x))}{(\sum_{k\in J} exp(-\epsilon f_{i,k}(x)))^2}}{\sum_{\rho\in I} \frac{1}{\sum_{k\in J} exp(-\epsilon f_{\rho,k}(x))}} = \frac{\frac{exp(-\epsilon(f_{i,j}(x)-\Phi(x)))}{(\sum_{k\in J} exp(-\epsilon(f_{i,k}(x)-\Phi(x))))^2}}{\sum_{\rho\in I} \frac{1}{\sum_{k\in J} exp(-\epsilon(f_{\rho,k}(x)-\Phi(x)))}}. \tag{7}$$

Example 1 Consider the problem

$$\min_{x\in\mathbb{R}} \max_{i\in\{1,2\}} \min_{j\in\{1,2,3\}} f_{i,j}(x)$$

with

$$f_{1,1}(x) = 10x^2 - 15, \quad f_{1,2}(x) = (x+2)^2 + 3, \quad f_{1,3}(x) = 10(x-6)^2 - 10,$$
$$f_{2,1}(x) = 2x^2 - 5, \quad f_{2,2}(x) = (3x-15)^2 - 10, \quad f_{2,3}(x) = x.$$

Figure 1 shows how the approximation function Φ_ϵ converges to Φ for increasing values of ϵ.

We run our algorithm with parameters $\sigma = 0.01$, $\beta = 0.5$, $s = 1$,$\epsilon_0 = 0.02$ and the updating parameter rule

$$\bar{\epsilon}_{k+1} = \left\{ \begin{array}{c} \epsilon_k \text{ if } ||\nabla\Phi_{\epsilon_k}(x_k)|| > 0.5 \\ \\ 2\epsilon_k \text{ if } ||\nabla\Phi_{\epsilon_k}(x_k)|| \le 0.5 \end{array} \right\}.$$

From the starting point $x_0 = -10$ our algorithm converges to $x^* = 0$ in 11 iterations. From the starting point $x_0 = 6$ the point $x^* = 5.51318$ is reached in 25 iterations.

Acknowledgment The authors wish to acknowledge support from EPSRC grant EP/C513584/1.

References

1. Mayne, D.Q., Polak, E., Trahan, R.: An outer approximations algorithm for computer-aided design problems. J. Optim. Theory Appl. **28**(3), 331–352 (1979)
2. Drezner, Z., Thisse, J.-F., Wesolowsky, G.O.: The minimax–min location problem. J. Reg. Sci. **26**(1), 87–101 (1986)
3. Polak, E., Royset, J.O.: Algorithms for finite and semi-infinite min–max–min problems using adaptive smoothing techniques. J. Optim. Theory Appl. **119**(3), 421–457 (2003)
4. Bertsekas, D.: Constrained Optimization and Lagrange Multiplier Methods. Athena Scientific, Nashua (1996)
5. Xingsi, L.: An entropy-based aggregate method for minimax optimization. Eng. Opt. **18**, 277–285 (1992)
6. Xu, S.: Smoothing method for minimax problems. Comput. Optim. Appl. **20**(3), 267–279 (2001)
7. Polak, E., Royset, J.O., Womersley, R.S.: Algorithms with adaptive smoothing for finite minimax problems. J. Optim. Theory Appl. **119**(3), 459–484 (2003)
8. De Bruijn, N.G.: Asymptotic methods in analysis, 2nd edn. Bibliotheca Mathematica, vol IV. North-Holland, Amsterdam (1961)
9. Parpas, P., Rustem, B., Pistikopoulos, E.N.: Linearly constrained global optimization and stochastic differential equations. J. Global Optim. **36**(2), 191–217 (2006)

10. Hwang, C.: Laplace's method revisited: weak convergence of probability measures. Ann. Probab. **8**(6), 1177–1182 (1980)
11. Polak, E.: Optimization. Applied Mathematical Sciences, vol. 124. In: Algorithms and consistent approximations. Springer, New York (1997)
12. Bertsekas, D.: Nonlinear Programming. Athena Scientific, Nashua (1999)