EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 3, No. 1, 2010, 98-106
ISSN 1307-5543 – www.ejpam.com

On the Cross-Entropic Regularization Method for Solving Min-Max Problems

Lili Zhang[1*], Jianyu Li[2] and Xingsi Li[3]

[1] *Department of Applied Mathematics, Dalian University of Technology, Dalian, P. R. China*
[2] *School of Mechanical Engineering, Tianjin University of Science and Technology, Tianjin, P. R. China*
[3] *State Key Laboratory of Structural Analysis for Industrial Equipment, Department of Engineering Mechanics, Dalian University of Technology, Dalian, P. R. China*

Abstract. A smoothing method of multipliers which is a natural result of cross-entropic regularization for min-max problems is analyzed. As a smoothing technique, we first show how the smooth approximation yields the first order information on the behavior of max function. Then under suitable assumptions, some basic properties including the Hessian are given. At last, the condition number is analyzed, and the results reveal that the smoothing method of multipliers is stable for any fixed smoothing parameter.

2000 Mathematics Subject Classifications: 90C47; 41A46; 57R12

Key Words and Phrases: Min-Max Problem; Cross-Entropic Regularization; Smooth Approximation; Subgradient; Condition Number

---

1. Introduction

A variety of scientific and engineering problems (such as the problems that arise in structural optimization, synthesis of filters, antenna design etc) can be formulated as the following min-max problem (see [1], [2], [3], [4]):

$$\min_{x \in X} \max_{1 \leqslant i \leqslant m} \{f_i(x)\} \tag{1}$$

where X is a common domain of the component functions $f_i(x), i = 1, 2, ...m$, which are usually assumed to be twice continuously differentiable. For a complete treatment of the min-max problems, see the books ([5], [6], [7]).

*Corresponding author.

Email address: `jnlilizhang@163.com` (L. Zhang)

One major difficulty encountered in developing solution methods is the non-differentiability of the max function:

$$F(x) := \max_{1 \leqslant i \leqslant m} \{f_i(x)\} \tag{2}$$

And along with non-smooth optimization methods, the smoothing technique has been used for the min-max since the early 70s([8]), (see also [3], [9], [10], [11], [12], [13]). Among them, a class called regularization methods has been developed base on approximating the max function by certain smooth functions([1], [2], [10], [15], [11], [13]). Gigola and Gomez ([10]), Hiriart-Urruty and Lemarechal ([14]) had given some regularization functions, but no explicit expressions of the smooth approximation functions were given in their work. Li ([15], [16]) used entropy function as the regularization function and derived a smooth and good approximation function in explicit expression for the max function $F(x) \approx \frac{1}{p} \ln \sum_{i=1}^{m} \exp(pf_i(x))$. This smooth approximation function is called aggregate function and instead of using the entropy regularization function, the cross-entropy regularization function can lead to another smooth and good approximation function which is a smooth approximation function of multipliers and is called cross-entropy aggregate function:

$$F_p(x, \lambda) = \frac{1}{p} \ln \sum_{i=1}^{m} \lambda_i \exp(pf_i(x)) \tag{3}$$

Both the two smoothing approximation functions have a natural interpretation which fits the special structure of the min-max problem and apply to related problems. The function (3) is also proposed by Bertsekas ([1]) by means of a tedious derivation using the multiplier methods with an exponential penalty function.

In this paper, we mainly investigate the approximation function (3), although the uniform convergence and the algorithmic convergence based on this function had been already given ([16], [1]), the main results in this paper have not appeared in either [16] or [1]. We first show how the smooth approximation yields the first order information on the behavior of max function. Then under suitable assumptions, some basic properties including the Hessian are given. As this smooth approximation function of multipliers is obtained by the cross-entropic regularization of the classical Lagrangian, if we treat this smooth approximation as a nonlinear Lagrangian function for the min-max problem, then based on the basic properties, a similar saddle point result can be obtained. At last, the condition number is analyzed, and the results reveal that the smoothing method of multipliers is stable for any fixed smoothing parameter.

The rest of this paper is arranged as follows: Section 2 gives the problem formulation and some basic assumptions. Some basic results including the first order information, basic properties and the condition number are given in section 3. Section 4 is the conclusion.

2. Problem Formulation and Basic Assumptions

Consider the min-max problem (1). The following are assumptions:

Assumption 1. *The optimal set X^* is not empty and bounded;*

Without loss of generality we can assume that $F(x^*) = 0$.

The min-max problem (1) has the classical Lagrangian function $L(x,\lambda) = \sum_{i=1}^{m} \lambda_i f_i(x)$ for each $x \in X$, where λ denotes the Lagrangian multipliers that are restricted to fall within the simplex $\Lambda \equiv \left\{ \lambda | \lambda \geqslant 0; \sum_{i=1}^{m} \lambda_i = 1 \right\}$. The following Karush-Kuhn-Tucker ($K-K-T$) conditions for problem (1) hold true:

$$\begin{cases} \nabla_x L(x^*,\lambda^*) = \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) = 0 \\ \lambda_i^* \geq 0, i = 1,2,\cdots,m, \qquad \sum_{i=1}^{m} \lambda_i^* = 1 \\ \lambda_i^* f_i(x^*) = 0 \qquad i = 1,2,\cdots,m \end{cases} \tag{4}$$

Let $I(x^*) = \{ i | f_i(x^*) = F(x^*) \}$ is the active set at point x^* and $|I(x^*)| = r$, where $|Q|$ is the cardinal number of the set Q. Respectively,
$\nabla f_I(x) = J(f_I(x)) = \left[\nabla f_{i_1}(x) \cdots \nabla f_{i_r}(x) \right]^T, i_k \in I(x) \quad k = 1,2,\cdots r;$
$\nabla f(x) = J(f(x)) = \left[\nabla f_1(x) \cdots \nabla f_m(x) \right]^T$ are their Jacobians. $\nabla f_I(x^*)$ linear independence, and the pair (x^*,λ^*) satisfies the second order optimality conditions:

$$\left\langle \nabla_{xx}^2 L(x^*,\lambda^*) d, d \right\rangle \geq \rho \langle d, d \rangle, \rho > 0, \forall d \neq 0 : \nabla f_I(x^*)^T d = 0 \tag{5}$$

the strictly complementary condition is true: $\lambda^* \in \Lambda$ and $\lambda_i^* > 0, i \in I(x^*)$.

Then the max function can be obtained as

$$F(x) = \max_{\lambda \in \Lambda} L(x,\lambda), \quad x \in X \tag{6}$$

Unfortunately, the above maximization rarely has an explicit solution. Therefore the regularization methods add some terms that are called regularization functions

$$L_p(x,\lambda) = L(x,\lambda) + \frac{1}{p} R(\lambda;\mu) \tag{7}$$

where $p > 0$ is a control parameter, R is a regularization function, and μ is an optional parameter vector of R. Then, by carefully choosing the function R, maximizing the $L_p(x,\lambda)$ could result in a smooth approximation function.

When $R(\lambda,\mu) = \sum_{i=1}^{m} \lambda_i \ln \frac{\lambda_i}{\mu_i}$, we have the following cross-entropic regularization formula:

$$F_p(x,\mu) \equiv \max_{\lambda \in \Lambda} \left\{ L(x,\lambda) - \frac{1}{p} R(\lambda;\mu) = \sum_{i=1}^{m} \lambda_i f_i(x) - \frac{1}{p} \sum_{i=1}^{m} \lambda_i \ln \frac{\lambda_i}{\mu_i} \right\} \tag{8}$$

where μ is an additional control vector representing some known as priori information (distribution).

A simple calculation results in the following solution:

$$\lambda_i^* = \frac{\mu_i \exp\left(pf_i\left(x\right)\right)}{Z}, i = 1, 2, \cdots, m \tag{9}$$

where $Z = \sum_{i=1}^{m} \mu_i \exp\left(pf_k\left(x\right)\right)$. Substituting (9) into (8), we have the smooth approximation function of multiplier (3): $F_p\left(x, \mu\right) = \frac{1}{p} \ln \sum_{i=1}^{m} \mu_i \exp\left(pf_i\left(x\right)\right)$. The reason for the name of the smooth approximation function of multiplier is that if the pair (x^l, μ^l) have been found already, we find the next approximation (x^{l+1}, μ^{l+1}) by the following formulas:

$$x^{l+1} = \arg\min_x \left\{F_p\left(x, \mu^l\right)\right\} \tag{10}$$

$$\mu_i^{l+1} = \frac{\mu_i^l \exp\left(pf_i\left(x\right)\right)}{\sum_{k=1}^{m} \mu_k^l \exp\left(pf_k\left(x\right)\right)}, i = 1, 2, \cdots, m \tag{11}$$

So μ is the Lagrangian multiplier in practice, we can just denote the smooth approximation function of multiplier as $F_p\left(x, \lambda\right) = \frac{1}{p} \ln \sum_{i=1}^{m} \lambda_i \exp\left(pf_i\left(x\right)\right)$.

3. Main Results

3.1. First Order Information: Subgradient

It is well known that the max function $F(x)$ has discontinuous first derivatives at points where two or more of the components $f_i(x)$ are equal to $F(x)$ even if each $f_i(x)$ is smooth with any order. So we first analyze whether the approximation function (3) could yield the first order information on the behavior of $F(x)$. The results reveal that for the convex situation the subgradients of $F(x)$ can be obtained from the gradients of smooth approximation function $F_p(x)$.

The first order behavior of convex non-smooth function around a point x is reflected by the the concepts of subgradients and subdifferential ([17] p214).

Definition 1. *A vector ω is said to be a subgradient of a convex function at a point x if*

$$f\left(z\right) - f\left(x\right) \geqslant \langle \omega, z - x \rangle \quad \forall z \in X \tag{12}$$

The set of all subgradients of f at x is called the subdifferential of f at x and is denoted by $\partial f\left(x\right)$, i.e.

$$\partial f\left(x\right) := \left\{\omega | f\left(z\right) - f\left(x\right) \geqslant \langle \omega, z - x \rangle \quad \forall z \in X\right\} \tag{13}$$

Theorem 1. *For the max function $F(x)$ and the smooth approximation function of multipliers $F_p(x, \lambda)$, the following statements are equivalent:*

$i)$ ω is a subgradient of F at x;
$ii)$ there is a weighted vector ξ, such that $\omega = \sum_{i \in I(x)} \xi_i \nabla f_i(x)$;
$iii)$ $\omega = \lim_{p \to \infty} \left\{ \nabla_x F_p(x, \lambda) \right\}$

Proof. The equivalence between (i) and (ii) is well known, and is obtained by applying standard calculus rules for computing subgradients ([17]).

$$\lim_{p \to \infty} \left\{ \nabla_x F_p(x, \lambda) \right\} = \lim_{p \to \infty} \left\{ \sum_{i=1}^{m} \frac{\lambda_i \exp\left(p f_i(x)\right)}{\sum\limits_{k=1}^{m} \lambda_k \exp\left(p f_k(x)\right)} \nabla f_i(x) \right\} = \sum_{i \in I(x)} \lambda_i \nabla f_i(x) \tag{14}$$

as $\exp\left(p f_i(x)\right) \underset{p \to \infty}{\to} 0, i \notin I(x)$ and $\lambda \in \Lambda$.

Remark 1. *In this proof, we use the assumption that $F(x^*) = 0$. In fact, without this assumption, we can also obtain the result in theorem 1 by subtracting $F(x)$ both in the numerator and denominator in the middle part in formula (14).*

3.2. Basic Properties

For approximation function (3), the following results establish the basic properties of this function at any $K - K - T$ pair (x^*, λ^*):

Lemma 1. *For any $K - K - T$ pair (x^*, λ^*), the following hold for any $p \geq 0$:*
$i)$ $F_p(x^*, \lambda^*) = F(x^*) = 0$;
$ii)$ $\nabla_x F_p\left(x^*, \lambda^*, p\right) = 0$

Proof. $i)$Using the $K - K - T$ condition (4) directly.
$ii)$ As the computation in subsection 3.1 we have

$$\begin{aligned} \nabla_x F_p(x, \lambda) &= \sum_{i=1}^{m} \frac{\lambda_i \exp\left(p f_i(x)\right)}{\sum\limits_{k=1}^{m} \lambda_k \exp\left(p f_k(x)\right)} \nabla f_i(x) \\ &= \sum_{i=1}^{m} \frac{\lambda_i \nabla f_i(x)}{\sum\limits_{k=1}^{m} \lambda_k \left(\exp\left(p f_k(x)\right) - \exp\left(p f_i(x)\right)\right)} \end{aligned} \tag{15}$$

Applying the $K - K - T$ condition, we have

$$\nabla_x F_p\left(x^*, \lambda^*\right) = \sum_{i=1}^{m} \frac{\lambda_i^* \nabla f_i(x^*)}{\sum\limits_{k=1}^{m} \lambda_k^* \left(\exp\left(p f_k(x^*)\right) - \exp\left(p f_i(x^*)\right)\right)} = 0 \tag{16}$$

Theorem 2. *Let (x^*, λ^*) be $K-K-T$ pair of min-max problem (1). Assume that $\nabla f_i(x^*), i \in I(x^*)$ are linearly independent and x^* satisfy the second-order sufficiency condition. Then the following hold: There exists $\rho > 0$ and $p_0 > 0$,*

$$\left\langle \nabla^2_{xx} F_p\left(x^*, \lambda^*\right) d, d \right\rangle \geq \rho \langle d, d \rangle \tag{17}$$

for any fixed $p \geq p_0$ and $d \in R^n$.

Proof.

$$\begin{aligned} \nabla^2_{xx} F_p(x, \lambda) &= \\ & \sum_{i=1}^{m} \frac{\lambda_i \exp\left(p f_i(x)\right)}{\sum\limits_{k=1}^{m} \lambda_k \exp\left(p f_k(x)\right)} \nabla^2 f_i(x) + p \sum_{i=1}^{m} \frac{\lambda_i \exp\left(p f_i(x)\right) \nabla f_i(x)^T \nabla f_i(x)}{\sum\limits_{k=1}^{m} \lambda_k \exp\left(p f_k(x)\right)} \\ &- p \frac{\left(\sum\limits_{i=1}^{m} \lambda_i \exp\left(p f_i(x)\right) \nabla f_i(x)\right)^T \sum\limits_{i=1}^{m} \lambda_i \exp\left(p f_i(x)\right) \nabla f_i(x)}{\left(\sum\limits_{k=1}^{m} \lambda_k \exp\left(p f_k(x)\right)\right)^2} \end{aligned} \tag{18}$$

and

$$\begin{aligned} \nabla^2_{xx} F_p\left(x^*, \lambda^*\right) &= \\ & \sum_{i=1}^{m} \frac{\lambda_i^* \exp\left(p f_i(x^*)\right)}{\sum\limits_{k=1}^{m} \lambda_k^* \exp\left(p f_k(x^*)\right)} \nabla^2 f_i\left(x^*\right) \\ &+ p \sum_{i=1}^{m} \frac{\lambda_i^* \exp\left(p f_i(x^*)\right) \nabla f_i(x^*)^T \nabla f_i(x^*)}{\sum\limits_{k=1}^{m} \lambda_k^* \exp\left(p f_k(x^*)\right)} \\ &- p \frac{\left(\sum\limits_{i=1}^{m} \lambda_i^* \exp\left(p f_i(x^*)\right) \nabla f_i(x^*)\right)^T \sum\limits_{i=1}^{m} \lambda_i^* \exp\left(p f_i(x^*)\right) \nabla f_i(x^*)}{\left(\sum\limits_{k=1}^{m} \lambda_k^* \exp\left(p f_k(x^*)\right)\right)^2} \end{aligned} \tag{19}$$

$$\begin{aligned} &= \nabla^2_{xx} L\left(x^*, \lambda^*\right) + p \sum_{i \in I(x^*)} \lambda_i^* \nabla f_i\left(x^*\right)^T \nabla f_i\left(x^*\right) \\ &- p \left(\sum_{i \in I(x^*)} \lambda_i^* \nabla f_i\left(x^*\right)\right)^T \left(\sum_{i \in I(x^*)} \lambda_i^* \nabla f_i\left(x^*\right)\right) \end{aligned} \tag{20}$$

as the convexity of the quadratic function, we have the inequality

$$\sum_{i\in I(x^*)} \lambda_i^* \nabla f_i\left(x^*\right)^T \nabla f_i\left(x^*\right) \geq \left(\sum_{i\in I(x^*)} \lambda_i^* \nabla f_i\left(x^*\right)\right)^T \left(\sum_{i\in I(x^*)} \lambda_i^* \nabla f_i\left(x^*\right)\right) \tag{21}$$

combining the second order optimality condition we have

$$\left\langle \nabla_{xx}^2 F_p\left(x^*,\lambda^*\right) d, d\right\rangle \geq \left\langle \nabla_{xx}^2 L\left(x^*,\lambda^*\right) d, d\right\rangle \geq \rho \left\langle d, d\right\rangle \tag{22}$$

Corollary 1. *if $f_i \in C^2$, $\nabla f_i(x^*), i \in I(x^*)$ are linearly independent and the second order sufficiently condition is satisfied, then for any $\lambda > 0$ and $p > 0$, the Hessian of $F_p(x,\lambda)$ is positive definite for any $x \in R^n$, i.e. $F_p(x,\lambda)$ is strictly convex in R^n and strongly convex on any bounded set in R^n.*

Remark 2. *As the approximation function (3) is obtained by the cross-entropic regularization of classical lagrangian, so when we treat function (3) as a nonlinear lagrangian function of the min-max problem, under the same conditions in theorem 2, for $p > p_0$ we can have the similar saddle point result*

$$F_p\left(x,\lambda^*\right) \geq F_p\left(x^*,\lambda^*\right) \geq F_p\left(x^*,\lambda\right), \forall x \in U\left(x,\varepsilon\right) \tag{23}$$

3.3. Condition Number

Based on the analysis in subsection 3.2, we give the following result about the condition number.

Theorem 3. *Let (x^*,λ^*) be $K-K-T$ pair of min-max problem (1), assume that $\nabla f_i(x^*), i \in I(x^*)$ are linearly independent and the second-order sufficiency condition is satisfied, then there exists p_0 and $M_0 \geq \tau_0 \geq 0$ that*

$$M_0 \left\langle d, d\right\rangle \geqslant \left\langle \nabla_{xx}^2 F_p\left(x^*,\lambda^*\right) d, d\right\rangle \geqslant \tau_0 \left\langle d, d\right\rangle \tag{24}$$

is true for any fixed $p \geq p_0$.

Proof. The right inequality is the result in theorem 2, and the left inequality follows from the formulas (20) and (21) if $p \geq p_0$ and p_0 large enough.

Corollary 2. *If $f_i(x)$ are twice continuous differentiable and $\varepsilon \geq 0$ is small enough then for any fixed $p \geq p_0$, there exists $M \geq \tau \geq 0$ such that for any pair $\omega = (x,\lambda) \in U(\omega^*,\varepsilon) = \{\omega | \omega - \omega^*| \leqslant \varepsilon\}$ the following inequalities*

$$M \left\langle d, d\right\rangle \geqslant \left\langle \nabla_{xx}^2 F_p\left(x^*,\lambda^*\right) d, d\right\rangle \geqslant \tau \left\langle d, d\right\rangle, \forall d \in R^n \tag{25}$$

hold true.

So the theorem 3 and the corollary 2 mean that in the neighborhood of the $K-K-T$ pair (x^*, λ^*) the condition number $cond \nabla^2_{xx} F_p(x, \lambda) \leqslant \frac{\tau}{M}$ is stable for any fixed $p \geq p_0$.

Remark 3. *Only the first order information results in subsection 3.1 need the convexity assumption. The results in subsection 3.2 and 3.3 are true whether $F(x)$ and all $f_i(x)$ are convex or not.*

4. Concluding remarks

As the uniform convergence and the algorithm are well known concepts, in this paper we have given some basic results that are required for the analysis of this research. It should be noticed that the efficiency of this smoothing method of multipliers depends on the parameter p. A similar parameter analysis in the nonlinear rescaling method for the constrained optimization problems is discussed in [18]. For this, some problems are investigated, such as, for the min-max problem to find an explicit general threshold value of p. The smoothing approximation has some properties that are related to the parameter p. The research (including the implementations, properties and so on) of the smoothing method of multipliers leads to an important subject for research.

ACKNOWLEDGEMENTS The authors would like to thank the referees for their helpful suggestions. This work is supported by the National Natural Science Foundation of China Grant(No. 10902077, 60675046).

References

[1] D.P.Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, Boston, MA, 1982.

[2] C.Charalammbous, A.R.Conn, An efficient method to solve the Minimax problem directly, *SIAM Journal on Numerical Analysis*, 15: 162-187 (1978).

[3] E.Polak, D.Q.Mayne, J.E.Higgins, Superlinearly convergent algorithm for Min-max problems, *Journal of Optimization Theory and Applications*, 69(3): 407-439 (1991).

[4] A.Ben-Tal, A.Nemirovsky, *Convex Optimization in Engineering: Modeling Analysis, Algorithms*, Technion, Israel, 1998.

[5] V.F.Demyanov, V.N.Molozemov, *Introduction to Minimax*, Wiley, New York, 1974.

[6] D.Z.Du, P.M.Pardalos, *Minimax and Applications*, Kluwer Academic Publishers, Dordrecht, 1995.

[7] E.Polak, *Optimization: Algorithm and Consistent Approximations*, Springer Verlag, New York, 1997.

[8] R.Polyak, *On the best convex Chebichev approximation*, Soviet Mathematics Doklady, 12: 1441-1444 (1971).

[9] G.Di.Phillo, P.L.Grippo, S.Lucidi, A smooth method for the finite Minimax problem, *Mathematical Programming*, 60: 187-214 (1993).

[10] C.Gigola, S.Gomez, A regularization method for solving the finite convex Min-max problem, *SIAM Journal on Numerical Analysis*, 27: 1621-1634 (1990).

[11] R.A.Polyak, Smooth optimization methods for Minimax problems, *SIAM Journal on Control and Optimization*, 26: 1274-1286 (1988).

[12] A.Vardi, New Minimax algorithm, *Journal of Optimization Theory and Applications*, 75: 613-633 (1992).

[13] I.Zang, A smoothing out technique for Min-max optimization, *Mathematical Programming*, 19: 61-77 (1980).

[14] J.B.Hiriart-Urruty, C.Lemarechal, *Convex analysis and minimization algorithm*, Springer-Verlag, Berlin, 1993.

[15] X S Li, Entropy and Optimization, Ph.D.Thesis, University of Livepool, United Kingdom, 1987.

[16] X S Li, S C Fang, On the entropic regularization method for solving Min-Max problems with applications, *Mathematical Methods of Operations Research*, 46: 119-130 (1997).

[17] R.T.Rockafellar, *Convex Analysis*, Princeton University Press, New Jersey, 1970.

[18] I.Griva, R.A.Polyak, Primal-dual nonlinear rescaling method with dynamic scaling parameter update, *Mathematical Programming*, 106: 237-259 (2006).