Sequential Convex Programming in Theory and Praxis

Dr. Christian Zillober
Mathematisches Institut, Universität Bayreuth, Postfach 10 12 51,
D-W 8580 Bayreuth, Germany

Abstract

In this paper, convex approximation methods, such as CONLIN, the method of moving asymptotes (MMA) and a stabilized version of MMA (Sequential Convex Programming), are discussed with respect to their convergence behaviour.
In an extensive numerical study they are finally compared with other well–known optimization methods at 72 examples of sizing problems.

1 Introduction

In the last few years the concept of convex approximations caused more and more interest in structural optimization. A structural optimization problem written in the form

$$\begin{aligned} \min \quad & f(x) \quad (x \in \mathbb{R}^n) \\ \text{s.t.} \quad & h_j(x) \leq 0 \ , \ j = 1..M \\ & x \in X \end{aligned} \tag{P1}$$

(where $X := \{x \mid \underline{x}_i \leq x_i \leq \overline{x}_i \ , \ i = 1..n\}$) is replaced by a sequence of easier to solve, convex, separable subproblems which approximate the original problem. The functions f and h_j $(j = 1..M)$ are assumed to be continuously differentiable and the feasible region is assumed to be non-empty.
The most general of these methods, the Method of Moving Asymptotes (MMA), is nowadays implemented in many software systems. But from a mathematical point of view it has one major drawback: It is not possible to show a global convergence theorem. Furthermore, there are simple examples showing that cycling of the method is possible. This was the reason for further investigations resulting in the paper of Zillober (1993). By adding a line search subject to a function measuring the global convergence, the behaviour of the method could be stabilized without loosing the known advantages.
In the following section we will describe the optimization methods CONLIN and MMA and outline some of their main features. In section 3 we will explain the stabilized method and state the most important result concerning global convergence. The numerical behaviour of these so–called *convex approximation methods* is illustrated by a study of ten optimization methods at 72 examples of sizing problems in the final section.

2 CONLIN and MMA

Based on the idea of using reciprocal variables Fleury and Braibant (1986) developed the optimization method CONLIN (convex linearization). An approximation of a function is defined by separate linearization for each component depending on the sign of the partial derivative at the expansion point. If the sign is positive then the linearization is performed with respect to the original variable, if the sign is negative then it is subject to the inverse variable, leading to a convex approximation of the original function.
The Method of Moving Asymptotes (MMA), however, is a generalization of CONLIN. Svanberg (1987) proposed a linearization with respect to substituted variables

$$\frac{1}{U_i - x_i} \quad \text{and} \quad \frac{1}{x_i - L_i}$$

respectively, where U_i and L_i are some chosen parameters.

Definition 2.1 *Let g be a continuously differentiable function on X. A MMA approximation $\tilde{g}$ of g is defined by*

$$\tilde{g}(\boldsymbol{x}) = g(\boldsymbol{x}^0) + \sum_{+} \left.\frac{\partial g}{\partial x_i}\right|_{x^0} \left(\frac{(U_i - x_i^0)^2}{U_i - x_i} - (U_i - x_i^0) \right) - \sum_{-} \left.\frac{\partial g}{\partial x_i}\right|_{x^0} \left(\frac{(x_i^0 - L_i)^2}{x_i - L_i} - (x_i^0 - L_i) \right)$$

where $\sum_{+}\left(\sum_{-}\right)$ means summation over all components i where the partial derivative $\frac{\partial g}{\partial x_i}$ at the expansion point $\boldsymbol{x}^0$ is non-negative (negative).
$\tilde{g}$ is defined on $D_{\tilde{g}} := \{\boldsymbol{x} \mid \max(L_i, \underline{x}_i) < x_i < \min(U_i, \overline{x}_i), i = 1..n\}$.

It is easy to verify that $\tilde{g}$ is a first order approximation of g that means

$$\tilde{g}(\boldsymbol{x}^0) = g(\boldsymbol{x}^0) \quad \text{and} \quad \nabla\tilde{g}(\boldsymbol{x}^0) = \nabla g(\boldsymbol{x}^0)\,,$$

and that $\tilde{g}$ is convex and separable, where g stands for objective or constraint function, respectively.

Remarks:

- The CONLIN method is obtained by letting $L_i = 0$ and $U_i \to \infty$ $(i = 1..n)$
- L_i and U_i are asymptotes for $\tilde{g}$
- A very efficient dual approach for solving the subproblem is applicable, cf. Svanberg (1987).

The algorithm proposed by Svanberg (1987) can be outlined as follows:

Step 0 : Choose a starting point $\boldsymbol{x}^0$, let $k := 0$
Step 1 : Compute $f(\boldsymbol{x}^k)$, $h_j(\boldsymbol{x}^k)$, $\nabla f(\boldsymbol{x}^k)$, $\nabla h_j(\boldsymbol{x}^k)$, $j = 1..M$
Step 2 : Define a subproblem, replacing f, h_j by $\tilde{f}$, $\tilde{h}_j$, $j = 1..M$, according to (2.1)
Step 3 : Solve this subproblem and let it's solution be denoted by $\boldsymbol{x}^{k+1}$. Let $k := k + 1$ and goto step 1

We neglect here more details, for example the choice of the asymptotes which are in general updated in each iteration. Some strategies can be found in Svanberg (1987) or Zillober (1993). For simplifying the notation we shall use the following abbreviation: By $SP(\boldsymbol{x}^k)$ (subproblem) we state the optimization problem resulting from step 2 of the proposed algorithm.
For the CONLIN method Nguyen et al. (1987) gave a convergence proof but only for the case that (P1) consists of concave functions which is of less practical interest. They indicated furthermore by some examples that a generalization to non-concave functions is not possible. This was the motivation to look for another way to prove convergence of CONLIN and MMA without loosing the good behaviour of the original method. These results are reported in the next section.

3 Theory of Sequential Convex Programming

It is well known that by adding a line search with respect to a function measuring the global convergence of an algorithm the behaviour of an optimization method is improved. Generally such a line search needs additional evaluations of the original functions of (P1). Therefore most people reject this idea to globalize MMA since its numerical performance is excellent even without a stabilization. In this section we proof that MMA together with a line search subject to an augmented Lagrange function leads to a globally convergent optimization method. But first we rewrite (P1) in order to get a simplified notation in this section:

$$\min f(\boldsymbol{x}) \text{ s.t. } h_j(\boldsymbol{x}) \leq 0,\ j = 1..m \qquad \textbf{(P2)}$$

where $m = M + 2n$, that means we write the box-constraints in the form $h_i(\boldsymbol{x}) \leq 0$, too. Now we introduce the augmented Lagrange function.

Definition 3.1 *The augmented Lagrange function $\Phi_r : \mathbb{R}^{n+m} \to \mathbb{R}$ associated to (P2) is for a fixed parameter $r > 0$ defined by:*

$$\Phi_r\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{u}\end{pmatrix} = f(\boldsymbol{x}) + \sum_{j=1}^{m}\begin{cases} u_j h_j(\boldsymbol{x}) + \frac{r}{2}h_j^2(\boldsymbol{x}) \ , & \text{if } -\frac{u_j}{r} \leq h_j(\boldsymbol{x}) \\ -\frac{u_j^2}{2r} \ , & \text{otherwise} \end{cases}$$

By J we denote the set of constraints $\left\{j \,|\, 1 \leq j \leq m : -\frac{u_j}{r} \leq h_j(\boldsymbol{x})\right\}$

This function is also used in the general purpose optimization method SQP (Sequential Quadratic Programming) by Schittkowski (1981) as a merit function and is known to work well for stabilizing the method.

Now we formulate the algorithm, which we call SCP (Sequential Convex Programming) in order to show the similarity to the SQP-method.

Step 0 : Choose $x^0 \in X$, $u^0 \geq 0$, $0 < c < 1$ (e.g. 0.001), $0 < \psi < 1$ (e.g. 0.5), $r > 0$ (e.g. 1), let $k := 0$

Step 1 : Compute $f(x^k)$, $\nabla f(x^k)$, $h_j(x^k)$, $\nabla h_j(x^k)$, $j = 1..m$

Step 2 : Compute L_i^k and U_i^k $(i = 1..n)$ by some scheme; define $\tilde{f}(x)$, $\tilde{h}_j(x)$, $j = 1..m$ (cf.(2.1))

Step 3 : Solve $SP(x^k)$; let $\begin{pmatrix} y^{k+1} \\ v^{k+1} \end{pmatrix}$ be the solution, where v^{k+1} denotes the corresponding vector of Lagrange multipliers

Step 4 : If $y^{k+1} = x^k$ stop; $\begin{pmatrix} x^k \\ u^k \end{pmatrix}$ is the solution

Step 5 : Let $s^k := \begin{pmatrix} x^k - y^{k+1} \\ u^k - v^{k+1} \end{pmatrix}$, $\delta^k := \| y^{k+1} - x^k \|$,

$$\eta^k := \min \left\{ \min_{i=1..n} \left\{ 2\epsilon \frac{(U_i^k - x_i^k)^2}{(U_i^k - L_i^k)^3} \right\} , \min_{i=1..n} \left\{ 2\epsilon \frac{(x_i^k - L_i^k)^2}{(U_i^k - L_i^k)^3} \right\} \right\} \quad (> 0)$$

Step 6 : Compute $\Phi_r \begin{pmatrix} x^k \\ u^k \end{pmatrix}$, $\nabla \Phi_r \begin{pmatrix} x^k \\ u^k \end{pmatrix}$, $\nabla \Phi_r \begin{pmatrix} x^k \\ u^k \end{pmatrix}^T s^k$

Step 7 : If $\nabla \Phi_r \begin{pmatrix} x^k \\ u^k \end{pmatrix}^T s^k < \frac{\eta^k (\delta^k)^2}{4}$ let $r := 10r$ and goto step 6; otherwise compute the smallest $j \in \mathbb{N}_0$, such that $\Phi_r \left(\begin{pmatrix} x^k \\ u^k \end{pmatrix} - \psi^j s^k \right) \leq \Phi_r \begin{pmatrix} x^k \\ u^k \end{pmatrix} - c\psi^j \nabla \Phi_r \begin{pmatrix} x^k \\ u^k \end{pmatrix}^T s^k$ (Armijo); let $\sigma^k := \psi^j$

Step 8 : Let $\begin{pmatrix} x^{k+1} \\ u^{k+1} \end{pmatrix} := \begin{pmatrix} x^k \\ u^k \end{pmatrix} - \sigma^k s^k$, $k := k + 1$, goto step 1

The major difficulties in proving global convergence for the new method were to show that the search direction defined in step 5 of the algorithm is a descent direction for Φ_r and that the resulting sequence of penalty parameters is bounded. These theorems can be found in Zillober (1993).
The main result of this chapter is the following:

Theorem 3.2 *Let the sequence $(x^k, u^k)_{k=0,1,2,...}$ be produced by SCP, all subproblems be uniquely solvable, gradients of active constraints at the optimal points of $SP(x^k)$ be linear independent and assume a continuous choice of asymptotes. For $\delta^k \neq 0$ we define: $\alpha^k := \frac{\|u^k - v^{k+1}\|^2}{(\delta^k)^2}$. Let the sequence fulfill the following two conditions:*

a) $j \in J$ if and only if $\tilde{h}_j(y^{k+1}) = 0$ $(j = 1..m)$ (that means, subproblems and original problem identify the same set of active constraints)

b) there is a $\alpha \in \mathbb{R}$ independent of k, such that $\alpha^k \leq \alpha < \infty$

Then the sequence either terminates at a stationary point, or it has at least one accumulation point and each accumulation point is a stationary point for (P2).

Proof: Zillober (1993)

A detailed description of the mathematical description of global convergence of the SCP–method can be found in Zillober (1993).

4 Numerical study

In our test we had 72 examples of sizing problems at our disposal, which have been a standard set used for testing in the Finite–Elements system LAGRANGE (Kneppe et al. (1987)). About 50% are pure test examples, the remaining examples are real life applications. The largest used dimensons are 385 nodes, 520 structural variables, 108 design variables and 522 constraints besides box–constraints. In 49 examples only displacement–dependent constraints were apparent, in 40 examples only stress–constraints appeared, which is important for the following discussion of the different methods. For all the examples we knew a best point from numerous previous runs. We compared the methods by the number of function evaluations, the number of gradient evaluations and the CPU–time they needed to fulfill

$$f(\boldsymbol{x}) \leq f(\boldsymbol{x}^*)(1+\epsilon) \quad \text{and}$$

$$\max_{i=1..m}\{h_i(\boldsymbol{x})\} \leq \epsilon,$$

where $\boldsymbol{x}^*$ denotes the best known point. We chose $\epsilon = 0.01$, that means, we allowed a difference of 1% in the objective function to the best known point, simultaneously allowing a maximum violation of 0.01 of the constraints (notice, that the constraints are scaled in LAGRANGE, such that values of different examples are comparable). All the methods were run on all examples. The tables below show the number of different rankings which the methods attained to fulfill the above conditions. Rank 1 means, this method needed the least number of function evaluations (gradient evaluations, CPU–time) to fulfill the above conditions. "F" denotes failure.
The following optimization methods were used:

1. IBF: Inverse Barrier Function method, cf. Fiacco and McCormick (1968)
2. MOM: Method of Multipliers, cf. Bertsekas (1982)
3. SLP: Sequential Linear Programming, cf. Haftka and Kamat (1985)
4. SRM: Stress Ratio Method, cf. Haftka and Kamat (1985)
5. SQP: Sequential Quadratic Programming, cf. Schittkowski (1981)
6. GRG: Generalized Reduced Gradients, cf. Bremicker (1986)
7. CON: CONLIN
8. QPR: Quadratic Programming with Reduced Line Search Technique, cf. Bremicker (1986)
9. SCP
10. MMA

Rankings of number of function evaluations:

Rank	1	2	3	4	5	6	7	8	9	10	F
IBF	2	–	–	2	1	1	1	4	3	2	56
MOM	2	1	4	2	2	3	8	13	12	–	25
SLP	18	3	10	14	8	2	1	1	–	–	15
SRM	5	4	3	4	1	4	–	–	–	–	51
SQP	8	5	17	12	12	4	2	2	–	–	10
GRG	4	2	5	2	5	11	10	10	1	–	22
CON	25	6	2	3	7	–	–	–	–	–	29
QPR	9	5	4	5	13	10	12	–	–	–	14
SCP	20	14	5	4	2	7	2	3	–	–	15
MMA	20	21	5	6	–	2	1	–	–	–	17

Rankings of number of gradient evaluations:

Rank	1	2	3	4	5	6	7	8	9	10	F
IBF	2	–	–	2	–	1	1	5	3	2	56
MOM	2	–	4	1	1	4	9	15	11	–	25
SLP	4	10	8	6	11	9	7	2	–	–	15
SRM	21	–	–	–	–	–	–	–	–	–	51
SQP	4	4	9	10	12	14	7	2	–	–	10
GRG	15	16	10	3	2	2	–	2	–	–	22
CON	8	8	11	5	1	7	2	1	–	–	29
QPR	30	17	5	3	3	–	–	–	–	–	14
SCP	12	4	9	10	8	5	1	7	1	–	15
MMA	11	5	11	11	10	5	2	–	–	–	17

Rankings of CPU-time:

Rank	1	2	3	4	5	6	7	8	9	10	F
IBF	–	–	–	4	1	1	1	4	3	2	56
MOM	–	1	4	2	1	6	8	16	9	–	25
SLP	12	6	11	7	11	3	5	–	2	–	15
SRM	18	2	1	–	–	–	–	–	–	–	51
SQP	2	2	7	5	9	12	15	8	2	–	10
GRG	4	4	6	4	8	12	7	5	–	–	22
CON	10	11	13	5	2	1	–	1	–	–	29
QPR	12	11	2	12	12	6	3	–	–	–	14
SCP	7	12	8	10	7	7	3	1	2	–	15
MMA	11	13	10	8	7	5	1	–	–	–	17

First of all, we have to mention that the SRM–method is only applicable for problems with stress–constraints that means it is not permitted for 32 examples which are contained in the "F"–column. The IBF–method needs a feasible initial point. This is true for 42 examples. The others are contained in the "F"–column, too.
Let's now focuse on the robustness of the different methods. SLP, SQP, QPR, SCP and MMA have about the same rates of failure. GRG fails more often, but it is still rather robust. The CONLIN–method has a notable larger number of non–succeeding runs, which confirms the theoretical results. MOM is not very robust, even if we need more accurate results. It is a typical phase 1–method which means it is applicable in a first rough approximation of the solution. The SRM–method fails in about the half of its possible applications. As for MOM the same conclusions can be drawn, i.e. its robustness decreases considerably with the desired accuracy. The IBF–method is not very robust, independent of the desired accuracy.
Considering the efficiency of the methods we conclude as follows. The SRM–method is very efficient. It doesn't need any gradient evaluation that means if it converges, it needs only function evaluations saving a lot of computation time. The convex approximation methods, SLP and QPR are comparably efficient. SQP and GRG are a bit worse. IBF and MOM have no chance from this point of view.
If we need more accurate results, then the SQP–method gets better and better. Although it is very robust, the excellent behaviour of the quadratic model of the original problem is a typical local property. Besides QPR which is a combination of SQP and GRG and therefore uses quadratic models itself, all other methods perform in this case worse as shown in the tables.
Finally, we want to focuse on the global convergence behaviour of SCP and MMA. In the tables, there is no considerably difference to observe. One might ask therefore, why there was a need to add the line–search as shown in ch. 3. But the rate of failure of MMA increases when using initial points which are more far away from the solution as the initial guesses of the shown examples (which are guesses of experienced designers). SCP, however, doesn't seem to be as dependent as MMA on the quality of initial guesses. The same conclusion can be drawn applying the methods to examples from other fields than sizing problems of structural optimization. With other words, SCP can be viewed as a general purpose method in contrast to MMA which is a typical method for structural optimization.

5 Conclusion

As most important result from our tests, we learned that there is no method which fulfills all possible requests of users. Even if we focuse on a more special purpose, e.g. on robustness for lower accuracies, it is not possible to predict which method is the best. Therefore it will always be a little bit a gamble to decide which method should be used for an optimization. But we can give some recommendations. If the solution doesn't has to be very accurate we could use SRM (if applicable) or CONLIN, if we take the risk of failure. These methods are very efficient in phase 1. If it is more important to get a result than to get it very fast, then it would be better to use SLP, SQP, QPR, SCP or MMA.

If we need more accurate results, then CONLIN is still very efficient. The efficiency of SRM decreases with the desired accuracy. But in this case too, the methods which are more robust in phase 1 don't loose this property in phase 2. Additionally, the efficiency of SQP and QPR increases, the more accurate the solution has to be.

Acknowledgement

This work was supported by the German Science Foundation (DFG) under grant Schi 173/6-1.

References

Bertsekas, D.P. (1982): *Constrained optimization and Lagrange multiplier methods.* Academic Press, New York

Bremicker, M. (1986): *Entwicklung eines Optimierungsalgorithmus der generalisierten reduzierten Gradienten, Anwendung auf Beispiele der Strukturdynamik.* Universität GH Siegen, Institut für Mechanik und Regelungstechnik

Fiacco, A.V.; McCormick, G.P. (1968): *Nonlinear Programming: Sequential unconstrained minimization techniques.* Wiley, New York

Fleury, C.; Braibant, V. (1986): *Structural Optimization – a new dual method using mixed variables.* Int. J. Num. Meth. Eng. 23, 409 - 428

Haftka, R.T.; Kamat, M.P. (1985): *Elements of Structural Optimization.* Martinus Nijhoff Publishers, Netherlands

Kneppe, G.; Krammer, J.; Winkler, F. (1987): *Structural Optimization of large scale problems using MBB LAGRANGE.* 5th World Congress and Exhibition on FEM, Salzburg, Austria

Nguyen, V.H.; Strodiot, J.J.; Fleury, C. (1987): *A mathematical convergence analysis for the convex linearization method for engineering design optimization.* Engineering Optimization 11, 195 - 216

Schittkowski, K. (1981): *The Nonlinear Programming method of Wilson, Han and Powell with an augmented Lagrangian type line search function.* Numerische Mathematik 38, 83 114

Svanberg, K. (1987): *The Method of Moving Asymptotes – a new method for Structural Optimization.* Int. J. Num. Meth. Eng. 24, 359 - 373

Zillober, Ch. (1993): *A globally convergent version of the Method of Moving Asymptotes.* To appear in Structural Optimization