Università
della
Svizzera
italiana

**Institute of
Computational
Science
ICS**

2020

Student: Niccolo Zuppichini

Notes on "Nash, A multigrid approach to discretized optimization problems"

1. Introduction

The goal of the paper is to provide a multigrid optimization scheme for non-linear optimization problems. The main idea of a multigrid approach to an optimization problem is to compute a search direction on the coarse grid and then apply in the finest grid. A multigrid approach to an optimization problem is often more straightforward than converting the problem into a system of equations. For instance, an optimization problem with inequality constraints is not trivially converted to a system of non-linear equations.
In particular, 4 are the direct aims of this paper:

1. Propose a line search approach to the multigrid method for optimization
2. Study convergence properties
3. Extend the method to non-convex problems
4. Report results on the performance of the proposed method

2. A multigrid approach to optimization problems

The paper presents a multigrid algorithm that works directly with the optimization problem 1. In this section, we present the algorithm for convex problems, and in the next section, we generalise for non-convex.

$$\min_x f(x) \tag{1}$$

The proposed multigrid can be divided into five steps.

Given an initial solution x_0 if you are on the coarsest grid, exactly solve $min_x f(x)$, otherwise:

1. Smoothing.
   Perform a N_{p_1} iterations of a smoother on $min_x \quad f(x)$ starting from x_0 and get x_1.

2. Solve sub-problem.

   Name $\tilde{x} = I_h^H x_1$
   Compute: $\tilde{v} = \nabla f_H(\tilde{x}) - I_h^H \nabla f_h(x_1)$

3. Compute the search direction p_k on the coarse grid.
   $min_x \quad f_H(\tilde{x}) - v^T \tilde{x}$ and obtain $\tilde{x}_2$

   Compute $p = I_H^h(\tilde{x}_2 - \tilde{x}_1)$

4. Perform a line search

   $min_x \quad f(x_1 + \alpha p)$ and obtain x_2

5. Post Smoothing Perform a N_{p_2} iterations of a smoother on $min_x \quad f(x)$ starting from x_2 and get x_3.

3. Convergence

The paper proves the convergence of the multigrid algorithm 5. It states that if:

1. the underlying optimization algorithm is globally convergent
2. at the least one parameter for the smoothers is positive
3. the direction p is descent

Then the algorithm 5 is globally convergent.

The Theorem (3) highly depends on whether the direction p is a descent direction. This is true if three conditions are satisfied:

1. The interpolation and restriction operators satisfy: $\exists C_{h,H} > 0 \quad I_h^h = C_{h,H}(I_h^H)^T$

2. Each individual optimization problem is convex.

3. The multigrid subproblem (step 3 of Algorithm 5) is solved accurately enough.

The proof that p is a descent direction is showed under the assumption that each individual subproblem is convex and that the interpolation and restriction operators satisfy the given condition.

When we minimise at step 3 of Algorithm 5, we obtain:

$$\nabla f_H(\tilde{x}_2) = \nabla f_H(\tilde{x}_1) - I_h^H \nabla f_h(x_1) + z \tag{2}$$

Where z is the error obtained by solving the problem in the coarse grid inexactly. We can make z as small as we want.
We want to check that the search direction p in Algorithm 5 is a descent direction i.e. $\nabla f_h(x_1)^T p < 0$
By definition of p in Algorithm 5

$$\nabla f_h(x_1)^T p = \nabla f_h(x_1)^T (I_H^h(\tilde{x}_2 - \tilde{x}_1))$$

We made the assumption in 3 that $\exists C_{h,H} > 0 \quad I_h^h = C_{h,H}(I_h^H)^T$.

$$= \nabla f_h(x_1)^T (C_{h,H}(I_h^H)(\tilde{x}_2 - \tilde{x}_1)) = C_{h,H}(\nabla f_h(x_1)^T I_h^H(\tilde{x}_2 - \tilde{x}_1)) =$$

By 2 it yields:

$$= C_{h,H}(\nabla f_H(\tilde{x}_1) - \nabla f_H(\tilde{x}_2) + z)^T(\tilde{x}_2 - \tilde{x}_1) =$$

By the mean value theorem $\nabla f_H(\tilde{x}_1) - \nabla f_H(\tilde{x}_2) \leq -\nabla^2 f(\varepsilon)(\tilde{x}_2 - \tilde{x}_1)$ for some $\varepsilon \in (\tilde{x}_1, \tilde{x}_2)$. It follows:

$$\leq -C_{h,H}(\tilde{x_2} - \tilde{x_1})^T \nabla^2 f_H(\varepsilon)(\tilde{x_2} - \tilde{x_1}) + z^T(\tilde{x_2} - \tilde{x_1})$$

If we are sufficiently close to the solution $(\tilde{x_2} - \tilde{x_1})^T \nabla^2 f_H(\varepsilon)(\tilde{x_2} - \tilde{x_1})$ is positive. The term $z^T(\tilde{x_2} - \tilde{x_1}))$ can be ignored if we are solving accurately enough the problem. Therefore:

$$\nabla f_h(x_1)^T p \leq 0$$

However, the paper shows that 3 is negative even if the subproblem is not solved to full accuracy.

$$z^T(\tilde{x_2} - \tilde{x_1})) < 0 \tag{3}$$

By recalling Algorithm 5, the point $\tilde{x_2}$ is the result of solving the subproblem on the coarse grid starting from an initial guess $\tilde{x_1}$. This means that $\tilde{x_2}$ is "closer" to a minimizer than $\tilde{x_1}$

$$f_H(\tilde{x_2}) - \tilde{v}^T \tilde{x_2} \leq f_H(\tilde{x_1}) - \tilde{v}^T \tilde{x_1} \tag{4}$$

$$f_H(\tilde{x_2}) - \tilde{v}^T \tilde{x_2} - f_H(\tilde{x_1}) - \tilde{v}^T \tilde{x_1} \leq 0$$

By applying the mean value theoreem to 4 it yields:

$$f_H(\tilde{x_2}) - \tilde{v}^T \tilde{x_2} - f_H(\tilde{x_1}) - \tilde{v}^T \tilde{x_1} = \nabla f_H(\tilde{\varepsilon})(\tilde{x_2} - \tilde{x_1}) - \tilde{v}^T (\tilde{x_2} - \tilde{x_1})$$

More formally:

$$\exists \varepsilon \in (\tilde{x_1}, \tilde{x_2})\text{s.t.} \quad (\nabla f_H(\tilde{\varepsilon}) - \tilde{v})^T (\tilde{x_2} - \tilde{x_1}) \leq 0$$

As multigrid converges the distance between $|\tilde{\varepsilon} - \tilde{x_2}|$ gets smaller. We can make the approximation :

$$(\nabla f_H(\tilde{x_2}) - \tilde{v})^T (\tilde{x_2} - \tilde{x_1}) \leq 0$$

We recall :

$$z = (\nabla f_H(\tilde{x_2}) - \tilde{v})$$

Then It directly follows:

$$z^T (\tilde{x_2} - \tilde{x_1}) \leq 0$$

This shows that even if the subproblem is not solved at full accuracy 3 is negative.

Then, $\nabla f_h(x_1)^T p < 0$ and therefore p is a descent direction.

3.1. Multigrid for optimization of non-convex problem

We can modify the Algorithm 5 to solve non-convex optimization problems by changing the step 3 into :

$$\min_x f_H(\tilde{x}) - v^T \tilde{x} + \frac{1}{2}(\tilde{x} - \tilde{x_1})^T D(\tilde{x} - \tilde{x_1})$$

with the result $\tilde{x_2}$ and continuing with Algorithm 5. This approach is analogous to a semi-Newton method where the matrix D is modified to be "appropriately" positive definite.

4. Computational results

The major point of the algorithm presented in this paper is that it's faster than the underlying algorithm used to solve the minimization subproblems (Figure ??). We consider the following optimization problem 5 as benchmark.

$$\min_{x(s,t)} \int\int \sqrt{1 + x_s^2 x_t^2} ds dt \tag{5}$$

TABLE I MG/OPT *versus* TN on minimal-surface problem

MG/OPT			*TN*		
Iterations	*Flops*	*Error*	*Iterations*	*Flops*	*Error*
0	0	4×10^1	0	0	4×10^1
1	0.6×10^7	1×10^0	16	6.5×10^7	4×10^1
2	0.9×10^7	1×10^0			
3	1.5×10^7	7×10^{-1}			
4	1.8×10^7	4×10^{-2}	20	8.0×10^7	4×10^{-2}
5	2.2×10^7	8×10^{-3}	25	9.6×10^7	8×10^{-3}
6	2.6×10^7	1×10^{-3}	27	10.2×10^7	1×10^{-3}
7	3.2×10^7	6×10^{-5}	32	11.5×10^7	6×10^{-5}
8	3.6×10^7	2×10^{-5}	35	12.6×10^7	2×10^{-5}

It's worth mentioning that employing a multigrid approach without line search would not converge in this problem showing the effectiveness line search approach to multigrid presented in this paper.

5. Conclusion

The paper showed how a high-performance multigrid approach can be used to solve optimization problems opening new areas of research. For example,

- the non-convex algorithm has not been tested in-depth and a trust-region approach may be preferable rather than the one proposed.
- The results of this paper are global and local convergence has not been studied.
- The problems analyzed in this paper have as variables discretized values of a continuous solution. But this class of problems are not the only one where a multigrid approach can be applied.