PAMM · Proc. Appl. Math. Mech. **5**, 735–736 (2005) / **DOI** 10.1002/pamm.200510342

On the convergence of the MG/OPT method

Alfio Borzì[*1]

[1] Institut für Mathematik und Wissenschaftliches Rechnen, Karl-Franzens-Universität Graz, Heinrichstr. 36, A-8010 Graz, Austria

Global convergence of the MG/OPT method for optimization is discussed.

1 Introduction

In some recent papers, Nash [5] and Lewis and Nash [4], propose a multigrid approach to optimization problems, called MG/OPT, which closely resembles the well known full approximation storage (FAS) scheme [1] and is similar to the nonlinear multigrid (NMGM) methods discussed in [2]. One novelty of the MG/OPT approach is the extension of the multigrid strategy to optimization problems.

In [4, 5] it is emphasized that under appropriate assumptions, the multigrid coarse-grid correction provides a descent direction and, therefore, combining this fact with a line search procedure and a minimizing 'smoothing' iteration, a globally convergent algorithm is obtained. Numerical experiments, e.g. [5], demonstrate that MG/OPT greatly improves the efficiency of the underlying optimization scheme used as 'smoother', suggesting that the MG/OPT scheme may be beneficial in combination with well known optimization algorithms. This claim appears to be true as far as a line search along the coarse-grid correction is performed. Also in [5] it is reported that MG/OPT without line search diverges in some cases. Therefore line search appears to be necessary for convergence.

In this note, we point out how the results of Hackbusch and Reusken [3] on the analysis of a damped nonlinear multigrid method for partial differential equations apply to the analysis of the MG/OPT scheme and suggest that an a priori choice of the coarse grid correction step-length can be made.

2 The MG/OPT method for optimization

Consider the following (locally) convex optimization problem

$$\min_{x_k} f_k(x_k) \tag{1}$$

where $k = 1, 2, \dots, L$, is the resolution or discretization parameter, L denotes the finest resolution, and x_k is the (unconstrained) optimization variable in the space V_k. For variables defined on V_k we introduce the inner product $(\cdot,\cdot)_k$ with associated norm $\|x\|_k = (x,x)_k^{1/2}$. Among spaces V_k, restriction operators $I_k^{k-1} : V_k \to V_{k-1}$ and prolongation operators $I_{k-1}^k : V_{k-1} \to V_k$ are defined. We require that $(I_k^{k-1}x, y)_{k-1} = (x, I_{k-1}^k y)_k$ for all $x \in V_k$ and $y \in V_{k-1}$.

On each space, denote with S_k an optimization algorithm. For example the truncated Newton scheme used in [5]. Given an initial approximation x_k^0 to the solution of (1), the application of S_k results in $f_k(S_k(x_k^0)) < f_k(x_k^0) - \eta\|\nabla f_k(x_k^0)\|^2$ for some $\eta \in (0,1)$.

The MG/OPT scheme is an iterative method. One cycle of this method is defined as follows. Let x_k^0 be the starting approximation at resolution k.

MG/OPT (k)

If $k = 1$ (coarsest resolution) solve (1) exactly.

Else if $k > 1$:

1. Pre-optimization. Define $x_k^1 = S_k(x_k^0)$.

2. Setup and solve a coarse-grid minimization problem. Define $x_{k-1}^1 = I_k^{k-1}x_k^1$ and $\tau_{k-1} = \nabla f_{k-1}(x_{k-1}^1) - I_k^{k-1}\nabla f_k(x_k^1)$. The coarse-grid minimization problem is given by

$$\min_{x_{k-1}} \left(f_{k-1}(x_{k-1}) - \tau_{k-1}^T\, x_{k-1}\right). \tag{2}$$

   Apply one cycle of MG/OPT(k-1) to (2) to obtain x_{k-1}^2.

* Corresponding author: e-mail: alfio.borzi@uni-graz.at, Phone: +43 316 380 5166, Fax: +43 316 380 9815

3. Line-search and coarse-grid correction. Perform a line search in the $I_{k-1}^{k}(x_{k-1}^{2}-x_{k-1}^{1})$ direction to obtain α_k. The coarse-grid correction is given by
$$x_k^2 = x_k^1 + \alpha_k\, I_{k-1}^{k}(x_{k-1}^{2}-x_{k-1}^{1})$$
4. Post-optimization. Define $x_k^3 = S_k(x_k^2)$.

Roughly speaking, the essential guideline for constructing f_k on coarse levels is that it must sufficiently well approximate the convexity properties of the functional at finest resolution. This property together with the following

$$\nabla\left(f_{k-1}(x_{k-1}) - \tau_{k-1}^{T}\, x_{k-1}\right)|_{x_{k-1}^{1}} = I_k^{k-1}\nabla f_k(x_k^1),$$

give an insight to the fact that the coarse-grid correction provides a descending direction.

3 Convergence of the MG/OPT method

Assume that for each k, f_k is twice Frechét differentiable and $\nabla^2 f_k$ is positive definite and satisfies the 'ellipticity' condition $(\nabla^2 f_k(x)y,y)_k \geq \beta\|y\|_k^2$ together with $\|\nabla^2 f_k(x) - \nabla^2 f_k(y)\| \leq \lambda\,\|x-y\|_k$ uniformly for some positive constants β and λ.

The discussion that follows is based on the following lemma [3].

Lemma 3.1 For $v,x,y \in V_k$ assume $(\nabla f_k(x),y)_k \leq 0$ and let γ be such that

$$0 \leq \gamma \leq -2\delta(\nabla f_k(x),y)_k \left[\int_0^1 (\nabla^2 f_k(x+t\gamma y)y,y)_k dt\right]^{-1} \qquad \text{for some } \delta \in [0,1].$$

Then

$$-(1-\delta)\gamma(\nabla f_k(x),y)_k \leq f_k(x) - f_k(x+\gamma y) \leq -\gamma(\nabla f_k(x),y)_k. \tag{3}$$

The next lemma provides an explicit estimate for the step-length α_k for the coarse-grid correction.

Lemma 3.2 For $v,x,y \in V_k$ assume $(\nabla f_k(x),y)_k \leq 0$ and let

$$\alpha(x,y) = \min\{2, \frac{-(\nabla f_k(x),y)_k}{(\nabla^2 f_k(x)y,y)_k + \lambda\,\|y\|_k^3}\} \tag{4}$$

Then

$$0 \leq -\frac{1}{2}\alpha(x,y)(\nabla f_k(x),y)_k \leq f_k(x) - f_k(x+\alpha(x,y)y). \tag{5}$$

The following lemma states that the coarse-grid correction with step-length α given by Lemma 3.2 is a minimizing step.

Lemma 3.3 Take $x \in V_k$ and define $\tilde{x} = I_k^{k-1}x$. Denote with $\hat{f}_{k-1}(x_{k-1}) = f_{k-1}(x_{k-1}) - \tau_{k-1}^{T}\, x_{k-1}$ where $\tau_{k-1} = \nabla f_{k-1}(\tilde{x}) - I_k^{k-1}\nabla f_k(x)$. Let $\tilde{y} \in V_{k-1}$ be such that $\hat{f}_{k-1}(\tilde{y}) \leq \hat{f}_{k-1}(\tilde{x})$ and define $y = I_{k-1}^{k}(\tilde{y}-\tilde{x})$. Then

$$f_k(x+\alpha(x,y)y) - f_k(x) \leq \frac{1}{2}\alpha(x,y)(\nabla f_k(x),y)_k, \tag{6}$$

where $\alpha(x,y)$ is defined in Lemma 3.2 (strict inequality holds if $\hat{f}_{k-1}(\tilde{y}) < \hat{f}_{k-1}(\tilde{x})$).

The following theorem states convergence of the MG/OPT method.

Theorem 3.4 The MG/OPT method described above provides a minimizing iteration and if f_L is strictly convex then

$$\lim_{i\to\infty} \|x_L^i - q_L\|_L = 0,$$

where $f_L(q) = \min_{x_L} f_L(x_L)$ and i is the MG/OPT cycle index.

It is clear that for optimization problems with an underlying geometrical and/or differential structure, approaches similar to that of geometrical multigrid methods applied to PDE problems can be applied for the construction of the coarse f_{k-1} functional.

References

[1] A. Brandt, Multi-level adaptive solutions to boundary-value problems, Math. Comp., **31**, 333–390 (1977).
[2] W. Hackbusch, Multi-grid Methods and Applications (Springer-Verlag, New York, 1985).
[3] W. Hackbusch and A. Reusken, Analysis of a damped nonlinear multilevel method, Numer. Math., **55**, 225–246 (1989).
[4] R.M. Lewis and S.G. Nash, Model problems for the multigrid optimization of systems governed by differential equations, to appear in SIAM Journal on Scientific Computing.
[5] S.G. Nash, A multigrid approach to discretized optimization problems, Optimization Methods and Software, **14**, 99–116 (2000).