Multilevel limited memory BFGS with application to Data Assimilation

Serge Gratton[1] Philippe Toint[2]

[1]CERFACS, France

[2]Department of Mathematics,University of Namur, Belgium
(philippe.toint@fundp.ac.be)

August 2007

Multilevel limited memory BFGS with application to Data Assimilation

Serge Gratton[1]   Philippe Toint[2]

[1]CERFACS, France

[2]Department of Mathematics,University of Namur, Belgium
(philippe.toint@fundp.ac.be)

August 2007

Outline

Motivation

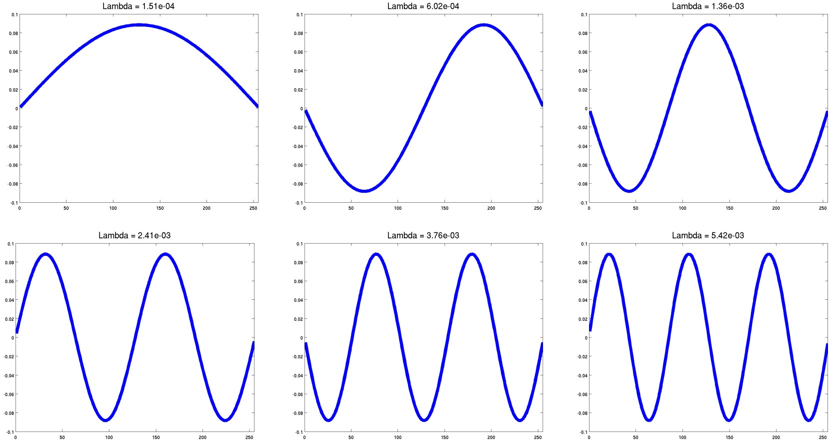
- PDE based optimization problems are present in a number of applications: shape optimization, Data Assimilation, control problems.
- Recent optimization methods have been designed to cope with these problems, including multigrid algorithms.
- These algorithms involve the computation of a hierarchy of discretization or Galerkin models, involving large algorithmic modification if the application was not originally designed for multigrid
- Idea: propose convergence improvements that take into account the grid structure and that does not imply rewriting the complete applications.

Outline

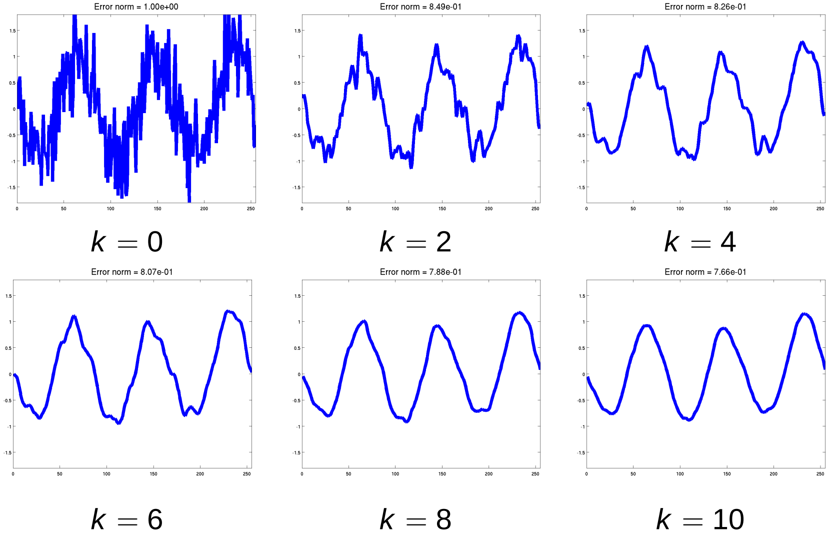
Linear systems arising from PDE

- Partial differential equations arise in many situations of scientific engineering. In particular the Laplace operator appear in many fields. The simple situation is the heat equation
$$\begin{cases} \frac{\partial u}{\partial t} - \Delta u = 0 \text{ in } \Omega, \\ +\text{boundary conditions on } \partial\Omega. \end{cases}$$
- In 1D with Dirichlet boundary conditions the steady state describes the heat distribution along a wire of length 1.
- Compute a finite difference discretization $Ax = b \Longleftrightarrow \min_x \frac{1}{2}x^T Ax - x^T b$
- Set $\theta_k = \frac{k\pi}{n+1}$. The eigenvalues of A are $\lambda_k = 2(1 - \cos\theta_k) = 4\sin^2\frac{\theta_k}{2}, \;\; k = 1, \cdots, n,$
- the associated eigenvectors are $v_k = \begin{pmatrix} sin\theta_k \\ sin(2\theta_k) \\ \vdots \\ sin(n\theta_k) \end{pmatrix}.$

Shape of the eigenvectors



Error at step k of CG



Slow convergence of the smooth modes

Outline

Basic idea

- Convergence of CG methods on model problems for multigrid:
  - The high frequency error component converge fast with many traditional iterative methods.
  - Use a coarse grid to compute the low frequency on a coarser grid.
  - Problem : discretization or Galerkin approximation needed at all considered level. May be problematic if the application was not original designed for multigrid.
- Use a preconditioning approach to obtain improve the smooth error components:
  - P and R are the prolongation and restriction operators.
  - The operator PR implements a grid-based smoothing.
  - Consider approximate secant information contained in $PRs = PR(x_{k+1} - x_k)$, and $PRy = PR(\nabla f(x_{k+1}) - \nabla f(x_k))$.

Approximate secant equations

- Secant approach $y \sim Hs$.

SECSM Smoothed y and s : $H\,PRs \sim PRy$

- Eigenvalue approach

EIG Use y for curvature information only: $H\,PRs \sim \theta_1 s$, $\theta_1 = \frac{y^T PRs}{\|PRs\|_2^2}$

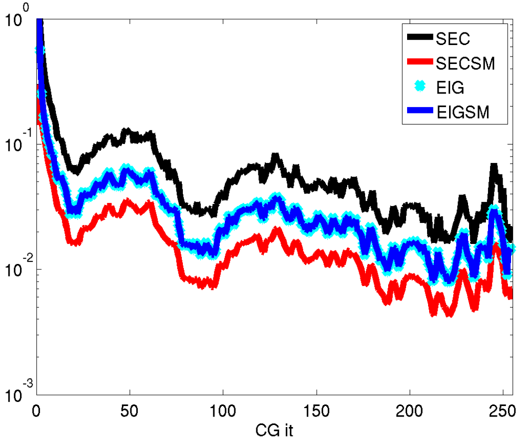
EIGSM Use smooth y for curvature information only: $H\,PRs \sim \theta_1 s$, $\theta_2 = \frac{(PRy)^T PRs}{\|PRs\|_2^2}$

- The approach is generalized to multilevel by replacing PR by $P_r P_{r-1} \ldots (P_j R_j) \ldots R_{r-1} R_r$ in the above expression.

For each secant approximation $H\tilde{s} = \tilde{y}$, we introduce the normalized error $\frac{\|H\tilde{s} - \tilde{y}\|}{\|H\|\|\tilde{s}\|}$.

Normalized error for the model problem

- We consider unpreconditioned CG (or L-BFGS) on our quadratic model problem, and plot the error corresponding to the 4 strategies.



- We include these additional information in a L-BFGS algorithm

Memory management

1. Target problems are large scale unconstrained problems. As for L-BFGS a maximum number of vectors to be stored is set: MEM.
2. Two strategies are considered:
   - EQ The memory contains the most recent coarse and fine secant information, with no a priori preference.
   - PF Only the secant information related to the current step is kept

Outline

Outline

Dirichlet-to-Neumann transfer DN

- It consists [Lewis,Nash,04] in finding the function $a(x)$ defined on $[0, \pi]$, that minimizes

$$\int_0^{\pi} \left(\partial_y u(x,0) - \phi(x)\right)^2 dx,$$

where $\partial_y u$ is the partial derivative of u with respect to y,
- and where u is the solution of the boundary value problem

$$\begin{array}{rcll} \Delta u & = & 0 & \text{in } S, \\ u(x,y) & = & a(x) & \text{on } \Gamma, \\ u(x,y) & = & 0 & \text{on } \partial S \backslash \Gamma. \end{array}$$

Multigrid model problem MG

- Consider here the two-dimensional model problem for multigrid solvers in the unit square domain S_2

$$\begin{aligned} -\Delta u(x,y) &= f \text{ in } S_2 \\ u(x,y) &= 0 \text{ on } \partial S_2, \end{aligned}$$

- f is such that the analytical solution to this problem is $u(x,y) = 2y(1-y) + 2x(1-x)$.
- This problem is discretized using a 5-point finite-difference scheme
- Our algorithm will be used on the variational minimization problem

$$\min_{x \in R^{n_r}} \frac{1}{2} x^T A_r x - x^T b_r,$$

Outline

The 4D-Var functional

- Consider a dynamical system $\dot{x} = f(t, x)$ with solution operator $x(t) = \mathcal{M}(t, x_0)$.
- Observations y_i at time t_i modeled by $y_i = \mathcal{H}x(t_i) + \epsilon$, where ϵ is a Gaussian noise with covariance matrix R_i.
- The a priori error error covariance matrix on x_0 is B.
- In data assimilation, we are looking for x_0 that minimizes

$$\frac{1}{2}\|x_0 - x_b\|^2_{B^{-1}} + \frac{1}{2}\sum_{i=0}^{N}\|\mathcal{H}\mathcal{M}(t, x_0) - y_i\|^2_{R_i^{-1}},$$

- The first term in the cost function is the background term, the second term is the observation term.

Data assimilation in the heat equation DA

- In this problem, the dynamical system is the heat equation in 2D, defined by

$$\left\{ \frac{\partial u}{\partial t} - \Delta u = 0 \text{ in } \Omega = [0,1] \times [0,1] \right.$$

- No a priori term is considered
- Observations: the state $x(t)$ is observed (i.e. $\mathcal{H}$ is a selection matrix) at every other point in spatial domain and at every time step.
- We perform twin experiments i.e. the observations are computed by
  - imposing a solution $x_0^\star$
  - by computing the system trajectory
  - and adding a Gaussian noise.

The shallow water system SW

- The shallow system is often considered as a good approximation of the dynamical systems used in ocean modeling.
- It is based on the Shallow Water equations

$$\begin{cases} \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial v}{\partial y} - fv + g\frac{\partial z}{\partial x} = \lambda\Delta u \\ \frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + g\frac{\partial z}{\partial y} = \lambda\Delta v \\ \frac{\partial z}{\partial t} + u\frac{\partial z}{\partial x} + v\frac{\partial z}{\partial y} + z\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = \lambda\Delta z \end{cases}$$

- Observations of the state exist at everything 5 points in the physical domain and every 5 time steps
- The a priori term is modeled using a diffusion operator as suggested in [Weaver, Courtier, 2001]
- The system is time integrated using a leapfrog scheme.
- The damping in $\lambda\Delta$ is implemented to improve the smoothness of the solution in the spatial domain

Outline

Algorithmic parameters

Pb	size	LMEM	conv.tol.	nblevels
DN	251	10	10^{-5}	7
MG	63^2	8	10^{-5}	5
DA	63^2	8	10^{-5}	5
SW	31^2	15	10^{-5}	3

We use linear interpolation for P and $R = \sigma P^T$, with $\sigma = 1/\|P\|$,

Problem DN

		cost	grad	it
LBFS	–	353	177	176
EQ	SECSM	536	268	267
EQ	EIG	243	107	106
EQ	EIGSM	303	148	146
PF	SECSM	309	155	154
PF	EIG	137	67	66
PF	EIGSM	151	75	73

Problem MG

		cost	grad	it
LBFS	–	222	128	127
EQ	SECSM	189	96	95
EQ	EIG	146	75	74
EQ	EIGSM	177	98	97
PF	SECSM	135	77	76
PF	EIG	166	89	88
PF	EIGSM	121	69	68

Data assimilation in the heat equation DA

		cost	grad	it
LBFS	–	14564	2751	2706
EQ	SECSM	3622	703	655
EQ	EIG	2972	581	532
EQ	EIGSM	2804	548	503
PF	SECSM	3267	631	598
PF	EIG	5599	1080	1034
PF	EIGSM	5796	1112	1071

The shallow water system SW

		cost	grad	it
LBFS	–	101	51	50
EQ	SECSM	107	54	53
EQ	EIG	109	55	54
EQ	EIGSM	111	56	55
PF	SECSM	101	51	50
PF	EIG	87	44	43
PF	EIGSM	89	45	44

Outline

Conclusion

- For problems where multigrid is expected to perform well (elliptic PDE based optimization), exploit the underlying structure.
- Use of smoothed approximate secant equations.
- Multiple variant with different memory costs.
- Encouraging results on models problems and on a more realistic data assimilation problem involving Shallow Water system of equations.