Preconditioner updates for sequences of symmetric positive definite linear systems arising in optimization

Stefania Bellavia [+] ,
Valentina De Simone[*], Daniela di Serafino[*], Benedetta Morini[+]

[+] Università degli Studi di Firenze
[*] Seconda Università degli Studi di Napoli

SC2011
October 10-14, 2011

The problem

- Consider the sequence of linear systems

$$(A + \Delta_k)x = b_k$$

where $A \in \Re^{n \times n}$ is large, sparse and positive definite (SPD), Δ_k is diagonal positive semidefinite.

The problem

- Consider the sequence of linear systems

$$(A + \Delta_k)x = b_k$$

where $A \in \Re^{n \times n}$ is large, sparse and positive definite (SPD), Δ_k is diagonal positive semidefinite.

- Special case: Shifted linear systems

$$(A + \alpha_k I)x = b_k \quad \alpha_k > 0$$

Applications in constrained optimization

- Affine scaling methods for convex bound constrained QP problems and bound constrained linear least squares require the solution of sequences of linear systems of the form:

$$(M_k Q M_k + D_k)s = b_k, \quad k = 0, 1, \ldots$$

where Q is the Hessian of the quadratic function, M_k is diagonal SPD and D_k is diagonal positive semidefinite.

[Coleman, Li 1996],[Bellavia, Macconi, Morini, 2006]

Applications in unconstrained optimization

Consider an unconstrained nonlinear least-squares problem

$$\min_{x \in \Re^n} \|F(x)\|_2^2, \quad F : \Re^n \to \in \Re^m$$

Computation of the step in elliptical trust-region methods:

$$\underset{p}{\text{minimize}}\, m(p) = \frac{1}{2}\|F + Jp\|_2^2, \quad \|Gp\|_2 \leq \Delta$$

where G is diagonal SPD, $J \in \Re^{m \times n}$ is the Jacobian of F, $\Delta > 0$.

Applications in unconstrained optimization

Consider an unconstrained nonlinear least-squares problem

$$\min_{x \in \Re^n} \|F(x)\|_2^2, \quad F : \Re^n \to\in \Re^m$$

Computation of the step in elliptical trust-region methods:

$$\underset{p}{\text{minimize}}\, m(p) = \frac{1}{2}\|F + Jp\|_2^2, \quad \|Gp\|_2 \leq \Delta$$

where G is diagonal SPD, $J \in \Re^{m \times n}$ is the Jacobian of F, $\Delta > 0$.

- For a certain $\lambda \geq 0$, the minimizer $p = p(\lambda)$ satisfies

$$(J^T J + \lambda G)p(\lambda) = -J^T F,$$

- If $\lambda > 0$, it solves a scalar nonlinear secular equation. A root finding method applied to the secular equation gives rise to a sequence of linear systems of the above form.

Applications in unconstrained optimization

- Recent regularization approaches [Nesterov, 2007; Cartis, Gould, Toint, 2009, 2010; Bellavia, Cartis, Gould, Morini, Toint, 2010]:

$$\underset{p}{\text{minimize}}\, m(p) = \|F + Jp\|_2 + \frac{1}{2}\sigma||p||_2^2,$$

$$\underset{p}{\text{minimize}}\, m(p) = \frac{1}{2}\|F + Jp\|_2^2 + \frac{1}{3}\sigma||p||_2^3,$$

  where $\sigma > 0$

- For a certain $\lambda > 0$, the minimizer $p = p(\lambda)$ satisfies

$$(J^T J + \lambda I)p(\lambda) = -J^T F.$$

  The computation of p calls for the solution of a sequence of shifted linear systems.

Preconditioning sequences of matrices

- Freezing the preconditioner often leads to slow convergence.
- Recomputing the preconditioner from scratch for each matrix is costly and pointlessly accurate.
- Updating strategies derive preconditioners from previous systems of the sequence in a cheap way.

Updating strategies

- Given a preconditioner for a specific matrix of the sequence (seed preconditioner), updating strategies update it in order to build a preconditioner for subsequent matrices of the sequence at a low computational cost.
- Minimum requirement: Inexpensive updates must have the ability to precondition sequences of slowly varying systems.
- Expected behaviour in terms of linear solver iterations: to be in between the the frozen and the recomputed preconditioner.

Existing approaches

- Sequences $A + \Delta_k$ based on incomplete factors of A^{-1}:
  [Benzi, Bertaccini, 2003],[Bertaccini, 2004]
- Sequences $A + \alpha_k I$ based on incomplete LDL^T factorization of A:
  [Meurant, 2001], [Bellavia, De Simone, di Serafino, Morini, 2011].
- Sequences of matrices differing for general matrices:
  [Morales-Nocedal 2000], [Bergamaschi, Bru, Martinez, Putti 2006], [Tebbens, Tuma, 2007, 2010], [Calgaro, Chehab, Saad, 2010], [Bellavia, Bertaccini, Morini, 2011].

Approaches based on LDL^T preconditioners, $\Delta_k = \alpha_k I$

[Bellavia, De Simone, di Serafino, Morini, 2011, Meurant 2001]

Let

$$A = LDL^T,$$

where L is unit lower triangular and $D = diag(d_1, \ldots, d_n)$.

Approaches based on LDL^T preconditioners, $\Delta_k = \alpha_k I$

[Bellavia, De Simone, di Serafino, Morini, 2011, Meurant 2001]

Let

$$A = LDL^T,$$

where L is unit lower triangular and $D = diag(d_1, \ldots, d_n)$.

A preconditioner P for matrix $A + \alpha_k I$ has the form

$$P = \tilde{L}\tilde{D}\tilde{L}^T,$$

with $\tilde{L}$ unit lower triangular and $\tilde{D} = diag(\tilde{d}_1, \ldots, \tilde{d}_n)$

- $\tilde{D} = D + \alpha_k I$;
- $off(\tilde{L}) = off(L)S$, with $S = D\tilde{D}^{-1}$. Column j of off(L) is scaled by the factor $d_j/\tilde{d}_j \in (0, 1)$.

- The update computational overhead is low.

- Given the Cholesky factorization of A, $P = \tilde{L}\tilde{D}\tilde{L}^T$ can be derived as an order 0 asymptotic expansions in terms of α of the Cholesky factor of $A + \alpha I$, [Meurant 2001].

- P is effective for a broad range of values of α.
  For small and large values of α the eigenvalues of $P^{-1}(A + \alpha I)$ are clustered in a neighbourhood of 1, [Bellavia, De Simone, di Serafino, Morini, 2011].
- Incomplete LDL^T factorizations of A can be used.

Updating factorization framework for $A + \Delta_k$

Let $A = LDL^T$ where L is unit lower triangular and $D = diag(d_1, \ldots, d_n)$.

UF (Updating Factorization) framework:

A preconditioner P for matrix $A + \Delta_k$ has the form

$$P = \tilde{L}\tilde{D}\tilde{L}^T,$$

- $\tilde{D} = diag(\tilde{d}_1, \ldots, \tilde{d}_n)$, $\tilde{d}_i \geq d_i$.
- $\|\tilde{D} - D\| \leq \tau \|\Delta_k\|$, for some $\tau > 0$.
- $\tilde{L}$ unit lower triangular, $off(\tilde{L}) = off(L)S$, with $S = D\tilde{D}^{-1}$.

- P is SPD.
- $\tilde{L}$ has the same sparsity pattern as L.

Slowly varying sequences of matrices

Theorem

Let P be an UF preconditioner for matrix $A+\Delta_k$. Then, for some positive ζ:

$$\|A+\Delta_k-P\| \leq \zeta\|\Delta_k\|.$$

Corollary

For $\|\Delta_k\|$ small enough, the eigenvalues of $P^{-1}(A+\Delta_k)$ are clustered in a neighbourhood of 1.

Preconditioner UF1

A practical preconditioner in the UF framework is obtained generalizing the preconditioner for shifted matrices in [Bellavia, De Simone, di Serafino, Morini, 2011, Meurant 2001].

Let

$$P = \tilde{L}\tilde{D}\tilde{L}^T$$

- $\tilde{D} = D + \Delta_k$.
- $\tilde{L}$ unit lower triangular, $off(\tilde{L}) = off(L)S$ with $S = D\tilde{D}^{-1}$.

The update computational overhead is low.

Preconditioner UF2

Fix $\tilde{D}$ so that $diag(P) = diag(A + \Delta_k)$.

Let

$$P = \tilde{L}\tilde{D}\tilde{L}^T$$

- $\tilde{d}_i = d_i + \delta_{k,i} + \sum_{j=1}^{i-1} l_{i,j}^2 (d_j - s_j^2 \tilde{d}_j)$
- $\tilde{L}$ unit lower triangular, $off(\tilde{L}) = off(L)S$ with $S = D\tilde{D}^{-1}$.

Unlike UF1 preconditioner, the computation of $\tilde{D}$ appears to be serial

Analysis of the preconditioners

- Let P be computed by the UF1 approach, then

$$\begin{aligned} \|A+\Delta_k-P\| &\le 2\|off(L)D(D+\Delta_k)^{-1}\Delta_k off(L)^T\| \\ &\le 4\|off(L)\|^2\|D\| \end{aligned}$$

$$\|diag(A+\Delta_k-P)\| \neq 0, \qquad \|off(A+\Delta_k-P)\| \neq 0$$

- Let P be computed by the UF2 approach, then

$$\begin{aligned} \|A+\Delta_k-P\| &\le 2\|off(off(L)S(\tilde{D}-D)off(L)^T)\| \\ &\le 2\|off(L)\|^2\|D\| \end{aligned}$$

$$\|diag(A+\Delta_k-P)\| = 0$$

$\|\Delta_k\|$ large

Let P be computed by the UF1 or UF2 approach.

Let ϵ be a small positive integer. Then for $\|\Delta_k\|$ sufficiently large,

$$\frac{\|A + \Delta_k - P\|}{\|A + \Delta_k\|} \leq \epsilon.$$

$\|\Delta_k\|$ large

Let P be computed by the UF1 or UF2 approach.

Let ϵ be a small positive integer. Then for $\|\Delta_k\|$ sufficiently large,

$$\frac{\|A + \Delta_k - P\|}{\|A + \Delta_k\|} \leq \epsilon.$$

Further, if Δ_k is SPD and and $\|\Delta_k^{-1}\|$ is sufficiently small, the eigenvalues of $P^{-1}(A + \Delta_k)$ are clustered in a neighbourhood of 1.

Practical case: $A \approx LDL^T$

- The quality of P depends on the quality of the seed preconditioner;
- A term depending on $\|A - LDL^T\|$ must be added to the upper bound on $\|A + \Delta_k - P\|$.
- The property of UF2 preconditioner

$$diag(P) = diag(A + \Delta_k)$$

is not longer valid but the discrepancy between the two diagonal depends on the error $diag(A - LDL^T)$:

$$diag(A + \Delta_k - P) = diag(A - LDL^T)$$

- The construction of both UF1 and UF2 does not break down.

Set1: Quadprog

- The Matlab function `Quadprog` available in the Matlab Optimization Toolbox implements the reflective Newton method for bound constrained QP problems:

$$min_x\{q(x) = \frac{1}{2}x^TQx + c^Tx : \quad l \leq x \leq u\}$$

Assume that QP is convex, $Q \in \Re^{n\times n}$ is symmetric positive semidefinite, $c \in \Re^n$, $l \in \{\Re \cup \{\infty\}\}^n$ and $u \in \{\Re \cup \{\infty\}\}^n$, $l < u$.

[Coleman, Li 1996].

- Quadprog generates a strictly feasible sequence $\{x_k\}$ and amounts to solve a sequence of linear systems of the following form:

$$\underbrace{(M_k Q M_k + D_k)}_{H_k} s = -M_k g(x_k), \quad k = 0, 1, \ldots$$

  where $g(x_k) = \nabla q(x_k) = Q x_k + c$, M_k is diagonal SPD and D_k^g is diagonal positive semidefinite.

- Preconditioned CG is employed to solve such linear systems

Preconditioners available in Quadprog

- Default preconditioner: DIAG:

$$P_{D,k} = diag\left(\|H_k(:,1)\|_2, \ldots, \|H_k(:,n)\|_2\right),$$

where $H_k(:,j)$ denotes the j-th column of H_k.

- Optional Preconditioner: TRID, Tridiagonal preconditioner, Cholesky factors of

$$\bar{H} = tril(triu(H_k, -1), 1),$$

computed using the Matlab built-in function `chol`. If $\bar{H}$ is not positive definite, a shift is applied and a new Cholesky factorization is attempted.

UF1 and UF2 in Quadprog

- Our updating procedures can be employed in quadprog to solve the sequences of linear systems

$$\underbrace{(M_k Q M_k + D_k)}_{H_k} s = -M_k g(x_k), \quad k = 0, 1, \ldots$$

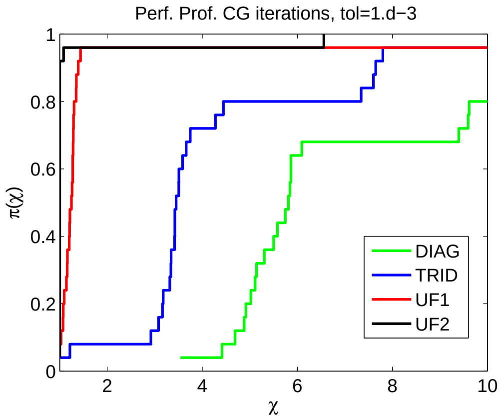
- Compute an incomplete $R^T R$ factorization of Q.
- The $R^T R$ factorization provides, for any k an incomplete LDL^T factorization of $M_k Q M_k$ given by $M_k R^T R M_k$.
- Then, applying UF1 or UF2 we obtain an $\tilde{L}\tilde{D}\tilde{L}^T$ preconditioner for $M_k Q M_k + D_k$.

Testing details

- Computational environment: `Intel Core 2 DUO U9600, 1.60 GHz, 3GB RAM, Matlab version 7.7`
- We compare the performance of UF1 and UF2 against `DIAG` and `TRID` within `Quadprog`
- Test set: strictly convex bound constrained QP of dimension $n > 500$ available in the CUTEr collection
- Matlab cholinc function to compute the incomplete $R^T R$ factorization of Q; drop tolerance=10^{-2}
- UF1 and UF2 have been implemented as mex-files with Matlab interface.
- Default stopping tolerance for the stopping criterions of `Quadprog`
- Stopping tolerance for `PCG` : `cg_tol`=10^{-3}.

Performance profile: total number of CG iterations

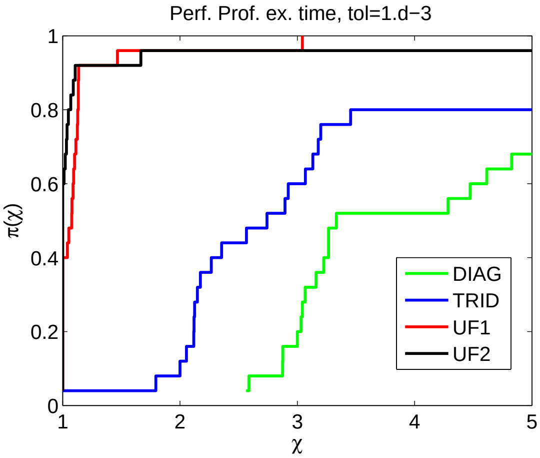
$\pi(\chi)$: Fraction of runs for which the preconditioner is within a factor χ of the best



All tests succesfully solved

The number of nonlinear iterations is not affected by the preconditioner.

Performance profiles: execution time



Execution time: time devoted to the linear algebra phase

Set 2: 8 sequences of shifted linear systems

Four systems of nonlinear equations of dimension $n = 10^4$ were solved by the RER algorithm [Bellavia, Cartis, Gould, Morini & Toint, 2010]

- Sequences of shifted systems arising in the first and second nonlinear iterations of RER; $\alpha \in (6.3195 \cdot 10^{-5},\ 58.4277)$

Set 2: 8 sequences of shifted linear systems

Four systems of nonlinear equations of dimension $n = 10^4$ were solved by the RER algorithm [Bellavia, Cartis, Gould, Morini & Toint, 2010]

- Sequences of shifted systems arising in the first and second nonlinear iterations of RER; $\alpha \in (6.3195 \cdot 10^{-5},\ 58.4277)$

UF1 and UF2 are compared with **NP:** no prec.; **RP**: prec. recomputed for each α; **FP:** fixed prec..

Set 2: 8 sequences of shifted linear systems

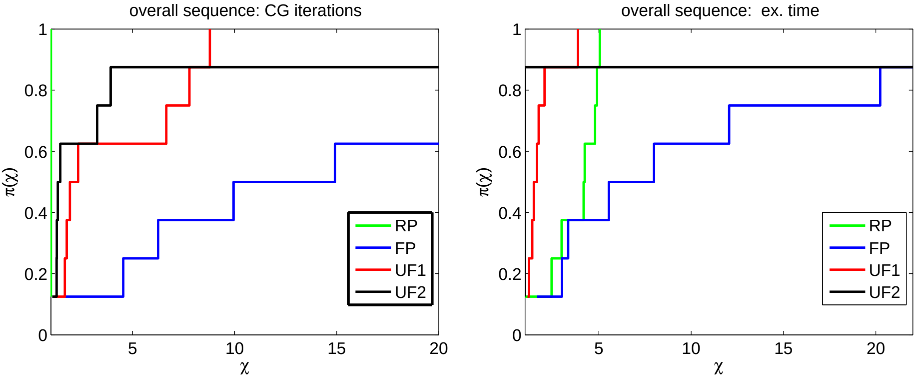
Four systems of nonlinear equations of dimension $n = 10^4$ were solved by the RER algorithm [Bellavia, Cartis, Gould, Morini & Toint, 2010]

- Sequences of shifted systems arising in the first and second nonlinear iterations of RER; $\alpha \in (6.3195 \cdot 10^{-5},\ 58.4277)$

UF1 and UF2 are compared with **NP:** no prec.; **RP**: prec. recomputed for each α; **FP:** fixed prec..

- Matlab pcg function with $tol = 10^{-6}$ and $maxit = 1000$;
- Matlab cholinc function to compute the incomplete LDL^T factorization; drop tolerance fixed by trial on the system $Ax = b$;

Test set 2: 8 sequences, all values of α



NP always fails in solving the first system of each sequence

FP and UF2 fail in solving one sequence

Conclusion

Given $A \approx LDL^T$, the update techniques:

1. preserve the sparsity pattern of the factor L.
2. are breakdown-free
3. do not need algorithmic parameters.
4. seem to be effective for a broad range of values of Δ_k (automatic adaptation to the size of the entries of Δ_k);

Further, preserving the diagonal of $A + \Delta_k$ gives a significant improvement in terms of CG iterations.

Many thanks for your attention!