Inexact Ideas and Asynchronous Methods

Trond Steihaug
Department of Informatics
University of Bergen
Norway

With contribution from
Erik Spildo
Baryamureeba Venansius
Yasemin Yalcinkaya

Fifth SIAM Conference on Optimization, Victoria May 20-22, 1996

This work is partially supported by NFR and VISTA - a research cooperation between the Norwegian Academy of Science and Letters and Den norske stats oljeselskap a/s (Statoil)

Overview

- Inexact v.s. Truncated
  - The simplex Method
- Inexact Interior point Methods
  - Computing the step
  - An LSQ formulation
- The nonlinear LSQ problem
  - The Gauss-Newton algorithm
  - Small residual or mildly nonlinear problems
  - Definition of mildly nonlinear
  - Inexact approach
- The linear LSQ problem
  - Block methods
  - Asynchronous implementations

Truncated v.s. Inexact

The efficiency of Newtons method using an iterative method to solve the inner (linear) system can be improved by balancing the the accuracy of the inner termination with properties (nonlinearity) of the outer iteration.

- Truncated method (explicit use of the inner iteration)
- Inexact method (only the approximate properties)

Inexact Newton method

Problem: Find x^* s.t. $F(x^*) = 0$ where $F : \mathbf{R}^n \to \mathbf{R}^n$.

$k = 0$
`While not` *converged* `do`
  Find *approximate* solution $\Delta\mathbf{x}^k$ of
    $F'(\mathbf{x}^k)\Delta\mathbf{x}^k = -F(\mathbf{x}^k)$
  $\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta\mathbf{x}^k$
  $k \leftarrow k+1$
`Continue`

The Simplex Method

Let $c \in \mathbf{R}^n, \; b \in \mathbf{R}^m, A \in \mathbf{R}^{m \times n}$

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

The Simplex method:

$$A = [B_k \; N_k], \quad x_k^T = (x_k^B, x_k^N)^T$$

where

$$B_k x_k^B = b, \; x_k^N = 0$$

Need to identify right variables to enter and leave the basis. The accuracy needed in the solution of the linear system depends on identification of the active set and computing the reduced prices.

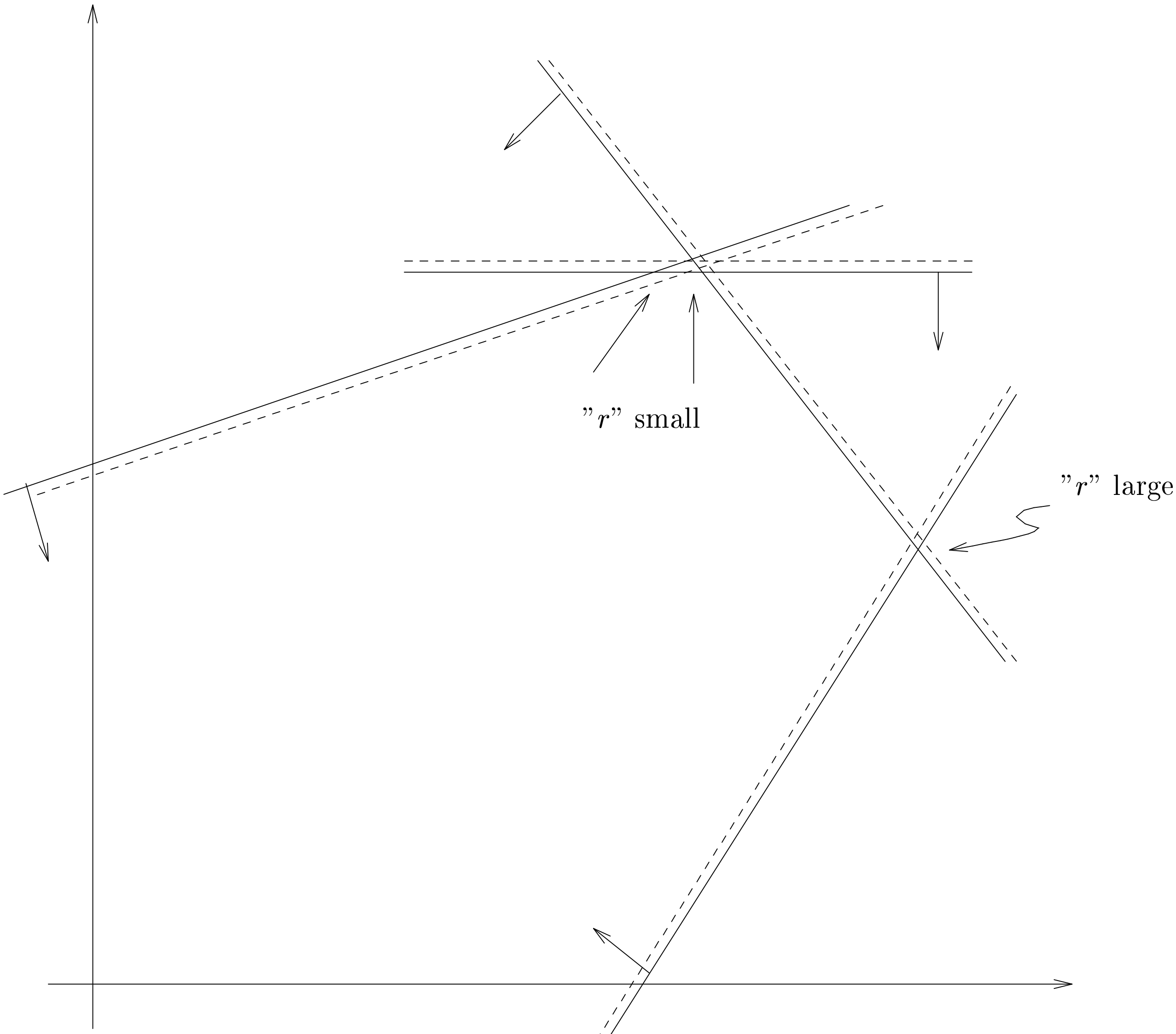
?????

Can we expect to find an inexact version of the Simplex method with apriori tolerances?

?????

Error in the primal-equation

$$Ax = b + r$$



Inexact Primal-Dual Interior Point Methods

`While not` *converged* `do`

$X = diag(x_1, \ldots, x_n)$ $Z = diag(z_1, \ldots, z_n)$

Find approximate solution

$$\begin{pmatrix} X^{-1}Z & A^T \\ A & \end{pmatrix} \begin{pmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{c} - A^T\mathbf{y} - \mu X^{-1}\mathbf{e} \\ A\mathbf{x} - \mathbf{b} \end{pmatrix} + \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{pmatrix},$$

$$\Delta\mathbf{z} = X^{-1}(\mu\mathbf{e} + Z\Delta\mathbf{x}) - \mathbf{z}$$

$$\alpha_x = \alpha \min\{|x_i/\Delta x_i| : \Delta x_i > 0,\ 1\}$$
$$\alpha_z = \alpha \min\{|z_i/\Delta z_i| : \Delta z_i < 0,\ 1\}$$

$$\mathbf{z}^+ = \mathbf{z} + \alpha_z\Delta\mathbf{z}$$
$$\mathbf{x}^+ = \mathbf{x} - \alpha_x\Delta\mathbf{x}$$
$$\mathbf{y}^+ = \mathbf{y} + \alpha_z\Delta\mathbf{y}$$
$$\mu^+ = \sigma(\mathbf{x}^+)^T\mathbf{z}^+$$

`Continue`

Inexact PD IPM

Find an approximate solution to the system

$$\begin{pmatrix} X^{-1}Z & A^T \\ A & \end{pmatrix} \begin{pmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{c} - A^T\mathbf{y} - \mu X^{-1}\mathbf{e} \\ A\mathbf{x} - \mathbf{b} \end{pmatrix} + \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{pmatrix}$$

such that

$$\left\| \begin{matrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{matrix} \right\|_2 \leq \eta \left\| \begin{matrix} A^T\mathbf{y} + \mathbf{z} - \mathbf{c} \\ A\mathbf{x} - \mathbf{b} \end{matrix} \right\|_2$$

Residuals or error vectors:

$$\begin{aligned} \mathbf{r}_1 &= A^T(\mathbf{y} + \Delta\mathbf{y}) + (\mathbf{z} + \Delta\mathbf{z}) - \mathbf{c} + \mu X^{-1}\mathbf{e} \\ \mathbf{r}_2 &= \mathbf{b} - A(\mathbf{x} - \Delta\mathbf{x}) \end{aligned}$$

Note that the residuals are (close to) the primal and dual infeasibility at the new point (except for the steplengths α_x and α_z)

Inexact PD IPM - Dual Feasibility

$$\begin{pmatrix} X^{-1}Z & A^T \\ A & \end{pmatrix}\begin{pmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{c} - A^T\mathbf{y} - \mu X^{-1}\mathbf{e} \\ A\mathbf{x} - \mathbf{b} \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbf{r}_2 \end{pmatrix}$$

Let $M = (X^{-1}Z)^{-1}$ then

$$AMA^T\Delta\mathbf{y} = AM(\mathbf{c} - A^T\mathbf{y} - \mu X^{-1}\mathbf{e}) + \mathbf{b} - A\mathbf{x} - \mathbf{r}_2$$

and

$$\Delta x = M(\mathbf{c} - A^T(\mathbf{y} + \Delta\mathbf{y}) - \mu X^{-1}\mathbf{e})$$

$$\|\mathbf{r}_2\| \leq \eta\|A\mathbf{x} - \mathbf{b}\|$$

From Ito, Kelley sand Sachs (1993) (wth minor modifications)
$\alpha < 1, \sigma$ sufficiently small

$$\eta < 1 \Rightarrow \text{q-linear convergence}$$

Extension of Tapia and Zhang (1992) to an inexact framework:
$\alpha \to 1, \sigma \to 0$

$$\eta \to 0 \Rightarrow \text{superlinear convergence}$$

Inexact PD IPM - Dual Feasibility (2)

Observation: x, y close to the solution and μ small (η small, but need not go to zero) then $\alpha_x \to 1$, $\alpha_z \to 1$

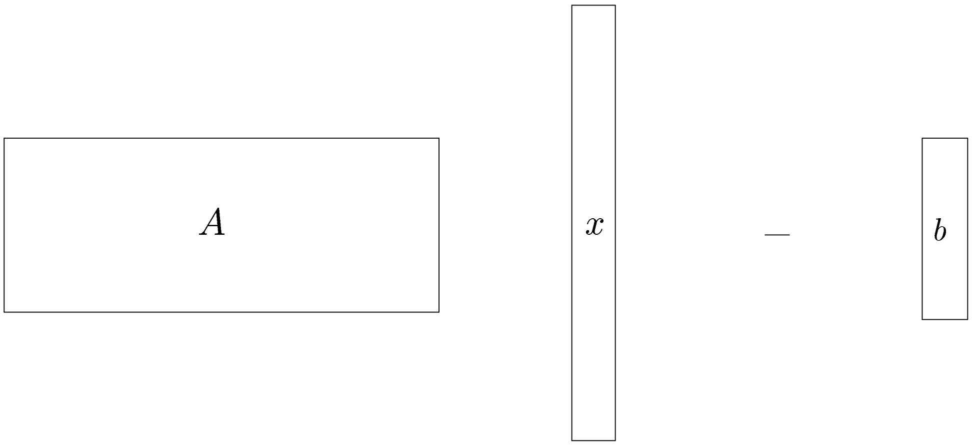
Observation Need to solve a linear LSQ problem:

$$\min \|A^T \Delta \mathbf{y} - \mathbf{d}\|_M$$

where

$$\mathbf{d} = \mathbf{c} - A^T \mathbf{y} - \mu X^{-1} \mathbf{e} + M^{-1}(\mathbf{x} - \hat{\mathbf{x}})$$

and $\hat{\mathbf{x}}$ is any feasible point $A\hat{\mathbf{x}} = \mathbf{b}$.

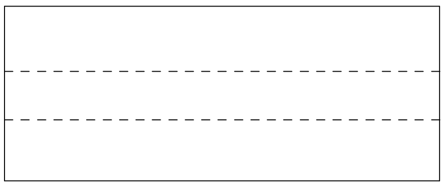


$$Ax - b$$

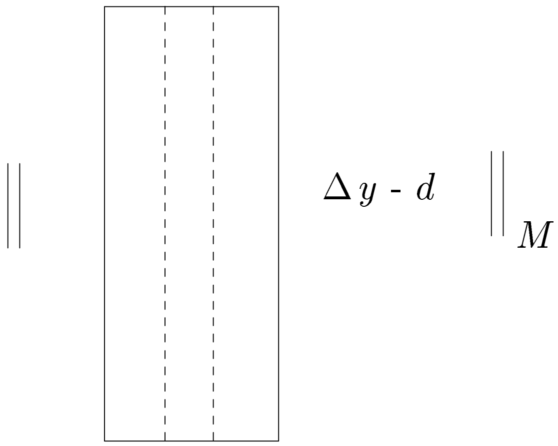
Rearrange the rows according to

$$\left| a_i x - b_i \right| \qquad i = 1, 2, .., m$$

a_i - row i in A



Linear LSQ problem



The Nonlinear Least Squares Problem

Given

$$F : \mathbf{R}^n \to \mathbf{R}^m, m \geq n, \text{ and } F = (F_1, ..., F_m)^T,$$

and define

$$\phi(x) = \frac{1}{2} F(x)^T F(x).$$

The nonlinear least squares problem is to find x^* so that for some norm and some $\epsilon > 0$,

$$\phi(x^*) \leq \phi(x) \text{ for all } ||x - x^*|| < \epsilon.$$

Gauss-Newton

Given x^0

FOR $k = 0$ STEP 1 UNTIL Convergence DO

Solve for s^k:

$$\min\{||F'(x^k)s + F(x^k)||_2 : s \in \mathbf{R}^n\}$$

Set $x^{k+1} = x^k + s^k$.

Convergence:

- Smoothness of F
- x^0 close to a solution x^*
- $F'(x^*)$ full rank
- Small residual or mildly nonlinear

Point of Attraction

Define

$$N(x) = x - [F'(x)^T F'(x)]^{-1} F'(x)^T F(x).$$

Gauss-Newton: $x^{k+1} = N(x^k)$, x^0 given.

A.1. F is twice continuously differentiable in an open neighborhood Ω of x^*;

A.2. $F'(x^*)^T F(x^*) = 0$;

A.3. $F'(x^*)$ has full rank;

A.4. x^* is a point of attraction (local convergence), i.e.

$$\rho(N'(x^*)) < 1$$

$\rho(\cdot)$ is the spectral radius,

$$N'(x^*) = -[F'(x^*)^T F'(x^*)]^{-1} S$$

$$S = \sum_{i=1}^{m} \nabla^2 F_i(x^*) F_i(x^*)$$

- Small residual ($\|F(x^*)\|$ small) $\Rightarrow \rho(N'(x^*))$ is small
- Each F_i is mildly nonlinear ($\|\nabla^2 F_i(x^*)\|$ is small) then $\rho(N'(x^*))$ is small

Small Residual or Mildly Nonlinear Problems

There exists $\gamma \geq 0$ so that for $x \in \Omega$,

$$\|(F'(x) - F'(x^*))^T F(x^*)\| \leq \gamma \|F'(x^*)^T (F(x) - F(x^*))\|$$

where $\|\cdot\|$ is any vectornorm.

$$\gamma < 1 \Rightarrow x^* \text{ is a point of attraction}$$

Let F be twice continuously differentiable in an open neighborhood Ω of x^*. If

$$\rho(N'(x^*)) \leq \gamma,$$

then there exists $\|\cdot\|$ and $\delta \geq 0$,

$$\bar{\gamma} = \gamma + \delta < 1$$

so that

$$\|(F'(x) - F'(x^*))^T F(x^*)\| \leq \bar{\gamma} \|F'(x^*)^T (F(x) - F(x^*))\|$$

For $x = x^* + \tau d$ consider:

$$\|(F'(x) - F'(x^*))^T F(x^*)\| \leq \gamma \|F'(x^*)^T (F(x) - F(x^*))\|$$

and let $\tau \to 0$ then

$$\forall d \in \mathbf{R}^n \quad \|Sd\| \leq \gamma \|F'(x^*)^T F'(x^*) d\|$$

$$S = \sum_{i=1}^{m} \nabla^2 F_i(x^*) F_i(x^*)$$

If $F'(x^*)$ has full rank, then for $y = (F'(x^*)^T F'(x^*))^{-1} d$ then $\forall y$

$$\forall y \in \mathbf{R}^n \quad \|N'(x^*)^T y\| \leq \gamma \|y\|$$

or

$$\|N'(x^*)^T\| \leq \gamma$$

Inexact Gauss-Newton

Gauss-Newton:

Solve for s^k:

$$\min\{||F'(x^k)s + F(x^k)||_2 : s \in \mathbf{R}^n\}$$

Need to find a suitable error:

$$F'(x^k)^T(F'(x^k)s + F(x^k)) \text{ should be small}$$

For a given real non-negative sequence $\{\eta_k\}$ and starting point x^0:

For $k = 0$ STEP 1 UNTIL Convergence DO

Find *some* approximate minimizer s^k so that

$$\|F'(x^k)^T r^k\| \leq \eta_k \|F'(x^k)^T F(x^k)\|$$

$$\text{where } r^k = F(x^k) + F'(x^k)s^k;$$

Set $x^{k+1} = x^k + s^k$.

Convergence

A.1. F is continuously differentiable in an open neighborhood Ω of x^*;

A.2. $F'(x^*)^T F(x^*) = 0$;

A.3. $F'(x^*)$ has full rank;

A.4. There exists $\gamma \geq 0$ so that for $x \in \Omega$,

$$\|(F'(x) - F'(x^*))^T F(x^*)\| \leq \gamma \|x - x^*\|_*$$

where $\|\cdot\|$ is any vectornorm and for every y

$$\|y\|_* = \|F'(x^*)^T F'(x^*) y\|.$$

Theorem: (Dennis and Steihaug 1983) Assume that

$$\gamma + \eta(1 + \gamma) < r < 1$$

for γ from A.4 and $0 \leq \eta_k \leq \eta$ in Inexact Gauss-Newton. Then there exists some $\epsilon > 0$ for which $||x^0 - x^*||_* \leq \epsilon$ implies that the sequence of inexact Gauss-Newton iterates $\{x^k\}$ is defined and converges at least linearly to x^* in the sense that

$$\|x^{k+1} - x^*\|_* \leq r \|x^k - x^*\|_*$$

The Linear Least Squares Problem

Let A be an m by n real matrix, $m \geq n, b \in \mathbf{R}^m$. The linear least-squares problem is:

$$\min_{x \in \mathbf{R}^n} ||Ax - b\,||_2.$$

Let the columns of A be divided into g groups $A_1, A_2, \ldots, A_g$, where A_i is a m by n_i submatrix and A_i may share columns with A_j.

$$\min\{||A_1x_1 + A_2x_2 + \cdots + A_gx_g - b||_2 :$$
$$x_1 \in \mathbf{R}^{n_1}, x_2 \in \mathbf{R}^{n_2}, \ldots, x_g \in \mathbf{R}^{n_g}\}.$$

This is an m by $\mathbf{n} = n_1 + n_2 + \cdots + n_g$ least-squares problem (using boldface):

$$\min_{\mathbf{x} \in \mathbf{R}^{\mathbf{n}}} ||\mathbf{A}\ \mathbf{x} - b\,||_2,$$

where $\mathbf{A} = (A_1|A_2|\cdots|A_g)$ is an $m \times \mathbf{n}$ matrix and $\mathbf{x} = (x_1, x_2, \ldots, x_g)^T$ is an $\mathbf{n}$ vector.

Successive Replacements

Let $x^k \approx x^*$ divided into $x_1^k, x_2^k, \ldots, x_g^k$.

FOR $i = 1, 2, \ldots, g$ DO

Solve for $x_i^{k+1} \in \mathbf{R}^{n_i}$

$$\min \| \sum_{j=1}^{i-1} A_j x_j^{k+1} + A_i x_i^{k+1} + \sum_{j=i+1}^{g} A_j x_j^k - b\|_2$$

Let s^k be the step and $r^k = Ax^k - b$ be the residual.

$$A_1 x_1^{k+1} + \sum_{j=2}^{g} A_j x_j - b = A_1 s_1^k + r^k.$$

The iteration is then:

FOR $i = 1, 2, \ldots, g$ DO
Solve for s_i^k : $\min\{||A_i s_i + r^{k+(i-1)/g}|| : s_i \in \mathbf{R}^{n_i}\}$;
Update the residual: $r^{k+i/g} = r^{k+(i-1)/g} + A_i s_i^k$.

The new approximate solution is

$$x_i^{k+1} = x_i^k + s_i^k, i = 1, \ldots, g; x^{k+1} = \sum_{i=1}^{g} \bar{x}_i^{k+1}.$$

Block-SOR

Subdivide A into g groups.
(Each A_i has full rank, $i = 1, 2, \ldots, g$.)
Choose $0 < \omega_i < 2, i = 1, 2, \ldots, g$.)

Choose $x_i^0, i = 1, \ldots, g$.
Compute $r^0 = Ax^0 - b$.
FOR $k = 0$ STEP 1 UNTIL Convergence DO
FOR $i = 1$ STEP 1 UNTIL g DO
$s_i^k = -(A_i^T A_i)^{-1} A_i^T r^{k+(i-1)/g}$;
$r^{k+1/g} = r^{k+(i-1)/g} + \omega_i A_i s_i^k$;
$x^{k+i/g} = x^{k+(i-1)/g} + \omega_i \bar{s}_i^k$;
Check for convergence.

Theorem: Let the columns of A be divided into g groups $A_1, .., A_g$ and let each A_i have full rank. Let $\{x^k\}$ be generated by the above algorithm with any choices $0 < \omega_i < 2$, and any $x_i^0 \in \mathbf{R}^{n_i}, i = 1, 2, \ldots, g$. Then $\{x^k\}$ converges to x^*, a solution to the original least-squares problem.

Gauss-Newton and SOR

For $s_i \in \mathbf{R}^{n_i}$, let $\bar{s}_i \in \mathbf{R}^n$ This suggests the following cycle in the inner loop:

For given x^k, let $r^k = F(x^k),\ y^k = x^k$;
FOR $c = 1$ STEP 1 UNTIL Termination DO
  FOR $i = 1$ STEP 1 UNTIL g DO
    Compute $A_i^k = F'(x^k)_i$ or $\Delta F(x^k)_i$;
    Solve for s_i^k :
      $\min\{\|A_i^k s_i + r^{k+(i-1)/g}\|_2 : s_i \in \mathbf{R}^{n_i}\}$;
    Set $r^{k+1/g} = r^{k+(i-1)/g} + \omega_i A_i^k s_i^k$;
    Set $y^{k+i/g} = y^{k+(i-1)/g} + \omega_i \bar{s}_i^k$.
  Check termination
    (with A^k in place of $F'(x^k)$).

The next iterate is now $x^{k+1} = y^{k+c} \in \mathbf{R}^n$ where c is the number of cycles, which corresponds to terminating the inner iteration after c sweeps through each of the column groups.

Nonlinear Substitution

Need to recompute $F'(x_i^k)$ or $\Delta F(x^k)_i$ when needed, or storing $F'(x^k)$ or $\Delta F(x^k)$. An alternative approach recomputes the Jacobian matrix of one particular group at a time and updates the nonlinear residual. This would suggest the following nonlinear substitution method:

Given x^0, compute $F(x^0)$
FOR $k = 0$ STEP 1 UNTIL Convergence DO
    FOR $i = 1$ STEP 1 UNTIL g DO
        Compute $A_i^{k+(i-1)/g} = F'(x^{k+(i-1)/g})_i$;
        Compute $r^{k+(i-1)/g} = F(x^{k+(i-1)/g})$;
        Solve for $s_i^k \in \mathbf{R}^{n_i}$:
            $\min\{\|A_i^{k+(i-1)/g} s_i + r^{k+(i-1)/g}\|_2 : s_i\}$;
        Set $x^{k+i/g} = x^{k+(i-1)/g} + \omega_i \bar{s}_i^k$;
    Check convergence.

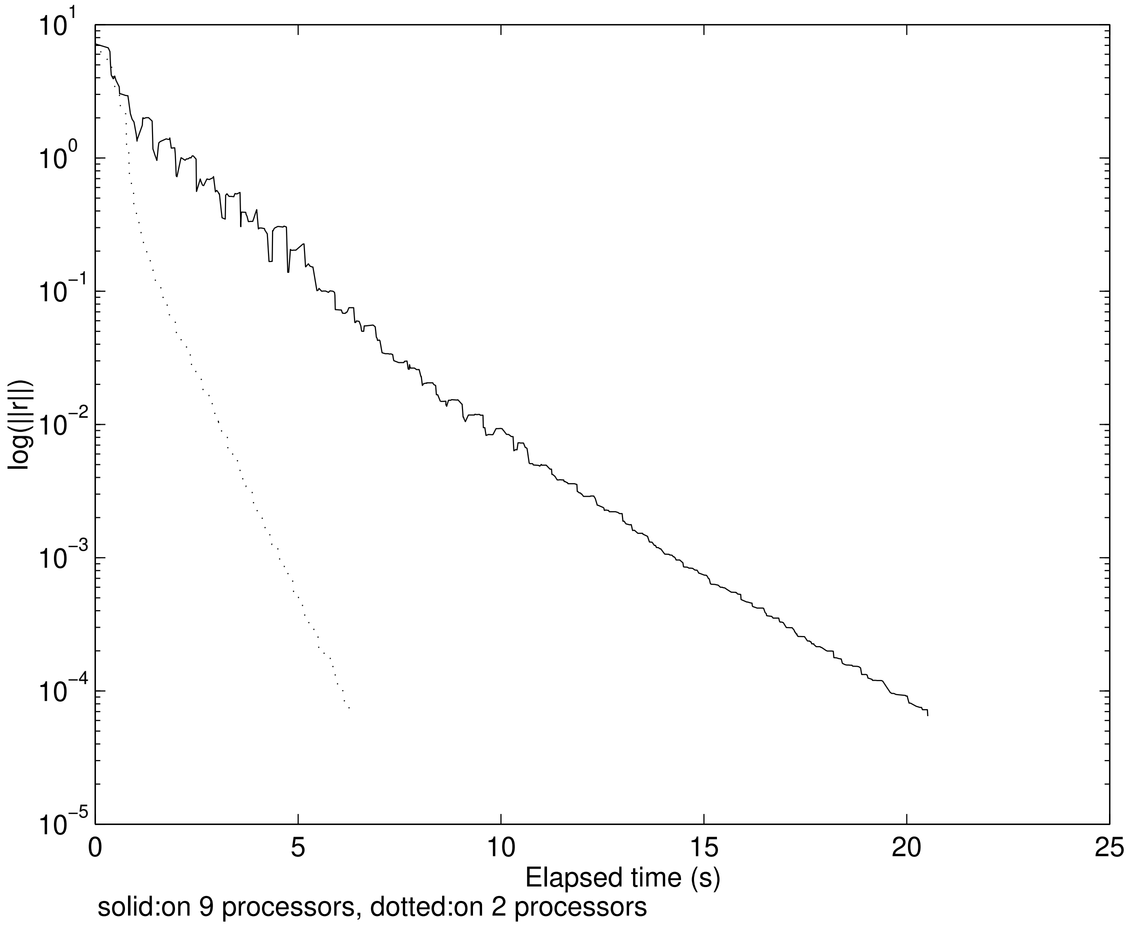
Asynchronuous Implementation

MPMD with $g+1$ processors. One master processor. **Put** and **Get** are communication with master processor. The inner FORALL loop runs until master process terminates each processor.

For given x^k, let $r^k = F(x^k),\ y^k = x^k$;
Put $r = r^k$; **Put** $y = y^k$;
Initiate each processor $i = 1, \dots, g$
  Compute $A_i^k = F'(x^k)_i$ and preprocess
WHILE NOT Termination DO
    FORALL $i = 1, \dots, g$ DO
        $r_i =$ **Get** r;
        Solve for s_i :
            $\min\{\|A_i^k s_i + r_i\|_2 : s_i \in \mathbf{R}^{n_i}\}$;
        Put $r =$ **Get** $r + \omega_i A_i^k s_i$;
        Put $y =$ **Get** $y + \omega_i \bar{s}_i^k$.
    Master processor checks for termination

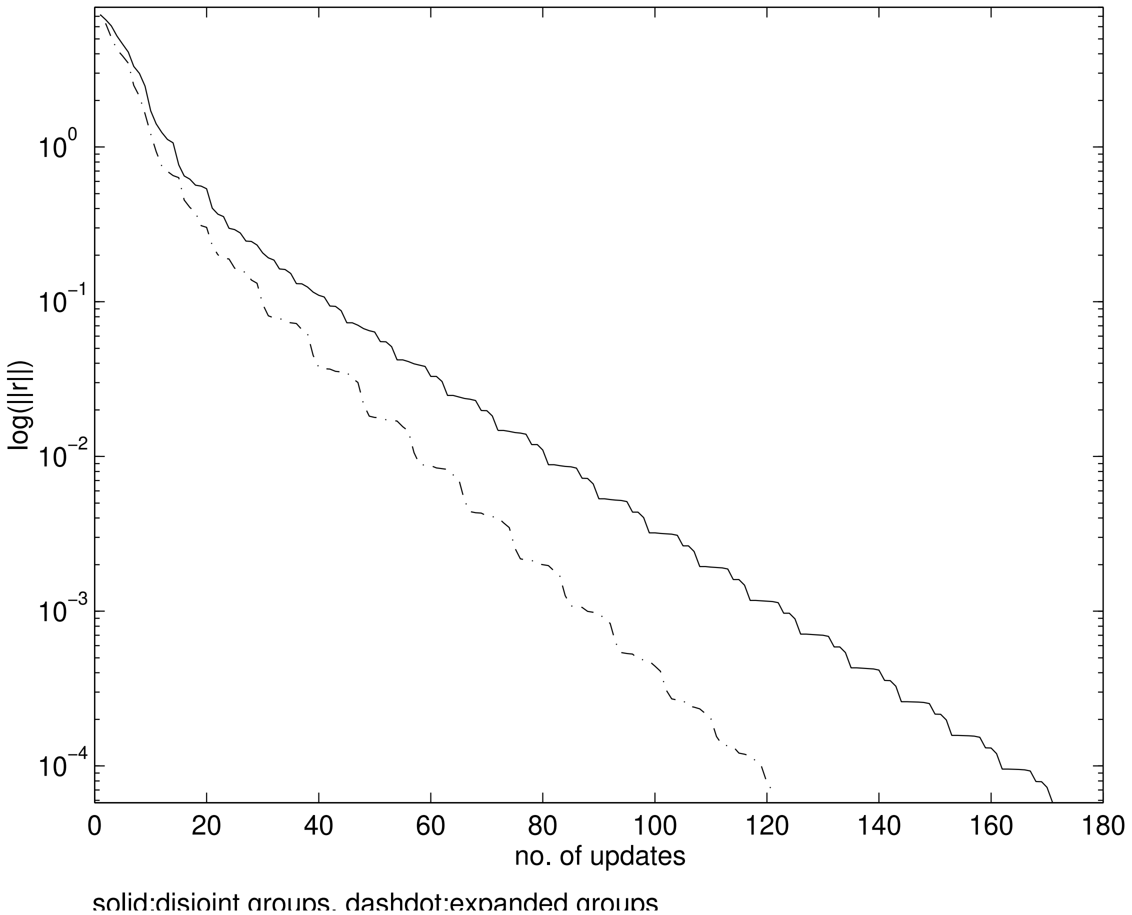
The next iterate is now $x^{k+1} = y \in \mathbf{R}^n$ on the Master processor.

Numerical Results



The residual for totally asynchronous iterations implemented with 2 (dotted) and 9 processor (disjoint groups) on a linear problem.

Numerical Results(2)



Sequential iterations on expanded and disjoint column groups. The asynchronous implementation shows a oscillatory behavior. The total number of updates in the asynchronous implementation example is 501 compared to 170 updates in the sequential process.

Concluding Remarks

- Inexact Primal Dual Interior Point method implemented using LIPSOL (Yin Zhang).
  - $\eta = 0.01$ reproduces "exact" solver (in most cases)
  - CG and QMR are so far the the most robust inner methods
  - No generally good precondtioners idenfied. (Use diag(AMA^T))
- Block Gauss-Newton is efficient
  - on small residual problems
  - "natural blocking"
- Asynchronous
  - Very slow on many processor
  - Need synchronization !
- Asynchronous in I PD IPM
  - Use multiplier estimate to find "important" equations
  - Scaling of A is important