ASTRAL: An Active Set ℓ^∞-Trust-Region Algorithm for Box Constrained Optimization

Liang (Linda) Xu

Department of Mathematics

University of Washington

Jan 9, 2007

Joint work with Jim Burke

8th US-Mexico Workshop on Optimization and its Applications

Minimization with simple constraints

$$(\mathcal{P}) \qquad \min \qquad f(x)$$
$$l \le x \le u.$$

General assumptions:

1. $f(x): \ \mathbb{R}^n \to \mathbb{R}$, are smooth.

2. n is large.

3. $l_i \in [-\infty, \infty)$, $u_i \in (-\infty, \infty]$, and $l_i \le u_i$.

$$\Omega = \{x \mid l \le x \le u\}.$$

Projected Gradient at x,

$$\nabla_\Omega f(x) = P_{T(x)}(-\nabla f(x)) \Leftrightarrow$$

$$(\nabla_\Omega f(x))_i = \begin{cases} -(\nabla f(x))_i & l_i < x_i < u_i; \\ \max(0, -(\nabla f(x))_i) & x_i = l_i; \\ \min(0, -(\nabla f(x))_i) & x_i = u_i. \end{cases}$$

Gradient Projection

$$p_x = P_\Omega(x - \nabla f(x)) - x \ .$$

Optimality condition

Minimization with simple constraints

$$\begin{aligned} \min \qquad & f(x) \\ & l \leq x \leq u. \end{aligned}$$

$\bar{x}$ is a local minima $\Rightarrow -\nabla f(\bar{x}) \in N_\Omega(\bar{x}) \Leftrightarrow \nabla_\Omega f(\bar{x}) = 0 \Leftrightarrow p_{\bar{x}} = 0.$

Active Constraints

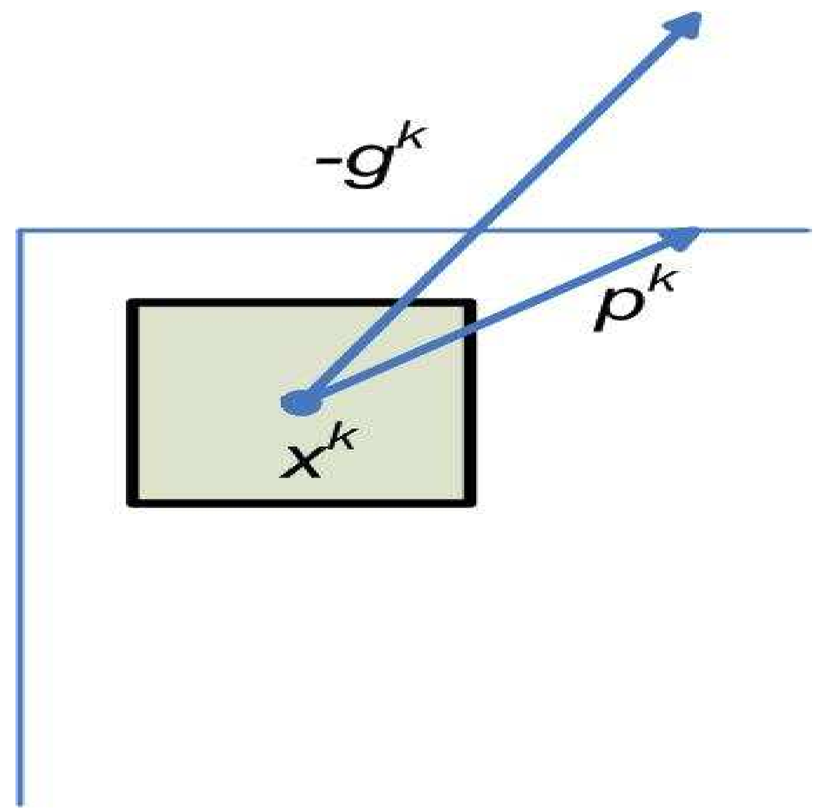
$$A(x) = \{i \mid x_i = l_i \text{ or } x_i = u_i\}$$

Binding Constraints

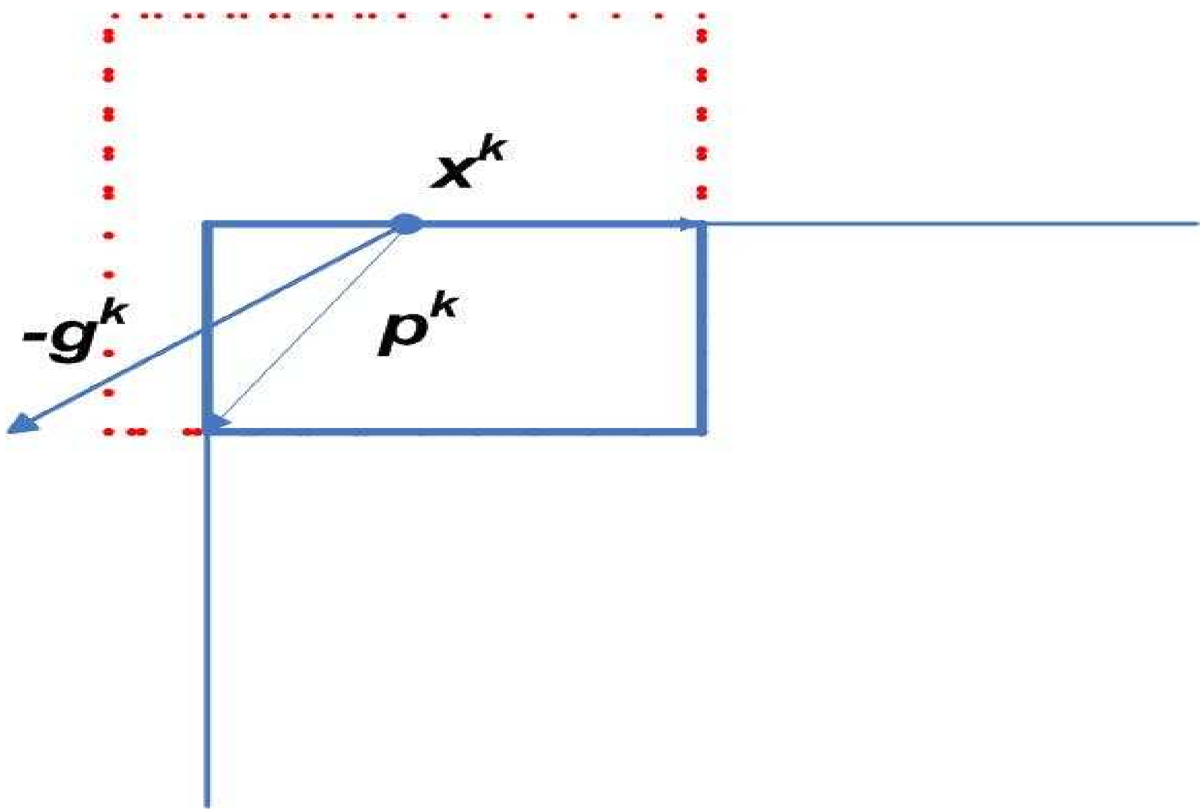
$$\mathcal{B}(x) = \left\{ i \left| \begin{array}{c} x_i = l_i \text{ and } (\nabla f(x))_i > 0, \\ \text{or } x_i = u_i \text{ and } (\nabla f(x))_i < 0 \end{array} \right. \right\}$$

Overview of Active-set Trust-Region Algorithm(ASTRAL)

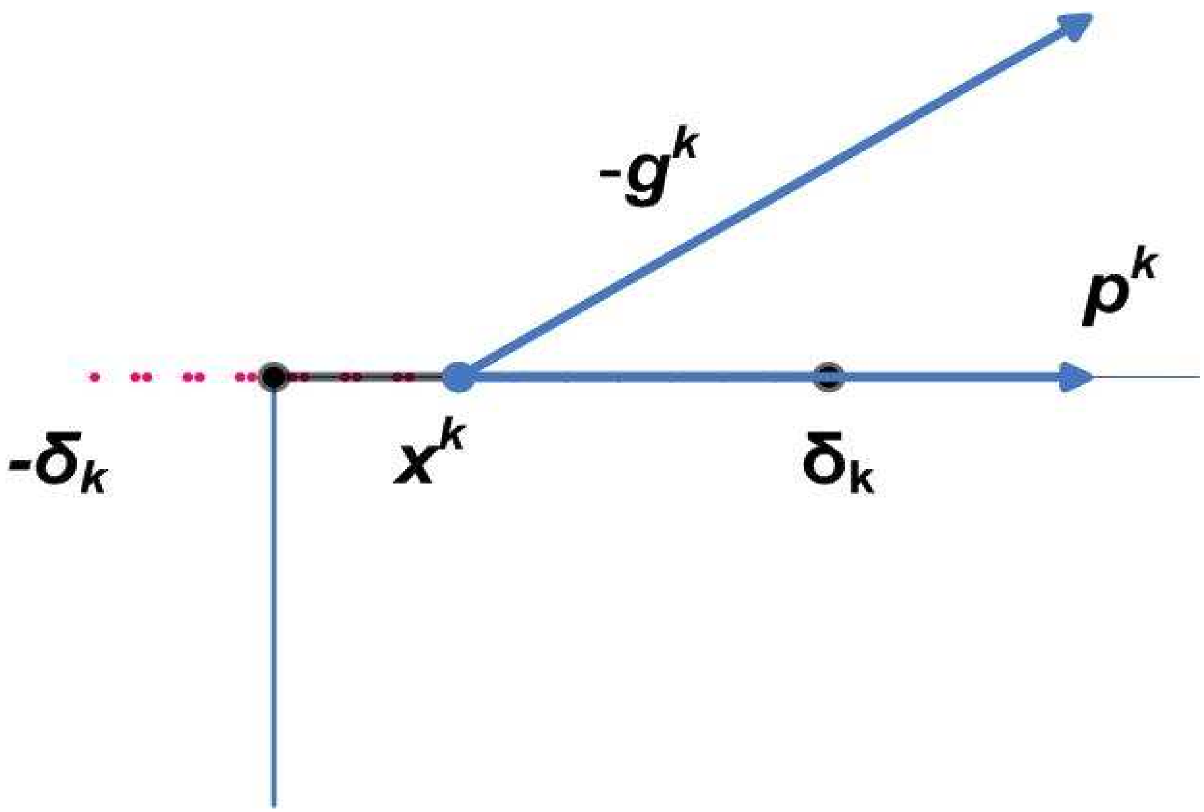
Case 1



Case 2



Case 3



Notation:

1. $\nu(x) = |\mathcal{B}^c(x)|$.

2. $\Phi(x)$ equal to the $n \times \nu(x)$ matrix whose columns are those of I corresponding to the non-binding constraints $\mathcal{B}^c(x)$.

3. $\Phi_k = \Phi(x^k), \quad \tilde{B}_k = \Phi_k^T B_k \Phi_k, \quad \tilde{g}^k = \Phi_k^T g^k,$

4. $(l^k)_j = \max\{(l - x^k)_j,\ -\delta_k\}, (u^k)_j = \min\{(u - x^k)_j, \delta_k\},$
$\tilde{l}^k = \Phi_k^T(l^k)$, and $\tilde{u}^k = \Phi_k^T(u^k)$

Active-set Trust-region Algorithm(Preliminary)

WHILE $||p^k||_\infty > \epsilon$

1. Identify the $\mathcal{B}(x^k)$. Solve trust-region subproblem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2} s^T \tilde{B}_k s + \tilde{g}^k s \\ \text{subject to} & \tilde{l}^k \leq s \leq \tilde{u}^k \end{array}$$

where $\tilde{B}_k = \Phi_k^T B_k \Phi_k, \quad \tilde{g}^k = \Phi_k^T g^k,$

2. Accept/Reject trial step $\tilde{s}$.

3. Update trust-region radius δ_k

ENDWHILE

Convergence Theory

Assumptions:

A1. There exists a M, such that $||B_k|| \leq M$.

A2. $\nabla f(x)$ satisfies Lipschitz condition over Ω.

Theorem 1 Under assumptions A1, A2. Let $\{x^k\}$ be a sequence generated by the algorithm ASTRAL. Then either the algorithm terminates finitely at a first-order stationary point of $\mathcal{P}$, or $f(x^k) \downarrow -\infty$, or $\lim_{k \to \infty} p^k = 0$.

L-BFGS Updating(Byrd, Nocedal, Schnabel).

Fix m to be a positive integer.

Define

$$s^k = x^{k+1} - x^k \text{ and } y^k = g^{k+1} - g^k$$

set

$$S = [s^{k-m+1}, \cdots, s^k], Y = [y^{k-m+1}, \cdots, y^k],$$

L-BFGS Hessian approximation at x^k is

$$B = \lambda I - \Psi \Gamma^{-1} \Psi^T.$$

where

$$\lambda = \frac{||y^k||_2^2}{{s^k}^T y}, \quad \Psi = [Y,\ \lambda S], \quad \Gamma = \begin{bmatrix} -D & L^T \\ L & \lambda S^T S \end{bmatrix},$$

and

$$S^T Y = L + D + R\ .$$

LM-BFGS and an Interior Point Method for the QP Suproblems

Let $w = s - \tilde{l}^k$, $h = \tilde{u}^k - \tilde{l}^k$.

$$\begin{array}{ll} \text{minimize} & \frac{1}{2} s^T \tilde{B}_k s + \tilde{g}^k s \\ \text{subject to} & \tilde{l}^k \le s \le \tilde{u}^k \end{array} \quad \Leftrightarrow \quad \mathcal{Q} \quad \begin{array}{ll} \text{minimize} & \frac{1}{2} w^T \tilde{B}_k w + \hat{g}^T w \\ \text{subject to} & 0 \le w \le h. \end{array} .$$

$$(\tilde{B}_k + U^{-1}V + W^{-1}Z)^{-1}$$

$$=$$

$$\Xi^{-1} + \Xi^{-1}(\Phi_k^T\Psi)(\Gamma - (\Phi_k^T\Psi)^T\Xi^{-1}(\Phi_k^T\Psi))^{-1}(\Phi_k^T\Psi)^T\Xi^{-1}$$

$$(\text{where } \Xi = \lambda I + U^{-1}V + W^{-1}Z)$$

Write $\Gamma - (\Phi_k^T\Psi)^T\Xi^{-1}(\Phi_k^T\Psi) = \begin{bmatrix} -\hat{D} & \hat{L}^T \\ \hat{L} & \hat{W} \end{bmatrix}$.

$$\Gamma - (\Phi_k^T\Psi)^T\Xi^{-1}(\Phi_k^T\Psi) = \begin{bmatrix} M & 0 \\ -\hat{L}M^{-T} & J \end{bmatrix} \begin{bmatrix} -M^T & M^{-1}\hat{L}^T \\ 0 & J^T \end{bmatrix}.$$

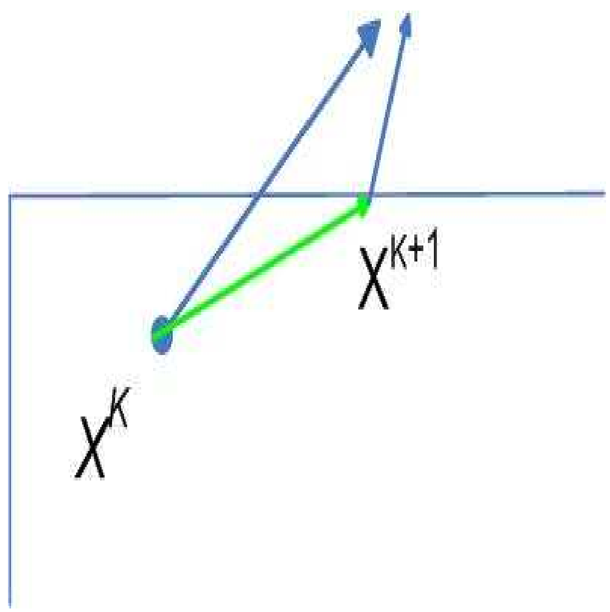
where $\hat{D} = MM^T$ and $JJ^T = \hat{W} + \hat{L}\hat{D}^{-1}\hat{L}^T$.

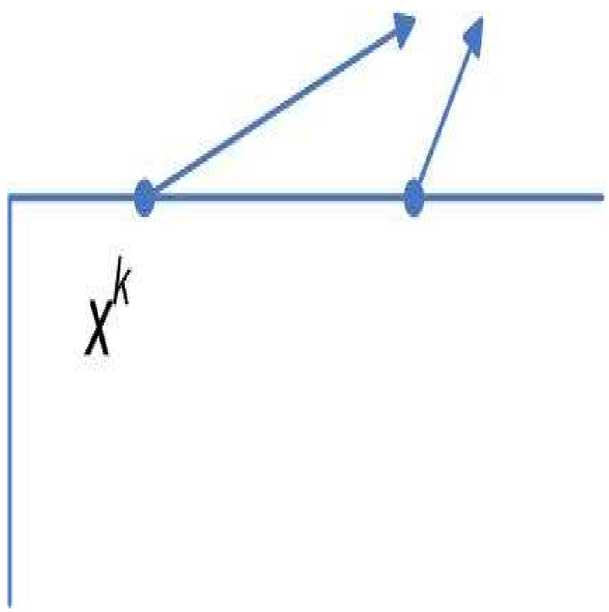
Restarting scheme and Local Convergence

(Gradient Projection Re-Start) If $\|p^k\| = 0$, **STOP**; if $\|p^k\| > \tau_k$, continue; otherwise, set

$$\tau_{k+1} = \gamma\tau_k \quad \hat{x} = P_\Omega(x^k - \nabla f(x^k)), \quad \hat{f} = f(\hat{x}), \text{ and } \hat{g} = \nabla f(\hat{x}) \ .$$

If $\mathcal{B}(\hat{x}) \neq \mathcal{B}(x^k)$, set
$x^{k+1} = \hat{x}, \quad f^{k+1} = \hat{f}, \quad g^{k+1} = \hat{g}, \quad \delta_{k+1} = \delta_k$, continue.





Active-set Trust-region Algorithm

WHILE $||\nabla_\Omega f(x^k)||_\infty > \epsilon$

0. Gradient Projection Re-start.

1. Identify the $\mathcal{B}(x^k)$. Solve trust-region subproblem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}s^T\tilde{B}_k s + \tilde{g}^k s \\ \text{subject to} & \tilde{l}^k \le s \le \tilde{u}^k \end{array}$$

where $\tilde{B}_k = \Phi_k^T B_k \Phi_k, \quad \tilde{g}^k = \Phi_k^T g^k,$

2. Accept/Reject trial step $\tilde{s}$.

3. Update trust-region radius δ_k

ENDWHILE

Let $\bar{x}$ be a stationary point of $\mathcal{P}$,

A3 (Non-Degeneracy) $-\nabla f(\bar{x}) \in \text{ri}\,(N_\Omega\,(\bar{x}))$.

A4 (Second-Order Sufficiency) $\Phi_k^T B_k \Phi_k$ is positive definite.

A5 (Uniform Positive Definiteness) There exists $\rho > 0$ and $\epsilon > 0$ such that for $x \in \mathbb{B}(\bar{x}, \epsilon)$, we have

$$s^T B_k s > \rho s^T s$$

where $s_i = 0$ if $i \in A(\bar{x})$.

Theorem 2

- Suppose A1-A2 hold and let $\{x^k\}$ be a sequence generated by ASTRAL with $\tau_0 > 0$. If $\bar{x}$ is a cluster point of the subsequence $J = \left\{k \mid \left\|p^{k-1}\right\| \leq \tau_{k-1}\right\}$, then $\bar{x}$ is a first-order stationary point of $\mathcal{P}$

- If in addition, $\bar{x}$ and the sequence of matrices $\{B_k\}$ satisfy A3-A5, then a gradient projection re-start occurs at most finitely many times, $x^k \to \bar{x}$, and $\nabla_\Omega f(x^k) \to 0$.

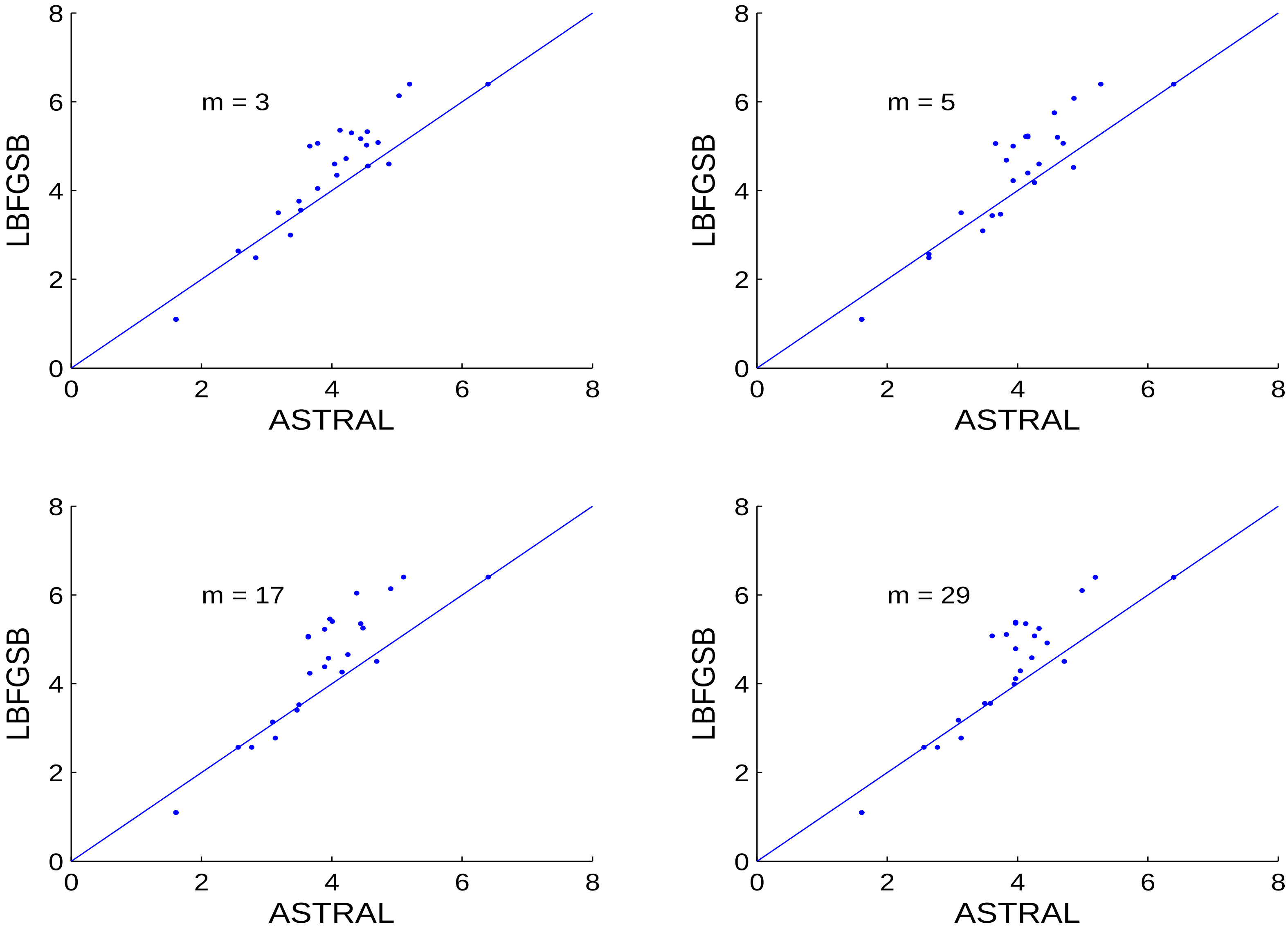
Corollary Suppose $B_k = \nabla^2 f(x^k)$ is positive definite and assume that $\nabla^2 f(\bar{x})$ is locally Lipschitz at $\bar{x}$. Then the ASTRAL iterates $\{x^k\}$ converge quadratically to $\bar{x}$.

Numerical Results(MATLAB Implementation)

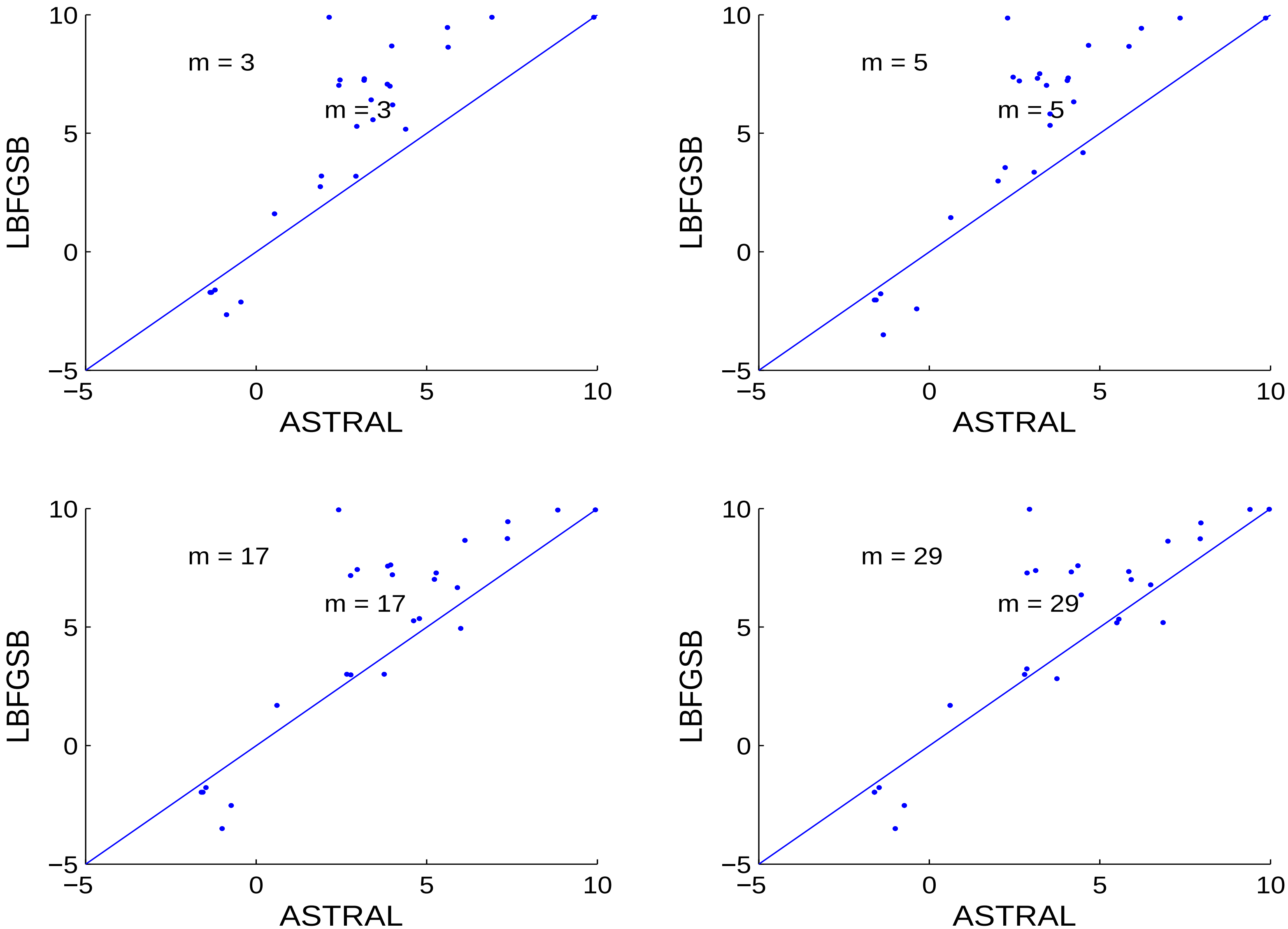
Termination Criteria:

1. $||\nabla_\Omega(f(x^k))||_\infty < 10^{-5}$.
2. $||s^k||_\infty < 10^{-10}$.
3. $|f^k - f_{\text{best}}|/\max(1, |f_{\text{best}}|) < 10^{-10}$.
4. nfg > 500.

Comparison of nfg(Matlab implementation)



Comparison of CPU time(Matlab implementation)



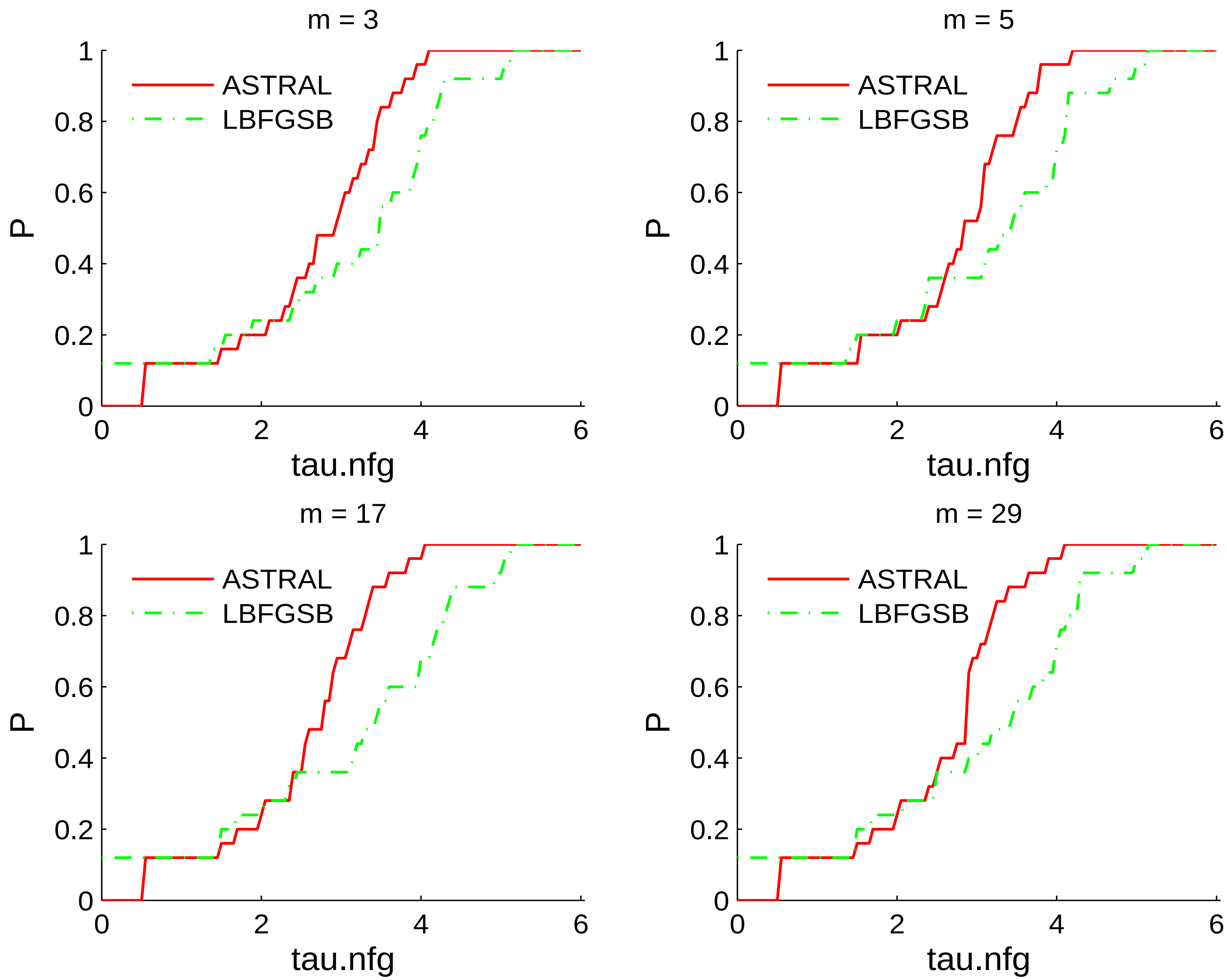
- Matlab implementation for ASTRAL available on www.math.washington.edu/lxu/

- Matlab implementation for L-BFGSB will soon be posted.

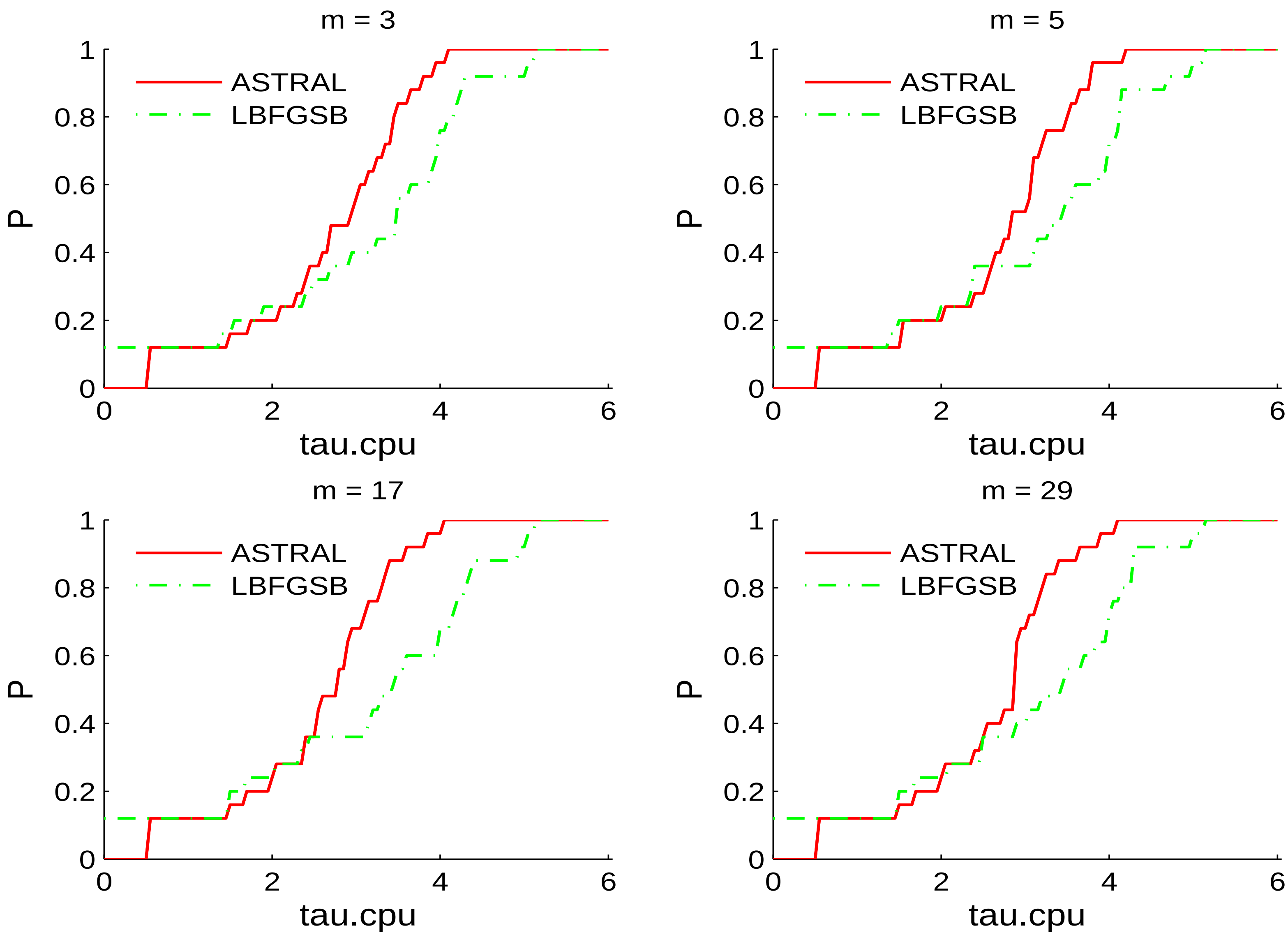
- Working on: C++ & Fortran implementation.

Thank you!!!

Performance profile (Dolen and Moré 2001) comparison nfg



Performance profile CPU time



The ASTRAL Algorithm

Step 0: (Initialization) Let $\hat{\delta}_0 > 0$, $x^0 \in \Omega$, and $B_0 \in \mathbb{R}^{n \times n}$ be symmetric and positive definite. Set

$$g^0 = \nabla f(x^0) \quad \text{and} \quad p^0 = p(x^0),$$

and choose the following constants:

$$0 < \beta_0 < \beta_1 < \beta_2 < 1, \qquad \text{(step acceptance parameters)}$$

$$0 < \sigma_1 < \sigma_2 < 1 < \sigma_3\ , \qquad \text{(trust-region scaling parameters)}$$

$$0 < \gamma < 1,\ 0 \leq \tau_0 \qquad \text{(re-start parameters)}$$

Set $\delta_0 = \min\{\max_i\{u_i - l_i\}, \hat{\delta}_0\}$ and $k = 0$.

Step 1: (Gradient Projection Re-Start) If $\|p^k\| = 0$, **STOP**; if $\|p^k\| > \tau_k$, go to Step 2; otherwise, set

$$\tau_{k+1} = \gamma\tau_k \quad \hat{x} = P_\Omega(x^k - \nabla f(x^k)), \quad \hat{f} = f(\hat{x}), \text{ and } \hat{g} = \nabla f(\hat{x}) \, .$$

If $\mathcal{B}(\hat{x}) \neq \mathcal{B}(x^k)$, set

$$x^{k+1} = \hat{x}, \quad f^{k+1} = \hat{f}, \quad g^{k+1} = \hat{g}, \quad \delta_{k+1} = \delta_k,$$

choose $B^{k+1} \in \mathbb{R}^{n\times n}$ to be symmetric and positive definite, compute

$$p^{k+1} = P_\Omega(x^{k+1} - g^{k+1}) - x^{k+1} \; ,$$

update k to $k+1$, and go to Step 2; otherwise, continue to Step 2.

Step 2 (Solve the Trust-Region Subproblem) Set

$$(l^k)_j = \max\{(l - x^k)_j,\ -\delta_k\}, \quad (u^k)_j = \min\{(u - x^k)_j, \delta_k\}$$

$$\text{for} \quad j = 1, 2, \ldots, n,$$

$$\Phi_k = \Phi(x^k), \quad \tilde{B}_k = \Phi_k^T B_k \Phi_k, \quad \tilde{g}^k = \Phi_k^T g^k,$$

$$\tilde{l}^k = \Phi_k^T (l^k), \text{ and } \tilde{u}^k = \Phi_k^T (u^k) \ .$$

Let $\tilde{s}$ be the solution to the $\nu(x^k)$-dimensional trust-region subproblem

$$\begin{array}{lll} (TR)_k & \text{minimize} & \frac{1}{2} s^T \tilde{B}_k s + \tilde{g}^k s \\ & \text{subject to} & \tilde{l}^k \le s \le \tilde{u}^k \end{array} .$$

Set

$$q_k = \frac{1}{2} \tilde{s}^T \tilde{B}_k \tilde{s} + \tilde{g}^k \tilde{s}, \quad \bar{s}^k = \Phi_k \tilde{s}, \text{ and } r_k = (f(x^k + \bar{s}^k) - f(x^k))/q_k .$$

Step 3: (Update the iterates) If $r_k \geq \beta_0$, set $s^k = \bar{s}^k$; otherwise, set $s^k = 0$. Update the iterate as follows:

$$x^{k+1} = x^k + s^k,\ g^{k+1} = \nabla f(x^{k+1}),\ p^{k+1} = p(x^{k+1}),\ \tau_{k+1} = \tau_k,$$

and choose $B^{k+1} \in \mathbb{R}^{n \times n}$ to be symmetric and positive definite.

Step 4: (Update the trust-region radius) Set

$$\delta_{k+1} = \begin{cases} \sigma_1 \delta_k, & \text{if } r < \beta_0; \\ \sigma_2 \delta_k, & \text{if } \beta_0 \leq r < \beta_1; \\ \delta_k, & \text{if } \beta_1 \leq r < \beta_2; \\ \min(\sigma_3 \delta_k,\ \delta_0), & \text{if } r \geq \beta_2, \end{cases} .$$

Step 5: Update k to $k+1$ and return to Step 1.