Understanding Asynchronous Parallel Pattern Search

Tamara G. Kolda
Sandia National Labs

Virginia Torczon
College of William & Mary

Thanks to the U.S. Department of Energy and the National Science Foundation.

Asynchronous Parallel Pattern Search

APPS (Hough, Kolda, Torczon; 2001) is the latest in the family of pattern search methods for solving problems of the form

$$\min f(x), \quad x \in \mathbb{R}^n.$$

Motivation: effective parallelization on current distributed computing platforms.

Key: removing the synchronization inherent in earlier definitions of pattern search.

Notation

Processes:	$\mathcal{P} = \{1, \ldots, p\}$
Search Directions:	$\mathcal{D} = \{d_1, \ldots, d_p\}$
Global Time Index:	$\mathcal{T} = \{0, 1, 2, \ldots\}$
Best Known Point:	x_i^t (at time t for process i)
Step Length Control:	Δ_i^t (at time t for process i)

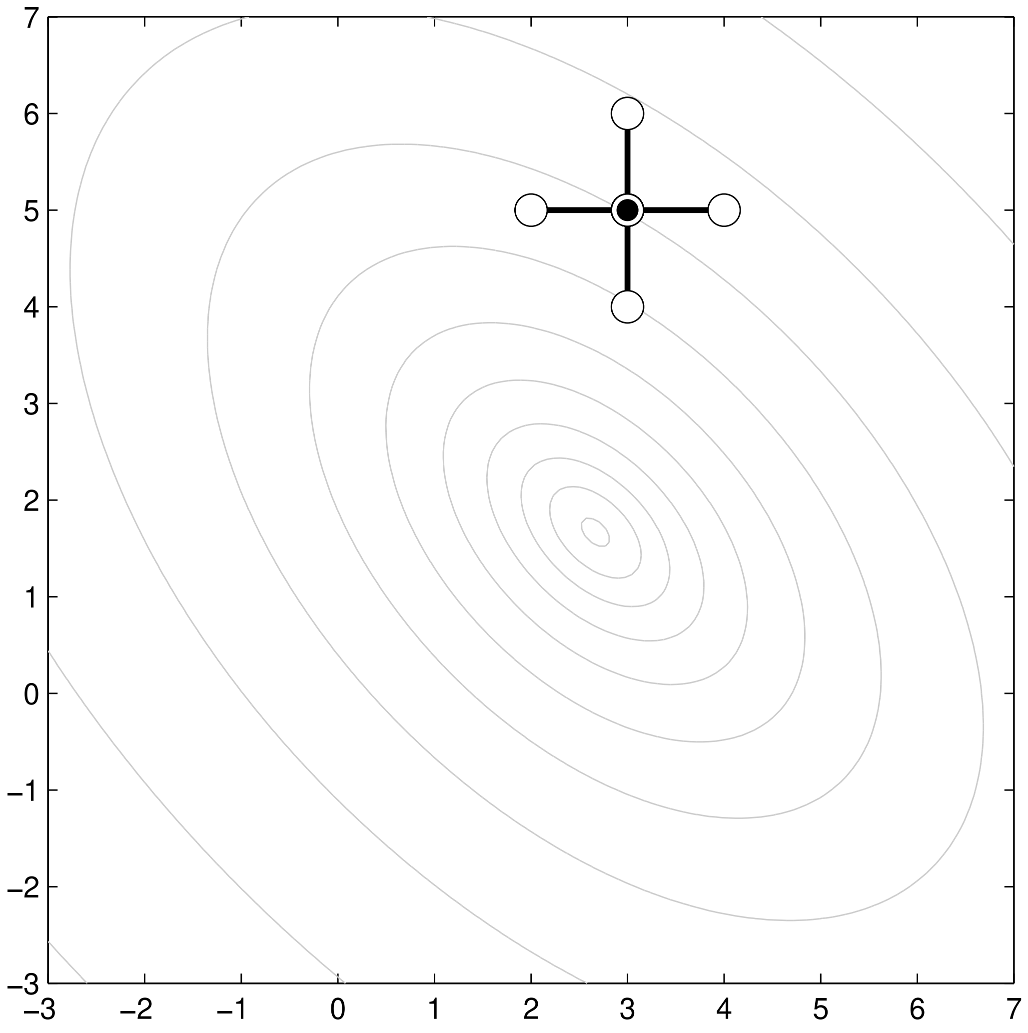
Starting the Search

At time 0, each process starts a function evaluation at the point $x_i^0 + \Delta_i^0 \, d_i$.

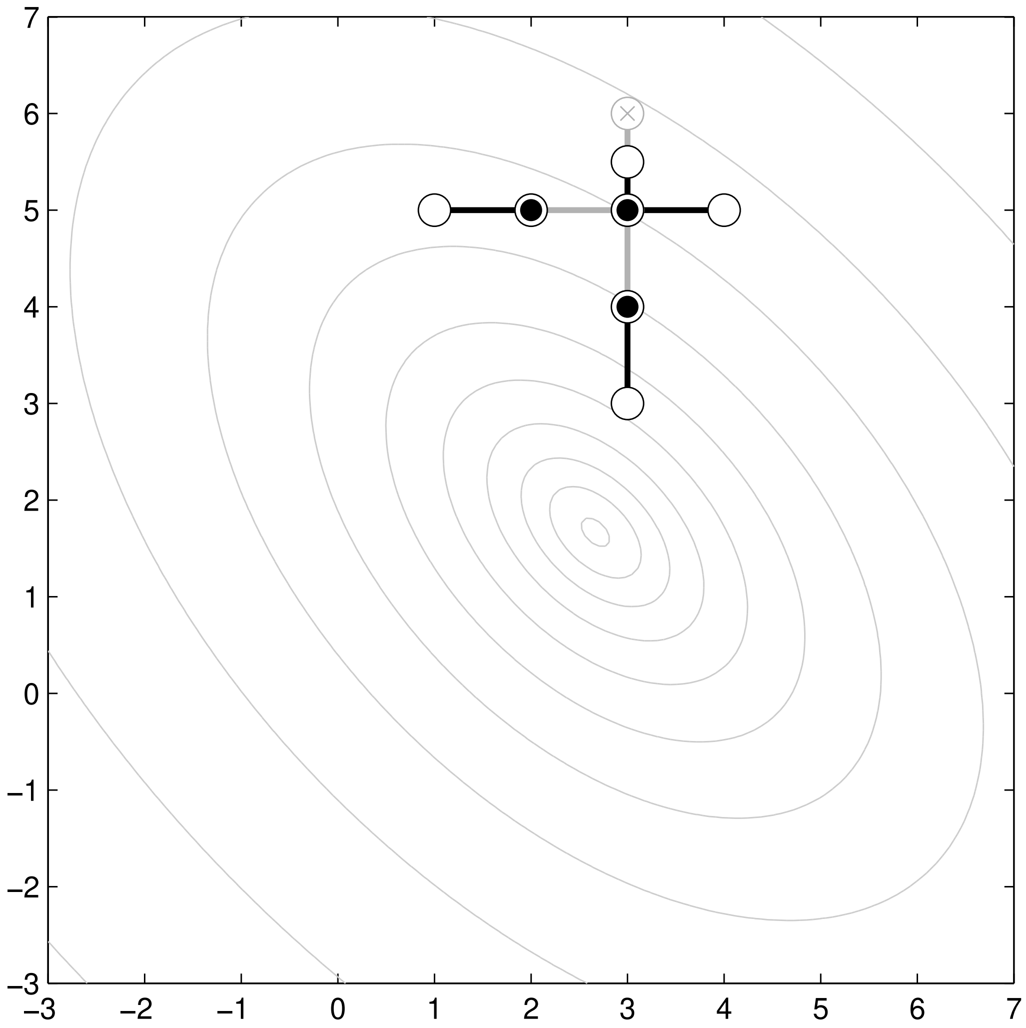
We assume

- $0 < \Delta_{\min} \leq \Delta^0 \leq \Delta_{\max} < +\infty$,
- $x^0 = x_1^0 = \cdots = x_p^0$,
- $\Delta^0 = \Delta_1^0 = \cdots = \Delta_p^0$, and
- $f(x_i^0)$ is known.

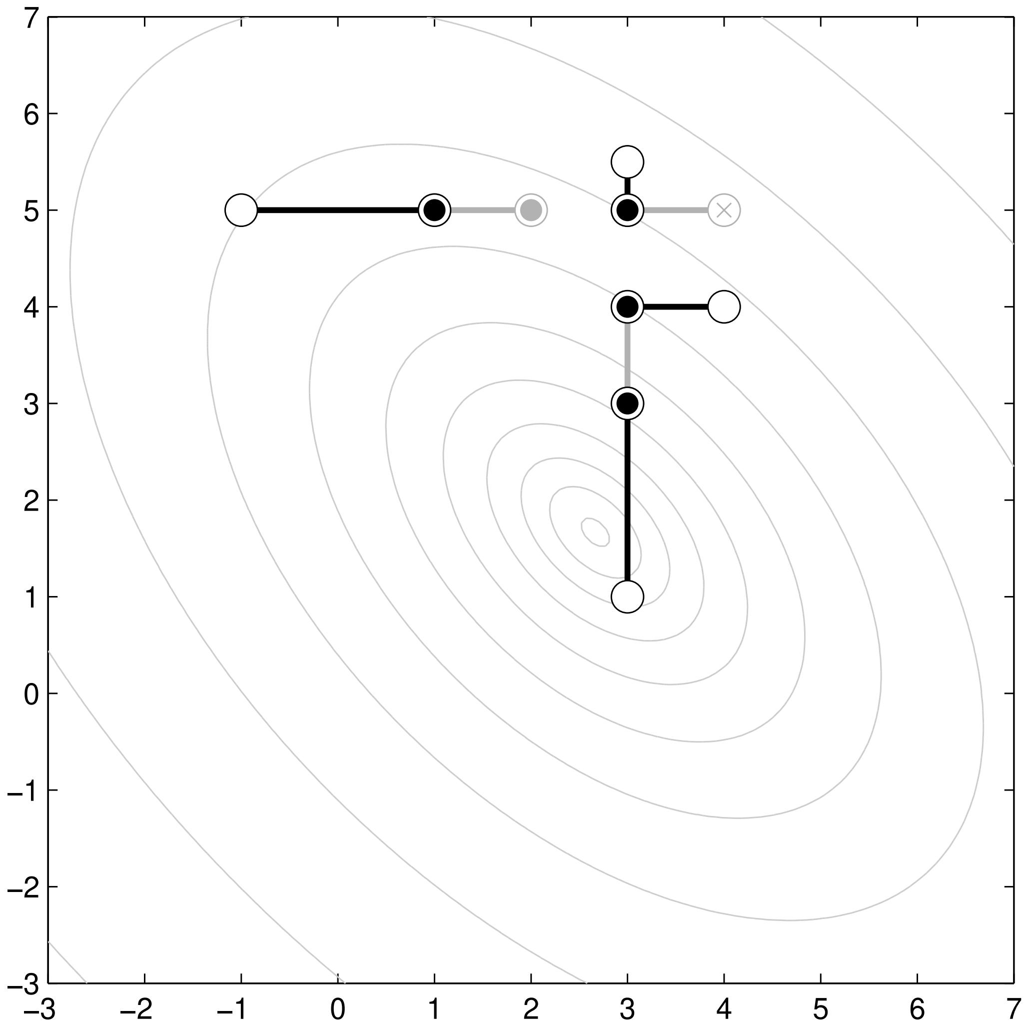
Example at Time Step $t = 0$



Example at Time Step $t = 1$



Example at Time Step $t = 2$



EXAMPLE AT TIME STEP $t = 3$

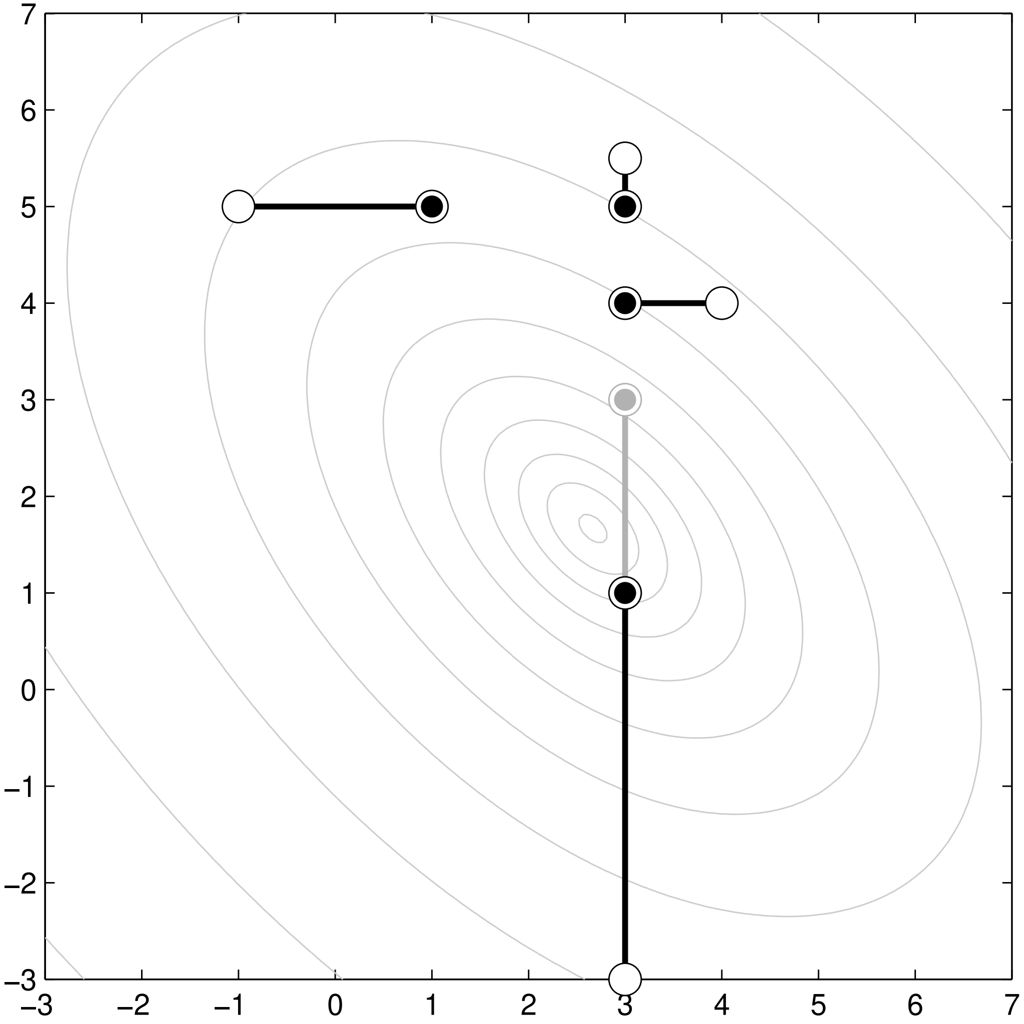
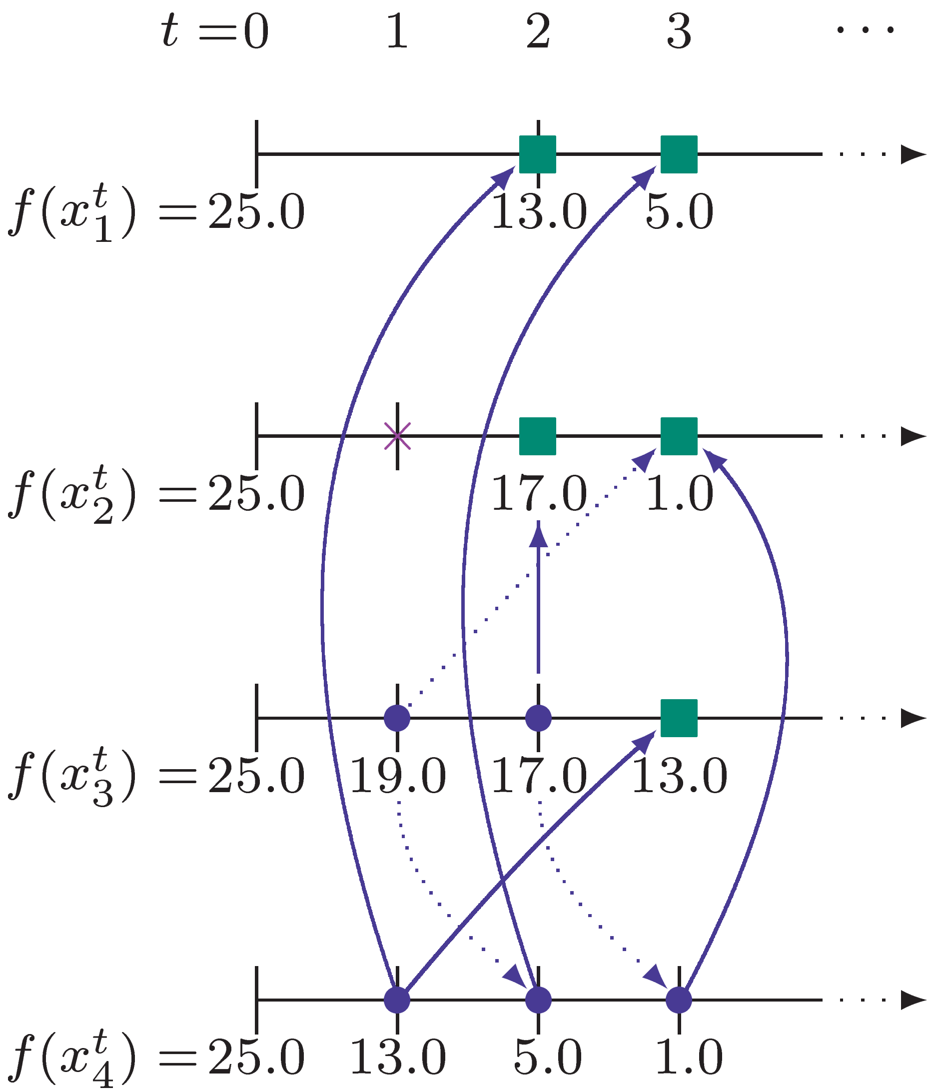


Illustration using Timelines



Caveat

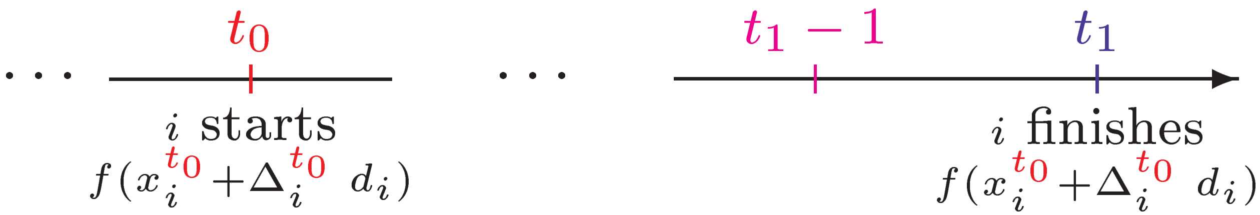
The bookkeeping we now introduce to analyze *APPS is* not *required for the* implementation.

Internal Successes

$\mathcal{I}_i$ = time steps that are internal successes on process i

t_0 = time function evaluation begins on process i

t_1 = time function evaluation finishes on process i



If $\quad f\left(x_i^{t_0} + \Delta_i^{t_0}\, d_i\right) < f\left(x_i^{t_1 - 1}\right)$,

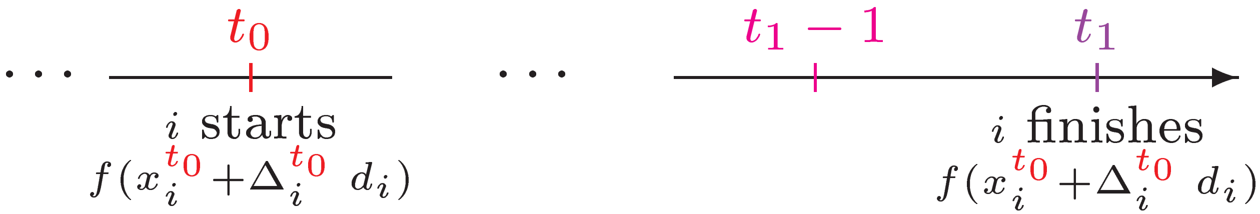
then $\quad x_i^{t_1} = x_i^{t_0} + \Delta_i^{t_0}\, d_i \quad$ and $\quad \Delta_i^{t_1} = \lambda_i^{t_1} \Delta_i^{t_0}$,

where $\quad \lambda_i^{t_1} \geq 1 \quad$ and $\quad \Delta_{\min} \leq \Delta_i^{t_1} \leq \Delta_{\max}$.

Broadcast $x_i^{t_1}, f(x_i^{t_1})$, and $\Delta_i^{t_1}$ to all $j \in \mathcal{P}, j \neq i$.

Contractions

$\mathcal{C}_i$ = time steps that are contractions on process i



If $\quad f\left(x_i^{t_1-1}\right) \leq f\left(x_i^{t_0} + \Delta_i^{t_0} \; d_i\right) \quad$ and $\quad x_i^{t_0} = x_i^{t_1-1}$

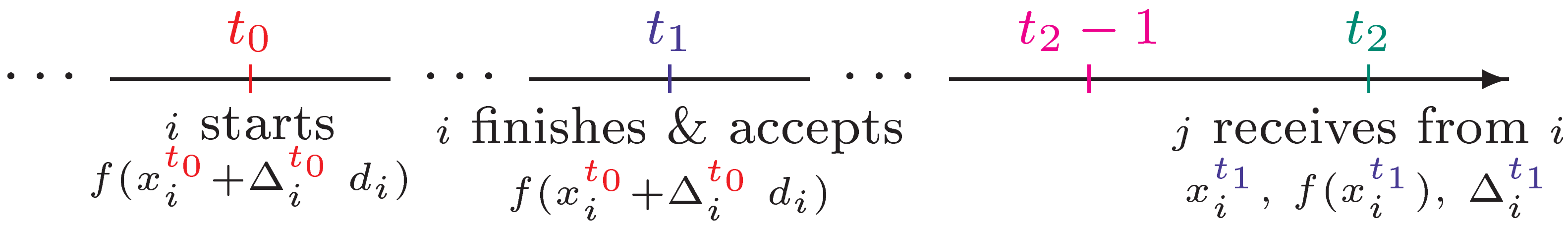
then $\quad x_i^{t_1} = x_i^{t_0} \quad$ and $\quad \Delta_i^{t_1} = \theta_i^{t_1} \Delta_i^{t_0}$,

where $\theta_i^{t_1} \in [\theta_{\min}, \theta_{\max}] \subset (0, 1)$.

External Successes

$\mathcal{E}_j$ = time steps that are external successes on process j

t_2 = time broadcast from process i reaches process j



If

$$
\begin{aligned}
f\left(x_i^{t_1}\right) &< f\left(x_j^{t_2-1}\right) \quad \text{or} \\
f\left(x_i^{t_1}\right) &= f\left(x_j^{t_2-1}\right) \quad \text{and} \quad x_i^{t_1} \prec x_j^{t_2-1},
\end{aligned}
$$

then

$$x_j^{t_2} = x_i^{t_1} \quad \text{and} \quad \Delta_j^{t_2} = \Delta_i^{t_1}.$$

Unchanged

$\mathcal{U}_i \quad = \quad$ time steps for which neither x_i^t nor Δ_i^t change on process i so that $x_i^t = x_i^{t-1}$ and $\Delta_i^t = \Delta_i^{t-1}$.

Possible reasons:

- $t_0 < t < t_1$ so $f(x_i^{t_0} + \Delta_i^{t_0})$ is still being computed on process i;
- no external candidate arrives from process j, $j \neq i$;
- an external candidate arrives but is discarded because it does not improve upon $f(x_i^{t-1})$; or
- $t = t_1$ but $f(x_i^{t_0} + \Delta_i^{t_0})$ does not improve upon $f(x_i^{t-1})$ and $x_i^{t-1} \neq x_i^{t_0}$.

Summary of Possible Outcomes

1. **Internal Success**: $t \in \mathcal{I}_i$. Both x_i^t and Δ_i^t change.
2. **Contraction**: $t \in \mathcal{C}_i$. Only Δ_i^t changes.
3. **External Success**: $t \in \mathcal{E}_i$. Both x_i^t and Δ_i^t change.
4. **Unchanged**: $t \in \mathcal{U}_i$. Neither x_i^t nor Δ_i^t change.

Also,

Successes: $t \in S_i = \mathcal{I}_i \cup \mathcal{E}_i$.

Changes: $t \in \mathcal{T}_i = \mathcal{S}_i \cup \mathcal{C}_i$.

Global Time Index: $t \in \mathcal{T} = \mathcal{T}_i \cup \mathcal{U}_i$.

Identifying the Source of a Change

$\omega_i(t)$ = the generating process index for the update at time step t on process i,

$\tau_i(t)$ = the time index for the initiation of the function evaluation that produced the update at time step t on process i, and

$\nu_i(t)$ = the time index for the completion of the function evaluation that produced the update at time step t on process i.

The Definitions for x_i^t and Δ_i^t

$$x_i^t = \begin{cases} x_{\omega_i(t)}^{\tau_i(t)} + \Delta_{\omega_i(t)}^{\tau_i(t)}\, d_{\omega_i(t)}, & \text{if } t \in \mathcal{S}_i, \text{ and} \\ x_i^{t-1}, & \text{otherwise.} \end{cases}$$

$$\Delta_i^t = \begin{cases} \lambda_{\omega_i(t)}^{\nu_i(t)}\, \Delta_{\omega_i(t)}^{\tau_i(t)}, & \text{if } t \in \mathcal{S}_i, \\ \theta_i^t \Delta_i^{t-1}, & \text{if } t \in \mathcal{C}_i, \text{ and} \\ \Delta_i^{t-1}, & \text{otherwise.} \end{cases}$$

Major New Features

- Semi-autonomous searches along each direction d_i.
- Changes to x_i^t and Δ_i^t based purely on knowledge local to process i regarding the current state of the search .
- External successes that require changes to x_i^t and Δ_i^t without regard for the history of past successes along d_i.
- Bounds on Δ_i^t after an internal success:

$$0 < \Delta_{\min} \leq \Delta_i^t \leq \Delta_{\max} < +\infty$$

in lieu of the classic Armijo-Goldstein-Wolfe step-length control conditions.

Bottom Line

Still preserve asymptotic convergence to a stationary point of f.

2001 *Understanding asynchronous parallel pattern search*. Submitted to *High Performance Algorithms and Software for Nonlinear Optimization*, G. Di Pillo and A. Murli, Editors, Kluwer Academic Publishers B.V.

Visit:

http://www.cs.wm.edu/~va/research/

http://csmr.ca.sandia.gov/~tgkolda/