SINBAD Consortium 2010

Hybrid Stochastic-Deterministic Methods

Michael Friedlander and Mark Schmidt

SLIM
University of British Columbia

Outline

- Stochastic vs. Deterministic Optimization
- Theoretical Analysis of a Hybrid Method
- Empirical Analysis of a Hybrid Method
- Future Work

Algorithm S vs. Algorithm D

ALGORITHM S

ALGORITHM D

- We want to solve an optimization problem.
- Should we use Algorithm S or Algorithm D?

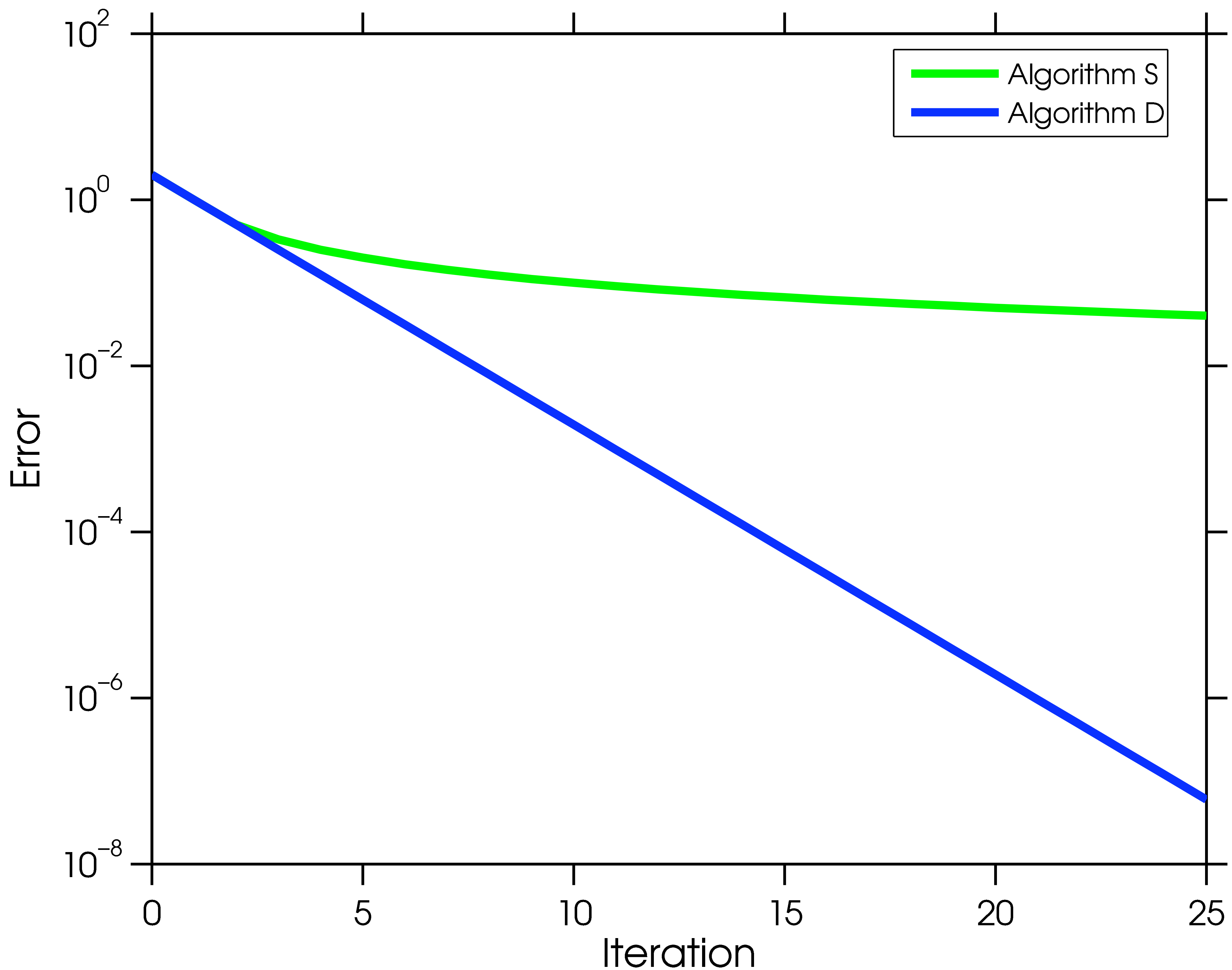
Algorithm S vs. Algorithm D

Algorithm S
Error = 1/k

Algorithm D
Error = $1/2^k$

- On iteration k, Algorithm S has an error of 1/k
- On iteration k, Algorithm D has an error of $1/2^k$

Algorithm S vs. Algorithm D



Algorithm S vs. Algorithm D

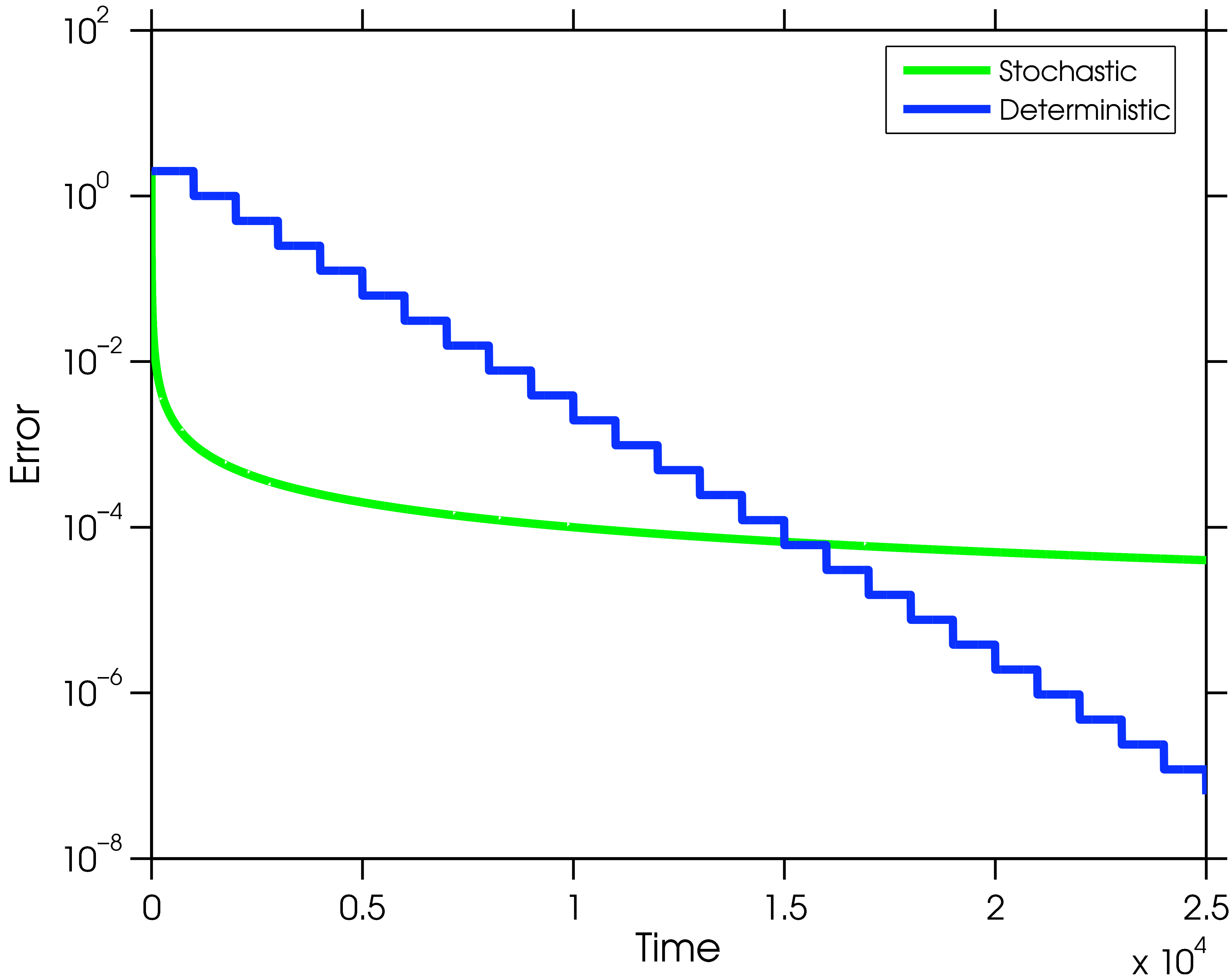
Algorithm S
Error = 1/k
Cost = 1

Algorithm D
Error = $1/2^k$
Cost = 1000

- Iterations of Algorithm S (Stochastic) are ***cheap***
- Iterations of Algorithm D (Deterministic) are 1000 times more expensive

SLIM

Stochastic vs. Deterministic



Stochastic vs. Deterministic

Stochastic	Deterministic
Error = 1/k	Error = $1/2^k$
Cost = 1	Cost = 1000

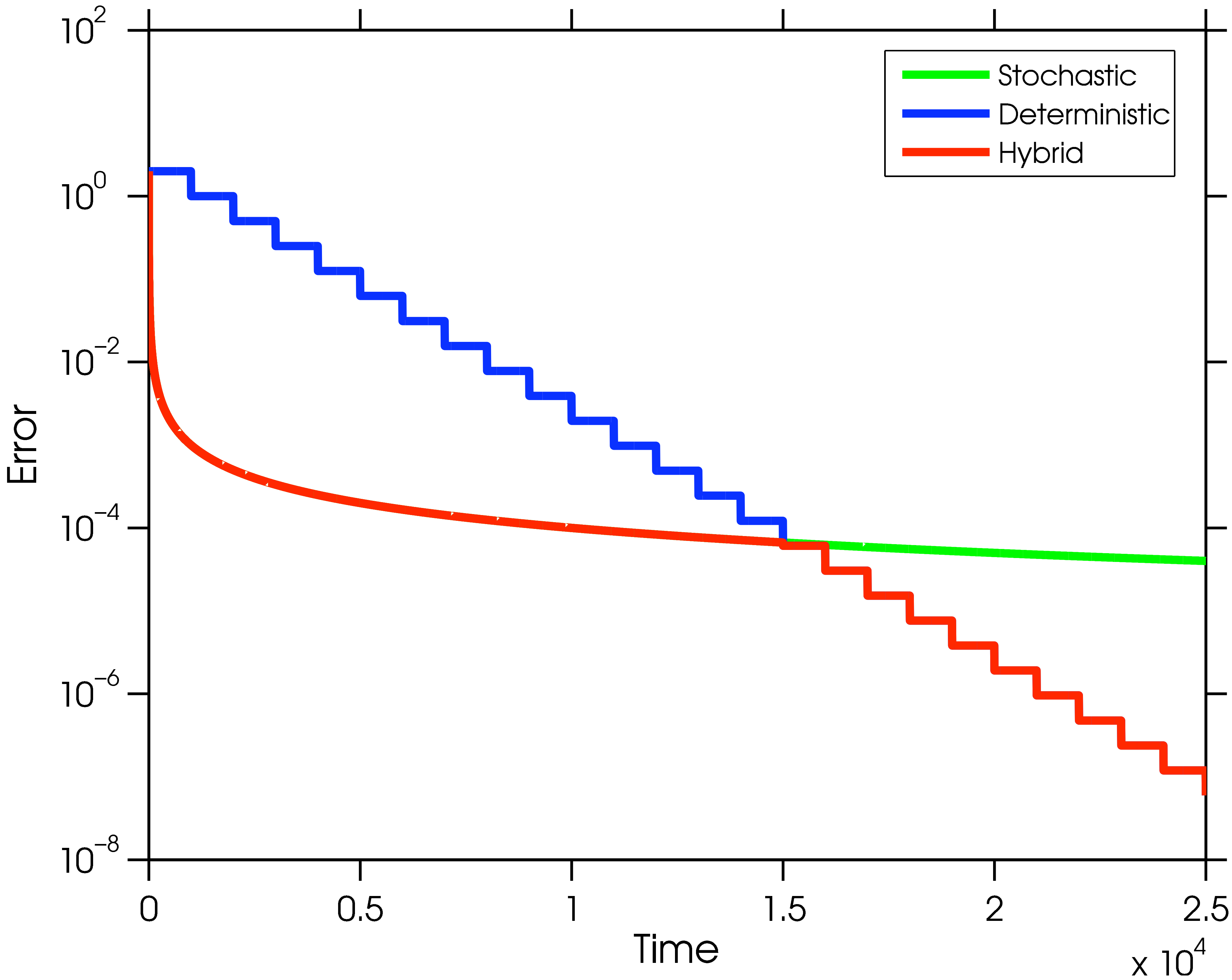
- Stochastic does better initially (*low* ***cost***)
- Deterministic catches up and passes (*low* ***error***)

Simple Hybrid Method

- Can we get the best of both worlds?
- Simple hybrid method:
  - Start with stochastic
  - Switch to deterministic

SLIM

Simple Hybrid Method

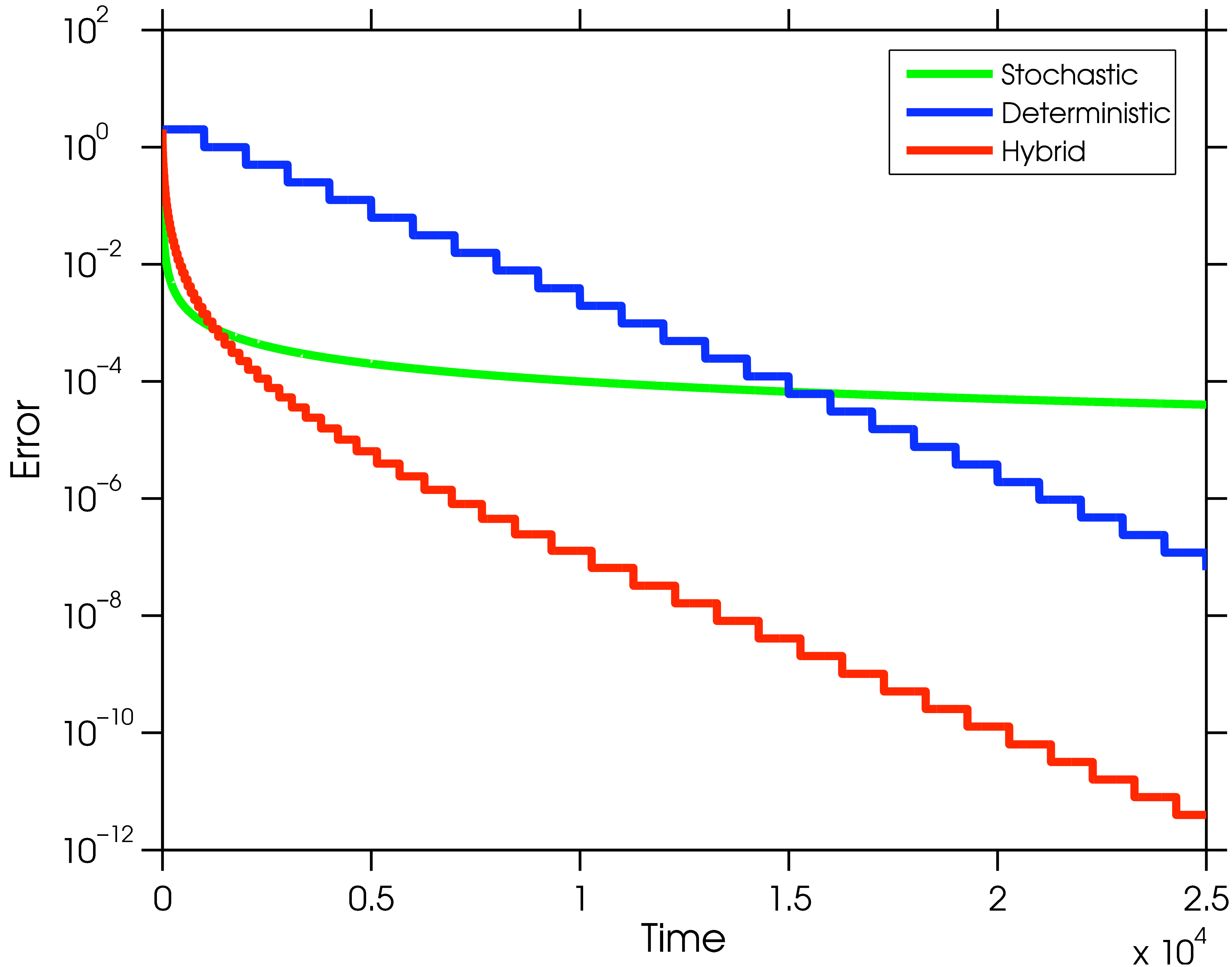


Better Hybrid Methods?

- The question underlying our work:
  - Can a hybrid method do ***better*** than both?

- Basic idea is similar:
  - Start with stochastic
  - Gradually become deterministic

Hybrid Method



Outline

- Stochastic vs. Deterministic Optimization
- **Theoretical Analysis of a Hybrid Method**
- Empirical Analysis of a Hybrid Method
- Future Work

Deterministic Algorithm

- We want to minimize a strongly convex f(x).
- Take gradient steps with a small step size:

$$\mathbf{x}^{k+1} \leftarrow \mathbf{x}^k - (1/M)\nabla f(\mathbf{x}^k).$$

- The convergence rate is ***linear***:

$$f(\mathbf{x}^k) - f(\mathbf{x}^*) \leq (1 - m/M)^k (f(\mathbf{x}^0) - f(\mathbf{x}^*)),$$

Stochastic Algorithm

- Replace the gradient with a noisy version:

$$g(\mathbf{x}^k) \leftarrow \nabla f(\mathbf{x}^k) + \omega^k.$$

- Noise must be zero-mean, finite-variance.
- This might be ***much*** cheaper to compute.
- But the convergence rate is ***sub-linear***:

$$\mathbb{E}[f(\mathbf{x}^k)] - f(\mathbf{x}^*) = \mathcal{O}(1/k).$$

SLIM

Hybrid Algorithm

- Use a gradient with error:

$$g(\mathbf{x}^k) \leftarrow \nabla f(\mathbf{x}^k) + \omega^k.$$

- We don't assume error is unbiased.
- But, we assume we can bound error size:

$$||\omega^k||^2 \leq B^k$$

- Can we achieve a ***linear*** convergence rate?

Hybrid Algorithm

- We have shown that if the bounds satisfy:

$$B^k \leq 2M(m/M - c)d^k$$

- Then you get a ***linear*** convergence rate:

$$f(\mathbf{x}^k) - f(\mathbf{x}^*) \leq (1-c)^k (f(\mathbf{x}^0) - f(\mathbf{x}^*))$$

- Classic result is special case where c = m/M.
- For c < m/M, ***never require the exact gradient***.

Hybrid Algorithm

- Theory agrees with intuition, you can have a big error if either:
  - You are far from the solution (d^k large).
  - The problem is well-behaved (m/M small).

Extensions

- Many assumptions can be relaxed:
  - You only need bounds on m, M, d^k.
  - You only need to bound $\mathbf{E}[||w^k||^2]$.
  - You don't need convexity (for local rates).
  - You don't need smoothness (for proximal).
  - You can analyze other algorithms.

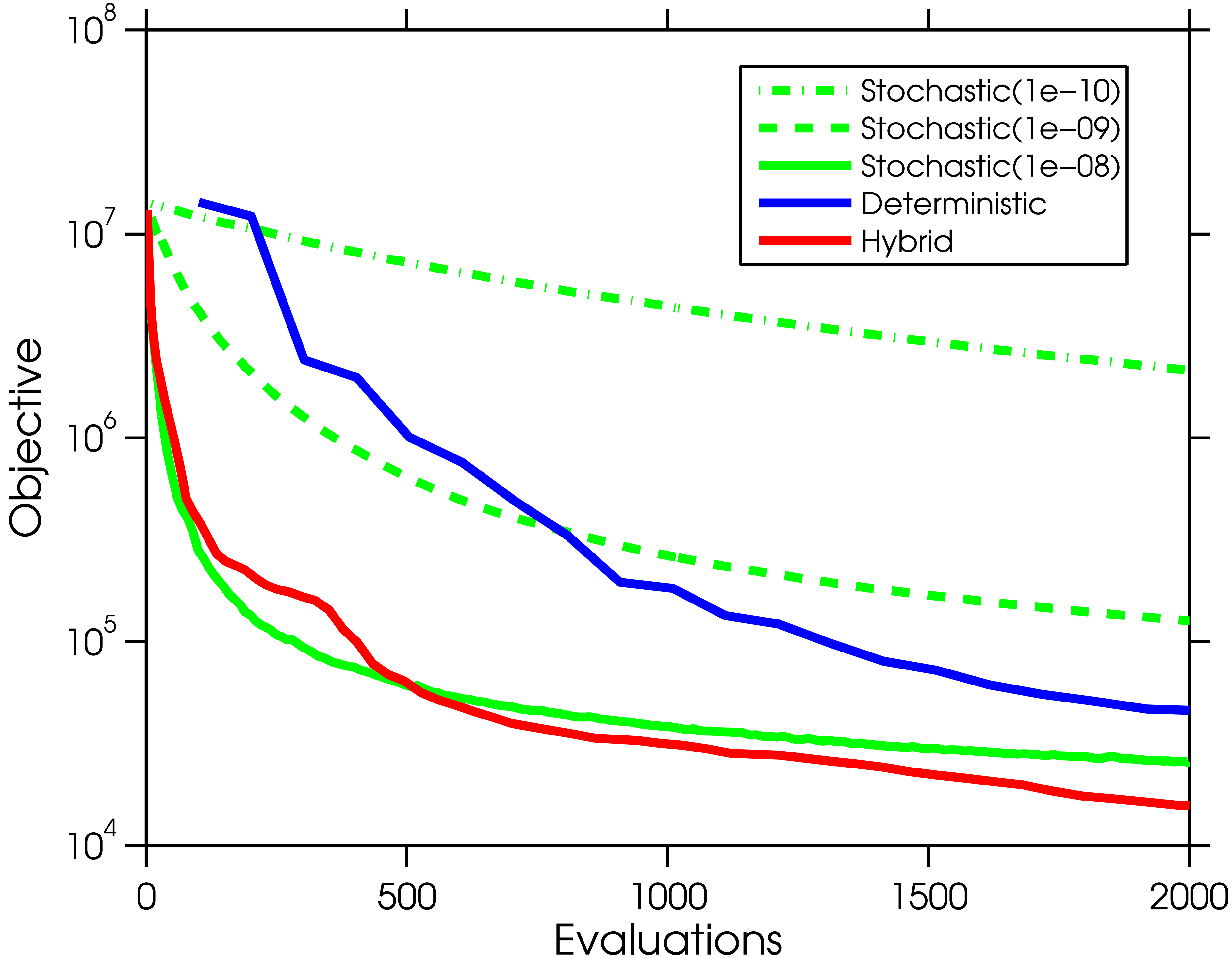
Outline

- Stochastic vs. Deterministic Optimization
- Theoretical Analysis of a Hybrid Method
- **Empirical Analysis of a Hybrid Method**
- Future Work

Cross-Well Tomography

- Synthetic cross-well tomography data set prepared by Tristan van Leeuwen.
- 101 point sources of acoustic energy (7 frequencies) and 101 receivers.
- Deterministic: All 101 sources (L-BFGS)
- Stochastic: 1 random source on each iteration.
- Hybrid: Start with 1 random source, on each iteration randomly add 1 extra source (L-BFGS)

Comparison of Methods



Outline

- Stochastic vs. Deterministic Optimization
- Theoretical Analysis of a Hybrid Method
- Empirical Analysis of a Hybrid Method
- **Future Work**

SLIM

Summary and Future Work

- Summary:
  - Linear convergence of a hybrid method.
  - Promising experimental results.
- Future work is connecting theory to practice:
  - Analysis of more complex algorithms.
  - Ways to control noise.