New Analysis of Adaptive Stochastic Optimization Methods via Supermartingales
Part II: Convergence analysis for stochastic line search

Courtney Paquette
Joint work with Katya Scheinberg

Waterloo

Lehigh University TRIPODS/DIMACS 2018
August 15, 2018
(Deterministic) Backtracking Line Search

Classical problem

$$\min_x f(x)$$

\[ f : \Omega \to \mathbb{R} \text{ is } C^1 \text{ smooth w/ } L\text{-Lipschitz continuous gradient, bounded below} \]

**Gradient descent:** \( x_{k+1} = x_k - \alpha \nabla f(x_k), \quad \alpha \in (0, 1/L] \)
(Deterministic) Backtracking Line Search

Classical problem

\[
\min_x f(x)
\]

\(f : \Omega \rightarrow \mathbb{R}\) is \(C^1\) smooth w/ \(L\)-Lipschitz continuous gradient, bounded below

Gradient descent: \(x_{k+1} = x_k - \alpha \nabla f(x_k), \quad \alpha \in (0, 1/L]\)

Backtracking Line Search Algorithm

- Compute \(f(x_k)\) and \(\nabla f(x_k)\)
- Check sufficient decrease (Armijo ’66)
  \[
  f(x_k - \alpha_k \nabla f(x_k)) \leq f(x_k) - \theta \alpha_k \|\nabla f(x_k)\|^2
  \]
- Successful: \(x_{k+1} = x_k - \alpha_k \nabla f(x_k)\) and \(\alpha_{k+1} = \alpha_k\)
- Unsuccessful: \(x_{k+1} = x_k\) and \(\alpha_k \downarrow\)

Stepsize \(\alpha_k \approx \frac{1}{L}\)
**Convergence Rate**

Sufficient Decrease: \[ f(x_k - \alpha_k \nabla f(x_k)) \leq f(x_k) - \theta \alpha_k \| \nabla f(x_k) \|^2 \]

<table>
<thead>
<tr>
<th>Smoothness</th>
<th>( | \nabla f(x_k) | &lt; \varepsilon )</th>
<th>( f(x_k) - f^* &lt; \varepsilon )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L )-smooth</td>
<td>( \frac{L}{\varepsilon^2} )</td>
<td>-</td>
</tr>
<tr>
<td>( L )-smooth/convex</td>
<td>( \frac{L}{\varepsilon} )</td>
<td>( \frac{L}{\varepsilon} )</td>
</tr>
<tr>
<td>( \alpha )-convex</td>
<td>( \frac{L}{\alpha} \cdot \log(\frac{1}{\varepsilon}) )</td>
<td>( \frac{L}{\alpha} \cdot \log(\frac{1}{\varepsilon}) )</td>
</tr>
</tbody>
</table>
Stochastic Line Search Question

Stochastic problem

\[
\min_{x \in \mathbb{R}^n} f(x) = \mathbb{E}_{\xi}[\tilde{f}(x; \xi)], \quad \xi \text{ is a random variable}
\]

Examples

- *Empirical risk minimization*: \(\xi_i\) is a uniform r.v. over training set
- *More generally*: \(\xi\) is any sample or set of samples from data distribution
Stochastic Line Search Question

Stochastic problem

$$\min_{x \in \mathbb{R}^n} f(x) = \mathbb{E}_\xi[\tilde{f}(x; \xi)], \quad \xi \text{ is a random variable}$$

Examples

- *Empirical risk minimization*: $\xi_i$ is a uniform r.v. over training set
- *More generally*: $\xi$ is any sample or set of samples from data distribution

(Stochastic) Backtracking Line Search Algorithm

- Compute stochastic estimates $\nabla f(x_k)$, $f(x_k)$, $f_0$, and $f(x_k - \alpha_k g_k)$
- Check sufficient decrease (Armijo ’66)
  $$f^s_k \leq f^0_k - \theta \alpha_k \|g_k\|^2$$
- Successful: $x_{k+1} = x_k - \alpha_k g_k$ and $\alpha_k \uparrow$
- Unsuccessful: $x_{k+1} = x_k$ and $\alpha_k \downarrow$

(Friedlander-Schmidt ’12; Mahsereci-Hennig ’17, ... )
\[ f^s_k \leq f^0_k - \theta \alpha_k \| g_k \|^2 \Rightarrow f(x_k - \alpha_k g_k) \leq f(x_k) - \theta \alpha_k \| g_k \|^2 \]

**Simple example:** Know exact function values

\[ f(x_{k+1}) \leq f(x_k) \]
\[ f_k^s \leq f_k^0 - \theta \alpha_k \| g_k \|^2 \Rightarrow f(x_k - \alpha_k g_k) \leq f(x_k) - \theta \alpha_k \| g_k \|^2 \]

**Simple example:** Know exact function values

\[ f(x_{k+1}) \leq f(x_k) \]

**Challenges**

- Bad function estimates may \( \uparrow \) objective value

Increase- \( \alpha_k^2 \| g_k \|^2 \)
\[ f_k^s \leq f_k^0 - \theta \alpha_k \|g_k\|^2 \Rightarrow f(x_k - \alpha_k g_k) \leq f(x_k) - \theta \alpha_k \|g_k\|^2 \]

**Simple example:** Know exact function values

\[ f(x_{k+1}) \leq f(x_k) \]

**Challenges**

- Bad function estimates may ↑ objective value
  
  \[ \text{Increase- } \alpha_k^2 \|g_k\|^2 \]

- Stepsizes, \( \alpha_k \), become arbitrarily small
\[ f^s_k \leq f^0_k - \theta \alpha_k \|g_k\|^2 \Rightarrow f(x_k - \alpha_k g_k) \leq f(x_k) - \theta \alpha_k \|g_k\|^2 \]

**Simple example:** Know exact function values

\[ f(x_{k+1}) \leq f(x_k) \]

**Challenges**

- Bad function estimates may \( \uparrow \) objective value
  
  Increase- \( \alpha_k^2 \|g_k\|^2 \)

- Stepsizes, \( \alpha_k \), become arbitrarily small

**Question**

Devise a line search for the stochastic problem with provable convergence guarantees using only knowable quantities.

**Knowable quantities:** e.g. bound on variance of \( \nabla \tilde{f}, \tilde{f} \)

(Bollapragada et al ’17, Cartis-Scheinberg ’17)
Proposed stochastic line search

Algorithm

- Compute random estimate of the gradient, $g_k$
- Compute random estimate of $f_k^0 \approx f(x_k)$ and $f_k^s \approx f(x_k - \alpha_k g_k)$
- Check the stochastic sufficient decrease

$$f_k^s \leq f_k^0 - \theta \alpha_k \|g_k\|^2$$

- Successful: $x_{k+1} = x_k - \alpha_k g_k$ and $\alpha_k \uparrow$

- Reliable step: If $\alpha_k \|g_k\|^2 \geq \delta_k^2$, $\uparrow \delta_k$
- Unreliable step: If $\alpha_k \|g_k\|^2 < \delta_k^2$, $\downarrow \delta_k$

- Unsuccessful: $x_{k+1} = x_k$, $\alpha_k \downarrow$, and $\delta_k \downarrow$
What is $\delta_k$?

Bad function estimates may $\uparrow$ objective value

$$\alpha_k \|g_k\|$$

$\delta \approx$ prediction of the size of $\alpha_k \|g_k\|$

$\approx$ size of a “trust region”

$\Rightarrow$ Largest $\uparrow$ in objective is at most $\delta_k^2$

- **Reliable step:** If $\alpha_k \|g_k\|^2 \geq \delta_k^2$, $\uparrow\delta_k$
- **Unreliable step:** If $\alpha_k \|g_k\|^2 < \delta_k^2$, $\downarrow\delta_k$
Stochastic Line Search

**Algorithm**

- Compute **random** estimate of the gradient, $g_k$
- Compute **random** estimate of $f_k^0 \approx f(x_k)$ and $f_k^s \approx f(x_k - \alpha_k g_k)$
- Check the **stochastic** sufficient decrease

$$f_k^s \leq f_k^0 - \theta \alpha_k \|g_k\|^2$$

- **Successful:** $x_{k+1} = x_k - \alpha_k g_k$ and $\alpha_k \uparrow$

- **Reliable step:** If $\alpha_k \|g_k\|^2 \geq \delta_k^2$, $\uparrow \delta_k$
- **Unreliable step:** If $\alpha_k \|g_k\|^2 < \delta_k^2$, $\downarrow \delta_k$

- **Unsuccessful:** $x_{k+1} = x_k$, $\alpha_k \downarrow$, and $\delta_k \downarrow$
Randomness assumptions

- **Accurate gradient** $G_k$ w/ prob. $p_g$:

  $$\Pr(\|G_k - \nabla f(X_k)\| \leq \kappa_g A_k \|G_k\| | \text{past}) \geq p_g$$
Randomness assumptions

- **Accurate gradient** $G_k$ w/ prob. $p_g$:
  \[
  \Pr \left( \|G_k - \nabla f(X_k)\| \leq \kappa_g A_k \|G_k\| \mid \text{past} \right) \geq p_g
  \]

- **Accurate function estimates** $F^0_k$ and $F^s_k$ w/ prob. $p_f$:
  \[
  \Pr \left( |f(X_k) - F^0_k| \leq \varepsilon_f A_k^2 \|G_k\|^2 \right)
  \]
  \[
  \text{and} \quad |f(X_k - A_k G_k) - F^s_k| \leq \varepsilon_f A_k^2 \|G_k\|^2 \mid \text{past} \right) \geq p_f
  \]
Randomness assumptions

- **Accurate gradient** $G_k$ w/ prob. $p_g$:
  \[ \Pr(\|G_k - \nabla f(X_k)\| \leq \kappa_g A_k \|G_k\| | \text{past}) \geq p_g \]

- **Accurate function estimates** $F^0_k$ and $F^s_k$ w/ prob. $p_f$:
  \[ \Pr(|f(X_k) - F^0_k| \leq \varepsilon_f A_k^2 \|G_k\|^2) \]
  \[ \text{and} \quad |f(X_k - A_k G_k) - F^s_k| \leq \varepsilon_f A_k^2 \|G_k\|^2 | \text{past}) \geq p_f \]

- **Variance condition**
  \[ \mathbb{E}[|F^0_k - F(X_k)|^2 | \text{past}] \leq \theta^2 \Delta^4_k \quad (\text{same for } F^s_k). \]

Want to choose these probabilities $(p_f, p_g)$ large enough
Randomness assumptions

- **Accurate gradient** $G_k$ w/ prob. $p_g$:
  \[
  \Pr\left(\|G_k - \nabla f(X_k)\| \leq \kappa_g A_k \|G_k\| \mid \text{past} \right) \geq p_g
  \]

- **Accurate function estimates** $F^0_k$ and $F^s_k$ w/ prob. $p_f$:
  \[
  \Pr\left(|f(X_k) - F^0_k| \leq \varepsilon_f A_k^2 \|G_k\|^2 \mid \text{past} \right) \geq p_f
  \]
  and
  \[
  |f(X_k - A_k G_k) - F^s_k| \leq \varepsilon_f A_k^2 \|G_k\|^2 \mid \text{past} \right) \geq p_f
  \]

- **Variance condition**
  \[
  \mathbb{E}[|F^0_k - F(X_k)|^2 \mid \text{past}] \leq \theta^2 \Delta_k^4
  \]
  (same for $F^s_k$).

Want to choose these probabilities $(p_f, p_g)$ large enough

$p_f, p_g \geq 1/2$ at least, but $p_f$ should be large.
Satisfying randomness assumptions

\[ \min_{x \in \mathbb{R}^n} f(x) = \mathbb{E}_\xi [\tilde{f}(x; \xi)] \]

and bound on variance

\[ \mathbb{E}(\| \nabla \tilde{f}(x, \xi_i) - \nabla f(x) \|^2) \leq V_g, \quad \mathbb{E}(|\tilde{f}(x; \xi_i) - f(x)|^2) \leq V_f. \]
Satisfying randomness assumptions

\[
\min_{x \in \mathbb{R}^n} f(x) = \mathbb{E}_\xi[\tilde{f}(x; \xi)]
\]

and bound on variance

\[
\mathbb{E}(\|\nabla \tilde{f}(x, \xi_i) - \nabla f(x)\|^2) \leq V_g, \quad \mathbb{E}(|\tilde{f}(x; \xi_i) - f(x)|^2) \leq V_f.
\]

Example: sampling

\[
g_k = \frac{1}{|S_g|} \sum_{i \in S_g} \nabla f(x_k; \xi_i), \quad f_k^0 = \frac{1}{|S_f|} \sum_{i \in S_f} f(x_k; \xi_i).
\]

How many samples do we need?
Satisfying randomness assumptions

\[
\min_{x \in \mathbb{R}^n} f(x) = \mathbb{E}_\xi [\tilde{f}(x; \xi)]
\]

and bound on variance

\[
\mathbb{E}(\|\nabla \tilde{f}(x, \xi_i) - \nabla f(x)\|^2) \leq V_g, \quad \mathbb{E}(|\tilde{f}(x; \xi_i) - f(x)|^2) \leq V_f.
\]

Example: sampling

\[
g_k = \frac{1}{|S_g|} \sum_{i \in S_g} \nabla f(x_k; \xi_i), \quad f^0_k = \frac{1}{|S_f|} \sum_{i \in S_f} f(x_k; \xi_i).
\]

How many samples do we need?

Idea: Chebyshev Inequality

\[
|S_g| \approx \tilde{O} \left( \frac{V_g}{A_k^2 \|G_k\|^2} \right), \quad |S_f| \approx \tilde{O} \left( \max \left\{ \frac{V_f}{A_k^4 \|G_k\|^4}, \frac{V_f}{\theta^2 \Delta_k^4} \right\} \right).
\]
### Dynamics of the stepsize

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>Stochastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_k \leq 1/L \Rightarrow$ successful step</td>
<td>Good gradient/function estimates &amp; stepsize $\leq 1/L$, $\Rightarrow$ success</td>
</tr>
</tbody>
</table>
Dynamics of the stepsize

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>Stochastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_k \leq 1/L \Rightarrow ) successful step</td>
<td>Good gradient/function estimates &amp; stepsize ( \leq 1/L ), ( \Rightarrow ) success</td>
</tr>
<tr>
<td>( \alpha_k ) bounded from 0</td>
<td></td>
</tr>
</tbody>
</table>

When \( \alpha_k \lesssim 1/L \), \( \alpha_k \) move ↑ and ↓ like random walk with probability \( p_g p_f \)
Dynamics of the stepsize

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>Stochastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_k \leq 1/L \Rightarrow$ successful step</td>
<td>Good gradient/function estimates &amp; stepsize $\leq 1/L$, $\Rightarrow$ success</td>
</tr>
<tr>
<td>$\alpha_k$ bounded from 0</td>
<td>$\Pr(\lim \sup_k A_k &gt; 0) = 1$</td>
</tr>
</tbody>
</table>

When $\alpha_k \lesssim 1/L$, $\alpha_k$ move ↑ and ↓ like random walk with probability $p_g p_f$
## Probability Viewpoint

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>Stochastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_k \leq 1/L \Rightarrow$ successful step</td>
<td>Good gradient/function estimates &amp; stepsize $\leq 1/L$, $\Rightarrow$ success</td>
</tr>
</tbody>
</table>
## Probability Viewpoint

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>Stochastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_k \leq 1/L \Rightarrow \text{successful step} )</td>
<td>Good gradient/function estimates &amp; stepsize ( \leq 1/L ), ( \Rightarrow ) success</td>
</tr>
<tr>
<td>( \alpha_k ) bounded from 0</td>
<td>( \Pr(\limsup_k A_k &gt; 0) = 1 )</td>
</tr>
</tbody>
</table>
## Probability Viewpoint

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>Stochastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_k \leq 1/L \Rightarrow$ successful step</td>
<td>Good gradient/function estimates &amp; stepsize $\leq 1/L$, $\Rightarrow$ success</td>
</tr>
<tr>
<td>$\alpha_k$ bounded from 0</td>
<td>$\Pr(\limsup_k A_k &gt; 0) = 1$</td>
</tr>
<tr>
<td>Function values decrease each iteration</td>
<td>$\Phi_k \approx f(X_k) - f^*$ such that $\mathbb{E}[\Phi_{k+1} - \Phi_k</td>
</tr>
</tbody>
</table>
## Probability Viewpoint

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>Stochastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_k \leq 1/L \Rightarrow$ successful step</td>
<td>Good gradient/function estimates &amp; stepsize $\leq 1/L$, $\Rightarrow$ success</td>
</tr>
<tr>
<td>$\alpha_k$ bounded from 0</td>
<td>$\mathbb{P}(\limsup_k A_k &gt; 0) = 1$</td>
</tr>
<tr>
<td>Function values decease each iteration</td>
<td>$\Phi_k \approx f(X_k) - f^*$ such that $\mathbb{E}[\Phi_{k+1} - \Phi_k</td>
</tr>
</tbody>
</table>

Convergence rate: number of iterations until nearly optimal (e.g. $\|\nabla f(x)\| < \varepsilon$, $f(x) - f^* < \varepsilon$) |

Convergence rate $\Rightarrow$ stopping times e.g. $T = \inf\{k > 0 : \|\nabla f(X_k)\| < \varepsilon\}$, $T = \inf\{k > 0 : f(X_k) - f^* < \varepsilon\}$

Interested in $\mathbb{E}[T]$
Renewal and reward process

Random process \( \{ \Phi_k, A_k, W_k \} \)

- \( \Phi_k \in [0, \infty) \) and \( A_k \in [0, \infty) \)
- \( W_k \) biased random walk with probability \( p > 1/2 \)

\[
\Pr(W_{k+1} = 1 | \text{past}) = p \quad \text{and} \quad \Pr(W_{k+1} = -1 | \text{past}) = 1 - p.
\]
Renewal and reward process

Random process \( \{\Phi_k, A_k, W_k\} \)

- \( \Phi_k \in [0, \infty) \) and \( A_k \in [0, \infty) \)
- \( W_k \) biased random walk with probability \( p > 1/2 \)

\[
\Pr(W_{k+1} = 1| \text{past}) = p \quad \text{and} \quad \Pr(W_{k+1} = -1| \text{past}) = 1 - p.
\]

Assumptions

(i) \( \exists \bar{A} \) with

\[
A_{k+1} \geq \min \left\{ A_k e^{\lambda W_{k+1}}, \bar{A} \right\}
\]

(ii) \( \exists \) nondecreasing \( h : [0, \infty) \to (0, \infty) \) and constant \( \Theta \) s.t.

\[
\mathbb{E}[\Phi_{k+1}| \text{past}] \leq \Phi_k - \Theta h(A_k).
\]
Renewal and reward process

Random process \( \{\Phi_k, A_k, W_k\} \)
- \( \Phi_k \in [0, \infty) \) and \( A_k \in [0, \infty) \)
- \( W_k \) biased random walk with probability \( p > 1/2 \)

\[
\Pr(W_{k+1} = 1|\text{past}) = p \quad \text{and} \quad \Pr(W_{k+1} = -1|\text{past}) = 1 - p.
\]

Assumptions

(i) \( \exists \bar{A} \) with

\[
A_{k+1} \geq \min \left\{ A_k e^{\lambda W_{k+1}}, \bar{A} \right\}
\]

(ii) \( \exists \) nondecreasing \( h : [0, \infty) \rightarrow (0, \infty) \) and constant \( \Theta \) s.t.

\[
\mathbb{E}[\Phi_{k+1}|\text{past}] \leq \Phi_k - \Theta h(A_k).
\]

Thm: (Blanchet, Cartis, Menickelly, Scheinberg ’17)

\[
\mathbb{E}[T_{\varepsilon}] \leq \frac{p}{2p - 1} \cdot \frac{\Phi_0}{\Theta h(\bar{A})} + 1.
\]
Convergence Result: Line search

Key observation

\[ \Phi_k = \nu(f(x_k) - f_{\text{min}}) + (1 - \nu)\alpha_k \| \nabla f(x_k) \|^2 + (1 - \nu)\theta \delta_k^2 \]

\[ \Rightarrow \Phi_{k+1} - \Phi_k = \nu(f(x_{k+1}) - f(x_k)) \]
\[ + (1 - \nu) \left( \alpha_{k+1} \| \nabla f(x_{k+1}) \|^2 - \alpha_k \| \nabla f(x_k) \|^2 \right) \]
\[ + (1 - \nu)\theta (\delta_{k+1}^2 - \delta_k^2) \]
Convergence Result: Line search

Key observation

\[ \Phi_k = \nu(f(x_k) - f_{\text{min}}) + (1 - \nu)\alpha_k \| \nabla f(x_k) \|^2 + (1 - \nu)\theta \delta_k^2 \]

\[ \Rightarrow \Phi_{k+1} - \Phi_k = \nu(f(x_{k+1}) - f(x_k)) \\
+ (1 - \nu) \left( \alpha_{k+1} \| \nabla f(x_{k+1}) \|^2 - \alpha_k \| \nabla f(x_k) \|^2 \right) \\
+ (1 - \nu)\theta(\delta_{k+1}^2 - \delta_k^2) \]

Thm: (P-Scheinberg ’18) If

\[ p_g p_f > 1/2 \quad \text{and} \quad p_f \text{ sufficiently large,} \]

\[ \mathbb{E}[\Phi_{k+1} - \Phi_k | \text{past}] \leq - \left( A_k \| \nabla f(X_k) \|^2 + \theta \Delta_k^2 \right) \]

Proof Idea:

- accurate gradient + accurate function estimates \( \Rightarrow \Phi_k \) always \( \downarrow \)
- all other cases \( \Phi_k \) \( \uparrow \) by same amount
Convergence result, nonconvex

Stopping Time

\[ T = \inf \{ k : \| \nabla f(x_k) \| < \varepsilon \} \]

Convergence rate, nonconvex (P-Scheinberg ’18)

If \( p_g p_f > 1/2 \) and \( p_f \) sufficiently large,

\[ \mathbb{E}[T] \leq O \left( \frac{1}{\varepsilon^2} \right). \]
Convex case

\[
\min_{x \in \Omega} f(x) = \mathbb{E}[\tilde{f}(x, \xi)]
\]

where

- \( f \) is convex and \( \|\nabla f(x)\| \leq L_f \) for all \( x \in \Omega \)
- \( \|x - x^*\| \leq D \) for all \( x \in \Omega \)

Stopping time: \( T = \inf\{k : f(x_k) - f^* < \varepsilon\} \)
Convex case

\[\min_{x \in \Omega} f(x) = \mathbb{E}[\tilde{f}(x, \xi)]\]

where

- \(f\) is convex and \(\|\nabla f(x)\| \leq L_f\) for all \(x \in \Omega\)
- \(\|x - x^*\| \leq D\) for all \(x \in \Omega\)

Stopping time: \(T = \inf\{k : f(x_k) - f^* < \varepsilon\}\)

Key observation:

\[\Psi_k = \frac{1}{\nu \varepsilon} - \frac{1}{\Phi_k}\]

(Convergence rate, convex) (P-Scheinberg ’18)

If \(p_g p_f > 1/2\) and \(p_f\) sufficiently large,

\[\mathbb{E}[T] \leq O\left(\frac{1}{\varepsilon}\right)\]
Strongly convex case

\[ \min_{x \in \Omega} f(x) = \mathbf{E}[\tilde{f}(x, \xi)] \]

where \( f \) is \( \mu \)-strongly convex

Stopping Time: \( T = \inf \{ k : f(x_k) - f^* < \varepsilon \} \)
Strongly convex case

$$\min_{x \in \Omega} f(x) = \mathbb{E}[\tilde{f}(x, \xi)]$$

where $f$ is $\mu$-strongly convex

Stopping Time:

$$T = \inf\{ k : f(x_k) - f^* < \varepsilon \}$$

Key observation:

$$\Psi_k = \log(\Phi_k) - \log(\nu \varepsilon)$$

Convergence rate, strongly convex (P-Scheinberg ’18)

If $p_g p_f > 1/2$ and $p_f$ sufficiently large,

$$\mathbb{E}[T] \leq \mathcal{O} \left( \log \left( \frac{1}{\varepsilon} \right) \right)$$
Thank You
References

A Stochastic Line Search Method with Convergence Rate Analysis.