High Probability Complexity Bounds for Trust-Region Methods with Noisy Oracles

Liyuan Cao 曹立元

Beijing International Center for Mathematical Research, Peking University
北京大学北京国际数学研究中心

April 8, 2023

Collaborators

Albert S. Berahas
University of Michigan

Katya Scheinberg
Cornell University

Liyuan Cao, Albert S. Berahas, and Katya Scheinberg. First-and second-order high probability complexity bounds for trust-region methods with noisy oracles. *arXiv preprint arXiv:2205.03667*, 2022.

The Problem

$$\min_{x \in \mathbb{R}^n} \phi(x)$$

ϕ follows common assumptions

$$\phi(x) \geq \hat{\phi} \text{ for all } x \in \mathbb{R}^n,$$
$$\|\nabla\phi(x) - \nabla\phi(y)\| \leq L_1\|x - y\| \text{ for all } (x, y) \in \mathbb{R}^n \times \mathbb{R}^n,$$

but we only have access to

$$\begin{cases} f_k \\ g_k \\ H_k \end{cases} \text{instead of} \begin{cases} \phi(x_k) \\ \nabla\phi(x_k) \\ \nabla^2\phi(x_k). \end{cases}$$

The Problem

$$\min_{x \in \mathbb{R}^n} \phi(x)$$

ϕ follows common assumptions

$$\phi(x) \geq \hat{\phi} \text{ for all } x \in \mathbb{R}^n,$$
$$\|\nabla \phi(x) - \nabla \phi(y)\| \leq L_1 \|x - y\| \text{ for all } (x, y) \in \mathbb{R}^n \times \mathbb{R}^n,$$

but we only have access to

$$\begin{cases} f_k \\ g_k \\ H_k \end{cases} \text{instead of} \begin{cases} \phi(x_k) \\ \nabla \phi(x_k) \\ \nabla^2 \phi(x_k). \end{cases}$$

A line of work:

- **algorithm** TR method modified to handle noise.
- **noise** Weaker assumptions in more recent work.
- **result** Stronger results in more recent work.

The Trust-Region Method

Algorithm: Modified First-Order Trust-Region Method

Inputs: Starting point x_0, initial trust region radius δ_0, tolerance parameter r, and hyperparameters $\eta_1 > 0, \eta_2 > 0$, $\gamma \in (0,1)$ for controlling the trust region radius.

for $k = 0, 1, 2, \ldots$ **do**

Build a quadratic model $m_k(x_k + s) = \phi(x_k) + \langle g_k, s\rangle + 0.5\,\langle H_k s, s\rangle$

Compute s_k by approximately minimizing m_k in $B(x_k, \delta_k)$ so that it satisfies the *Cauchy decrease condition*

$$m_k(x_k) - m_k(x_k + s_k) \geq \frac{1}{2}\|g_k\| \min\left\{\frac{\|g_k\|}{\|H_k\|}, \delta_k\right\}.$$

Compute

$$\rho_k = \frac{f_k - f_k^+ + r}{m_k(x_k) - m_k(x_k + s_k)}$$

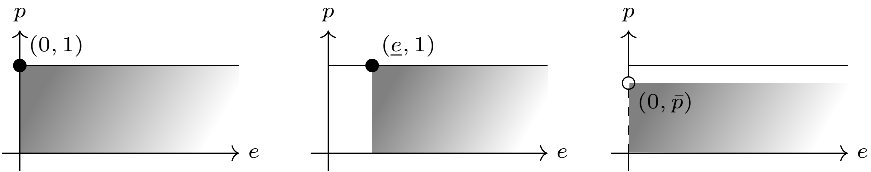
and update x and δ

$$(x_{k+1}, \delta_{k+1}) = \begin{cases} (x_k + s_k, \gamma^{-1}\delta_k) & \text{if } \rho_k \geq \eta_1 \text{ and } \|g_k\| \geq \eta_2\delta_k \\ (x_k + s_k, \gamma\delta_k) & \text{if } \rho_k \geq \eta_1 \text{ and } \|g_k\| < \eta_2\delta_k \\ (x_k, \gamma\delta_k) & \text{if } \rho_k < \eta_1. \end{cases}$$

Stochastic Oracle

Let $\varphi^{(j)}\left(x_k, \xi_k^{(j)}, \mathcal{S}_k^{(j)}\right)$ be the jth-order oracle that returns an estimate of $\nabla^j \phi(x_k)$ such that for all $(e, p) \in \mathcal{S}_k^{(j)} \subseteq [0, \infty) \times [0, 1]$,

$$\mathbb{P}_{\xi_k^{(j)}}\left\{\left\|\varphi^{(j)}\left(x_k, \xi_k^{(j)}, \mathcal{S}^{(j)}\right) - \nabla^j \phi(x_k)\right\| \leq e \,\middle|\, \mathcal{F}_k\right\} \geq p$$



(a) $\mathcal{S}_k^{(j)} \ni (0, 1)$
exact

(b) $\mathcal{S}_k^{(j)} = [\underline{e}, +\infty) \times [0, 1]$
bounded

(c) $\mathcal{S}_k^{(j)} = (0, +\infty) \times [0, \bar{p}]$
probabilistically sufficiently accurate

The Goal of the Analysis

ⓒ convergence

$$\liminf_{k\to\infty} \|\nabla\phi(x_k)\| \le \epsilon$$

ⓔ expected complexity

$$\mathbb{E}\min\{k : \|\nabla\phi(X_k)\| \le \epsilon\} = \mathcal{O}(1/\epsilon^2)$$

ⓗ high probability convergence

$$\mathbb{P}\left\{\min\{\|\nabla\phi(X_k)\| :\ 0 \le k \le T-1\} < \epsilon\right\}$$
$$\ge \text{a function of } T \text{ the converges to 1 as } T \text{ increase}$$

for some sufficiently large ϵ.

$\mathcal{S}^{(0)} = [0, \infty) \times [0, 1]$ and $\mathcal{S}^{(1)} = (0, \infty) \times [0, \bar{p}_1]$ for sufficiently large $\bar{p}_1$:

- ⓒ Afonso S Bandeira, Katya Scheinberg, and Luis Nunes Vicente. Convergence of trust-region methods based on probabilistic models. *SIAM Journal on Optimization*, 24(3):1238–1264, 2014.
- ⓗ Serge Gratton, Clement W Royer, Luis N Vicente, and Zaikun Zhang. Complexity and global rates of trust-region methods base on probabilistic models. *IMA Journal of Numerical Analysis*, 38(3):1579-1597, 2018.

$\mathcal{S}^{(j)} = (0, \infty) \times [0, \bar{p}_j]$ for sufficiently large $\bar{p}_j$, $j = 0, 1$:

- ⓒ Ruobing Chen, Matt Menickelly, and Katya Scheinberg. Stochastic optimization using a trust-region method and random models. *Mathematical Programming*, 169(2):447–487, 2018.

$\mathcal{S}^{(j)} = (0, \infty) \times [0, \bar{p}_j]$ for sufficiently large $\bar{p}_j$, $j = 0, 1, 2$ and $\mathbb{E}_{\xi_0}|f_k - \phi(x_k)| \leq C_0$:

- ⓔ Jose Blanchet, Coralia Cartis, Matt Menickelly, and Katya Scheinberg. Convergence rate analysis of a stochastic trust-region method via supermartingales. *INFORMS Journal on Optimization*, 1(2):92–119, 2019.

$\mathcal{S}^{(0)} = [\epsilon_f, \infty) \times [0, 1]$ and $\mathcal{S}^{(1)} = [\epsilon_g, \infty) \times [0, 1]$:

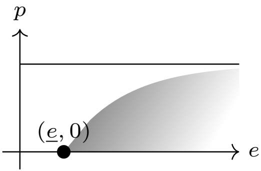
- ⓔ Shigeng Sun and Jorge Nocedal. A trust region method for the optimization of noisy functions. *arXiv preprint arXiv:2201.00973*, 2022.

We assume for all k:

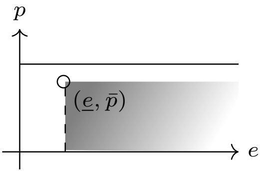
$$\left.\begin{array}{l}\mathbb{P}\left\{|f_k-\phi(x_k)|\le e\right\}\\ \mathbb{P}\left\{|f_k^+-\phi(x_k+s_k)|\le e\right\}\end{array}\right\}\ge \exp(a(\epsilon_f-e)) \qquad \text{unbounded noise,}$$

$$\mathbb{P}\left\{\|g_k-\nabla\phi(x_k)\|\le \kappa_{\text{eg}}\delta_k+\epsilon_g\right\}\ge p_1 \qquad \text{irreducible noise,}$$

$\|H_k\|\le\kappa_{\text{bhm}}$ for some constant κ_{bhm} (bound on hessian of model).



(a) $\mathcal{S}_k^{(0)}=\{(e,p):e\ge\underline{e}=\epsilon_f, p\le 1-\exp(a(\underline{e}-e))\}$



(b) $\mathcal{S}_k^{(1)}=(\underline{e},+\infty)\times[0,\bar{p}]$ with $\underline{e}=\kappa_{\text{eg}}\delta_k+\epsilon_g$ and $\bar{p}=p_1$

The Trust-Region Method

Algorithm: Modified First-Order Trust-Region Method

Inputs: Starting point x_0, initial trust region radius δ_0, tolerance parameter r, and hyperparameters $\eta_1 > 0, \eta_2 > 0$, $\gamma \in (0,1)$ for controlling the trust region radius.

for $k = 0, 1, 2, \dots$ **do**

Build a quadratic model $m_k(x_k + s) = \phi(x_k) + \langle g_k, s \rangle + 0.5 \langle H_k s, s \rangle$

Compute s_k by approximately minimizing m_k in $B(x_k, \delta_k)$ so that it satisfies the *Cauchy decrease condition*

$$m_k(x_k) - m_k(x_k + s_k) \geq \frac{1}{2} \|g_k\| \min\left\{ \frac{\|g_k\|}{\|H_k\|}, \delta_k \right\}.$$

Compute

$$\rho_k = \frac{f_k - f_k^+ + r}{m_k(x_k) - m_k(x_k + s_k)}$$

and update x and δ

$$(x_{k+1}, \delta_{k+1}) = \begin{cases} (x_k + s_k, \gamma^{-1}\delta_k) & \text{if } \rho_k \geq \eta_1 \text{ and } \|g_k\| \geq \eta_2 \delta_k \\ (x_k + s_k, \gamma \delta_k) & \text{if } \rho_k \geq \eta_1 \text{ and } \|g_k\| < \eta_2 \delta_k \\ (x_k, \gamma \delta_k) & \text{if } \rho_k < \eta_1. \end{cases}$$

random variables:	X_k	X_k^+	$\mathcal{E}_k$	$\mathcal{E}_k^+$	
realizations:	x_k	$x_k + s_k$	$\lvert f_k - \phi(x_k)\rvert$	$\lvert f_k^+ - \phi(x_k + s_k)\rvert$	
random variables:	M_k	∇M_k	$\nabla^2 M_k$	Δ_k	ρ_k
realizations:	m_k	g_k	H_k	δ_k	ρ_k

Define

$$I_k = \mathbb{1}\{\|\nabla M_k - \nabla\phi(X_k)\| \le \kappa_{\text{eg}}\Delta_k + \epsilon_g\} \qquad \text{gradient sufficiently accurate}$$

$$J_k = \mathbb{1}\{\mathcal{E}_k + \mathcal{E}_k^+ \le r\} \qquad \text{zeroth-order noise compensated}$$

$$\Lambda_k = \mathbb{1}\{\Delta_k > \bar{\Delta}\} \qquad \text{large TR radius}$$

$$\Theta_k = \mathbb{1}\{\rho_k \ge \eta_1 \text{ and } \|\nabla M_k\| \ge \eta_2\Delta_k\} \qquad \text{successful step}$$

$$\Theta_k' = \mathbb{1}\{\rho_k \ge \eta_1\} \qquad \text{accepted step}$$

where $\bar{\Delta} = C_1 \min_{0\le k\le T-1} \|\nabla\phi(X_k)\| - C_2\epsilon_g$.

Classification of Iterations

<table>
<tr><th></th><th colspan="3">$I_k = 1, J_k = 1$</th><th colspan="3">$I_k = 1, J_k = 0$</th><th colspan="3">$I_k = 0, J_k = 1$</th><th colspan="3">$I_k = 0, J_k = 0$</th></tr>
<tr><th></th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th></tr>
<tr><td>$\Delta_k \in (\bar{\Delta}, \infty)$</td><td>1</td><td>4</td><td>5</td><td>6</td><td>9</td><td>11</td><td>13</td><td>16</td><td>18</td><td>20</td><td>23</td><td>25</td></tr>
<tr><td>$\Delta_k \in (\gamma\bar{\Delta}, \bar{\Delta}]$</td><td>2</td><td colspan="2" rowspan="2"></td><td>7</td><td rowspan="2">10</td><td rowspan="2">12</td><td>14</td><td rowspan="2">17</td><td rowspan="2">19</td><td>21</td><td rowspan="2">24</td><td rowspan="2">26</td></tr>
<tr><td>$\Delta_k \in (0, \gamma\bar{\Delta}]$</td><td>3</td><td>8</td><td>15</td><td>22</td></tr>
</table>

Lemma (sufficient condition for successful step)

If $I_k J_k = 1$ *and* $\Lambda_k = 0$ *then* $\Theta_k = 1$.

Lemma (progress made in each iteration)

Let $h(\delta) = C_3\delta^2$. *Then we have*

$$\phi(X_k) - \phi(X_{k+1}) \geq \begin{cases} h(\Delta_k) - \mathcal{E}_k - \mathcal{E}_k^+ - r & \text{if } \Theta_k = 1 \text{ (successful)} \\ -\mathcal{E}_k - \mathcal{E}_k^+ - r & \text{if } \Theta'_k = 1 \text{ (accepted)} \\ 0 & \text{if } \Theta'_k = 0 \text{ (rejected).} \end{cases}$$

Lemma (total progress)

$$h(\gamma\bar{\Delta}) \sum_{k=0}^{T-1} \Theta_k \Lambda_k \leq \phi(x_0) - \hat{\phi} + \sum_{k=0}^{T-1} \Theta'_k \left(\mathcal{E}_k + \mathcal{E}_k^+ + r\right).$$

Lemma (total loss)

For any $t \geq 0$,

$$\mathbb{P}\left\{\sum_{k=0}^{T-1}\left(\mathcal{E}_k + \mathcal{E}_k^+ + r\right) \geq T(4/a + 2\epsilon_f + r) + t\right\} \leq \exp\left(-\frac{a}{4}t\right).$$

Let $t = rT$.

Lemma (total progress)

$$\begin{aligned} h(\gamma\bar{\Delta})\sum_{k=0}^{T-1}\Theta_k\Lambda_k &\leq \phi(x_0) - \hat{\phi} + \sum_{k=0}^{T-1}\Theta'_k\left(\mathcal{E}_k + \mathcal{E}_k^+ + r\right) \\ &< \phi(x_0) - \hat{\phi} + T(4/a + 2\epsilon_f + 2r) \end{aligned}$$

with probability at least $1 - \exp\left(-\frac{ar}{4}T\right)$.

$$h(\gamma\bar{\Delta}) = \gamma^2 C_3\left(C_1 \min_{0\leq k\leq T-1}\|\nabla\phi(X_k)\| - C_2\epsilon_g\right)^2$$

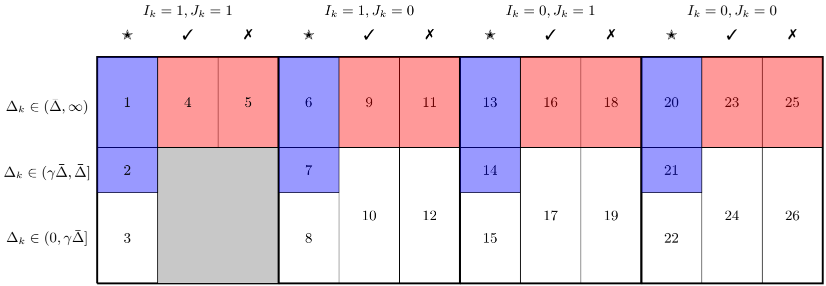
Classification of Iterations

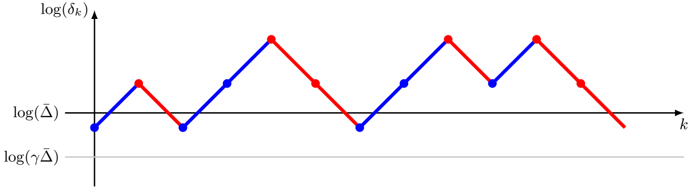
<table>
<tr><th></th><th colspan="3">$I_k = 1, J_k = 1$</th><th colspan="3">$I_k = 1, J_k = 0$</th><th colspan="3">$I_k = 0, J_k = 1$</th><th colspan="3">$I_k = 0, J_k = 0$</th></tr>
<tr><th></th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th></tr>
<tr><td>$\Delta_k \in (\bar{\Delta}, \infty)$</td><td>1</td><td>4</td><td>5</td><td>6</td><td>9</td><td>11</td><td>13</td><td>16</td><td>18</td><td>20</td><td>23</td><td>25</td></tr>
<tr><td>$\Delta_k \in (\gamma\bar{\Delta}, \bar{\Delta}]$</td><td>2</td><td colspan="2" rowspan="2"></td><td>7</td><td rowspan="2">10</td><td rowspan="2">12</td><td>14</td><td rowspan="2">17</td><td rowspan="2">19</td><td>21</td><td rowspan="2">24</td><td rowspan="2">26</td></tr>
<tr><td>$\Delta_k \in (0, \gamma\bar{\Delta}]$</td><td>3</td><td>8</td><td>15</td><td>22</td></tr>
</table>

Lemma (**total progress**)

$$h(\gamma\bar{\Delta}) \sum_{k=0}^{T-1} \Theta_k \Lambda_k \leq \phi(x_0) - \hat{\phi} + T(4/a + 2\epsilon_f + 2r).$$

Ups and Downs of the Radius

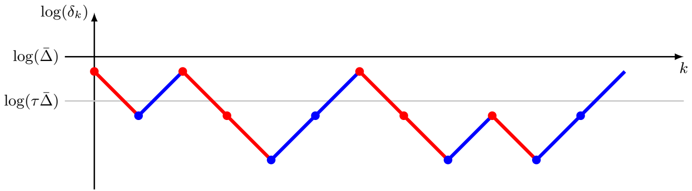




$$\sum_{k=0}^{T-1} (1-\Theta_k)\Lambda_k < \sum_{k=0}^{T-1} \Theta_k \Lambda_k + \min\left\{\log_\gamma\left(\frac{\delta_0}{\bar{\Delta}}\right), 0\right\} + 1$$

Downs and Ups of the Radius

	$I_k=1, J_k=1$			$I_k=1, J_k=0$			$I_k=0, J_k=1$			$I_k=0, J_k=0$		
	★	✓	✗	★	✓	✗	★	✓	✗	★	✓	✗
$\Delta_k \in (\bar{\Delta}, \infty)$	1	4	5	6	9	11	13	16	18	20	23	25
$\Delta_k \in (\gamma\bar{\Delta}, \bar{\Delta}]$	2			7	10	12	14	17	19	21	24	26
$\Delta_k \in (0, \gamma\bar{\Delta}]$	3			8			15			22		



$$\sum_{k=0}^{T-1} \Theta_k (1-\Lambda_k) < \sum_{k=0}^{T-1} (1-\Theta_k)(1-\Lambda_k) + \min\left\{ \log_\gamma \left(\frac{\bar{\Delta}}{\delta_0} \right), 0 \right\} + 1$$

Iterations with Sufficiently Accurate Gradient Estimate

<table>
<tr><th></th><th colspan="3">$I_k = 1, J_k = 1$</th><th colspan="3">$I_k = 1, J_k = 0$</th><th colspan="3">$I_k = 0, J_k = 1$</th><th colspan="3">$I_k = 0, J_k = 0$</th></tr>
<tr><th></th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th></tr>
<tr><td>$\Delta_k \in (\bar{\Delta}, \infty)$</td><td>1</td><td>4</td><td>5</td><td>6</td><td>9</td><td>11</td><td>13</td><td>16</td><td>18</td><td>20</td><td>23</td><td>25</td></tr>
<tr><td>$\Delta_k \in (\gamma\bar{\Delta}, \bar{\Delta}]$</td><td>2</td><td colspan="2" rowspan="2"></td><td>7</td><td rowspan="2">10</td><td rowspan="2">12</td><td>14</td><td rowspan="2">17</td><td rowspan="2">19</td><td>21</td><td rowspan="2">24</td><td rowspan="2">26</td></tr>
<tr><td>$\Delta_k \in (0, \gamma\bar{\Delta}]$</td><td>3</td><td>8</td><td>15</td><td>22</td></tr>
</table>

Lemma

Assume $\mathbb{P}\{I_k = 1 \mid \mathcal{F}_k\} \geq p_1$ *holds. By Azuma-Hoeffding inequality, for any positive integer* T *and any* $\hat{p}_1 \in [0, p_1]$ *we have*

$$\mathbb{P}\left\{\sum_{k=0}^{T-1} I_k > \hat{p}_1 T\right\} \geq 1 - \exp\left(-\frac{(1 - \hat{p}_1/p_1)^2}{2} T\right).$$

<table>
<tr><th></th><th colspan="3">$I_k = 1, J_k = 1$</th><th colspan="3">$I_k = 1, J_k = 0$</th><th colspan="3">$I_k = 0, J_k = 1$</th><th colspan="3">$I_k = 0, J_k = 0$</th></tr>
<tr><th></th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th><th>★</th><th>✓</th><th>✗</th></tr>
<tr><td>$\Delta_k \in (\bar{\Delta}, \infty)$</td><td>1</td><td>4</td><td>5</td><td>6</td><td>9</td><td>11</td><td>13</td><td>16</td><td>18</td><td>20</td><td>23</td><td>25</td></tr>
<tr><td>$\Delta_k \in (\gamma\bar{\Delta}, \bar{\Delta}]$</td><td>2</td><td colspan="2" rowspan="2"></td><td>7</td><td rowspan="2">10</td><td rowspan="2">12</td><td>14</td><td rowspan="2">17</td><td rowspan="2">19</td><td>21</td><td rowspan="2">24</td><td rowspan="2">26</td></tr>
<tr><td>$\Delta_k \in (0, \gamma\bar{\Delta}]$</td><td>3</td><td>8</td><td>15</td><td>22</td></tr>
</table>

Lemma

Assume both $\mathbb{P}\{\mathcal{E}_k > t\}$ *and* $\mathbb{P}\{\mathcal{E}_k^+ > t\}$ *are* $\leq \exp(a(\epsilon_f - t))$*. Let* $p_0 = 1 - 2\exp(a[\epsilon_f - r/2])$*. For any positive integer* T *and any* $\hat{p}_0 \in [0, p_0]$*, we have*

$$\mathbb{P}\left\{\sum_{k=0}^{T-1} J_k > \hat{p}_0 T\right\} \geq 1 - \exp\left(-\frac{(1-\hat{p}_0/p_0)^2}{2}T\right).$$

Analysis

$$\sum_{k=0}^{T-1}(1-\Theta_k)\Lambda_k < \sum_{k=0}^{T-1}\Theta_k\Lambda_k + \min\left\{\log_\gamma\left(\frac{\delta_0}{\bar{\Delta}}\right), 0\right\} + 1$$

$$\sum_{k=0}^{T-1}\Theta_k(1-\Lambda_k) < \sum_{k=0}^{T-1}(1-\Theta_k)(1-\Lambda_k) + \min\left\{\log_\gamma\left(\frac{\bar{\Delta}}{\delta_0}\right), 0\right\} + 1$$

$$\mathbb{P}\left\{\sum_{k=0}^{T-1} I_k > \hat{p}_1 T\right\} \geq 1 - \exp\left(-\frac{(1-\hat{p}_1/p_1)^2}{2}T\right)$$

$$\mathbb{P}\left\{\sum_{k=0}^{T-1} J_k > \hat{p}_0 T\right\} \geq 1 - \exp\left(-\frac{(1-\hat{p}_0/p_0)^2}{2}T\right)$$

$$\Downarrow$$

$$\mathbb{P}\left\{\sum_{k=0}^{T-1}\Theta_k\Lambda_k > \left(\hat{p}_0 + \hat{p}_1 - \frac{3}{2}\right)T - \frac{1}{2}\left|\log_\gamma\frac{\bar{\Delta}}{\delta_0}\right| - \frac{1}{2}\right\}$$

$$\geq 1 - \exp\left(-\frac{(1-\hat{p}_1/p_1)^2}{2}T\right) - \exp\left(-\frac{(1-\hat{p}_0/p_0)^2}{2}T\right)$$

Main Result

Let $t = rT$.

Theorem

Let assumptions hold. Given any $\epsilon > \sqrt{\frac{4\epsilon_f + 8/a + 2r}{C_3\gamma^2 C_1^2(2p_0 + 2p_1 - 3)}} + \frac{C_2}{C_1}\epsilon_g$*, we have*

$$\mathbb{P}\left\{\min\{\|\nabla\phi(X_k)\| : \ 0 \le k \le T-1\} \le \epsilon\right\} \ge$$
$$1 - \exp\left(-\frac{(1-\hat{p}_1/p_1)^2}{2}T\right) - \exp\left(-\frac{(1-\hat{p}_0/p_0)^2}{2}T\right) - \exp\left(-\frac{ar}{4}T\right)$$

for any $\hat{p}_0$ *and* $\hat{p}_1$ *such that* $\hat{p}_0 + \hat{p}_1 \in \left(\frac{3}{2} + \frac{2\epsilon_f + 4/a + r}{C_3\gamma^2(C_1\epsilon - C_2\epsilon_g)^2}, p_0 + p_1\right]$*, any* $t \ge 0$*, and any*

$$T \ge \left(\hat{p}_0 + \hat{p}_1 - \frac{3}{2} - \frac{2\epsilon_f + 4/a + 2r}{C_3\gamma^2(C_1\epsilon - C_2\epsilon_g)^2}\right)^{-1}$$
$$\left[\frac{\phi(x_0) - \hat{\phi}}{C_3\gamma^2(C_1\epsilon - C_2\epsilon_g)^2} + \frac{1}{2}\left|\log_\gamma \frac{C_1\epsilon - C_2\epsilon_g}{\delta_0}\right| + \frac{1}{2}\right] = \bar{\mathcal{O}}(\epsilon^{-2}).$$

Other Results

- Analyses under bounded noise assumption.
- Second-order TR method and analysis.
- Numerically testing the strength of the theoretical results.
- Experimenting with different values for r.