Summary for single funciton optimization f

B stands for B-Lipschitz of f, R is the diameter of X, L stands for L-smoothness, μ stands for μ-strongly convexity, and $\kappa := L/\mu$.

name	γ_t	convex	condition	last/average iterate convergence	convergence rate w.r.t T	ϵ optimal iteration
GD / PGD	$\frac{R}{B\sqrt{T}}$	True	B	Avg	$\mathcal{O}(1/\sqrt{T})$	$\mathcal{O}(1/\varepsilon^2)$
GD / PGD / Mirror GD	$\frac{1}{L}$	True	L	Last	$\mathcal{O}(1/T)$	$\mathcal{O}(1/\varepsilon)$
Nesterov Accelerated GD	Addaptive	True	L	Last	$\mathcal{O}(1/T^2)$	$\mathcal{O}(1/\sqrt{\varepsilon})$
GD / PGD	$\frac{1}{L}$	True	L, μ	Last	$\mathcal{O}((1-\frac{\mu}{L})^T)$	$\mathcal{O}(\kappa \ln(\frac{1}{\varepsilon}))$
GD	$\frac{1}{L}$	True	L, μ-PL	Last	$\mathcal{O}((1-\frac{\mu}{L})^T)$	$\mathcal{O}(\kappa \ln(\frac{1}{\varepsilon}))$
Coorinate GD	$\frac{1}{L_i}$	True	L, μ	Last	$\mathcal{O}(\left(1-\frac{\mu}{d\bar{L}}\right)^T)$	$\mathcal{O}(d\kappa \ln(\frac{1}{\varepsilon}))$
Sub GD / Mirror GD	$\frac{R}{B\sqrt{T}}$	True	B	Avg	$\mathcal{O}(\frac{1}{\sqrt{T}})$	$\mathcal{O}(1/\varepsilon^2)$
Sub GD	$\frac{f(x_t)-f^*}{\|g(x_t)\|_2^2}$, $\frac{2}{\mu(t+1)}$	True	B, μ	Avg	$\mathcal{O}(\frac{1}{T})$	$\mathcal{O}(1/\varepsilon)$
Frank-Wolfe	$\frac{2}{t+2}$	True	L	Last	$\mathcal{O}(\frac{1}{T})$	$\mathcal{O}(1/\varepsilon)$
Newton	Auto	True	Hessian B-Lipschitz, μ	Last	$\mathcal{O}((\frac{1}{2})^{2^T-1})$	$\mathcal{O}(\log\log(1/\varepsilon)$
Newton	Auto	True	L, μ	Last	$\mathcal{O}\left((1-\frac{\mu^2}{L^2})^t\right)$	$\mathcal{O}(\kappa^2 \ln(\frac{1}{\varepsilon}))$

Summary for stochastic funciton optimization of $F = \mathbb{E}[f_\xi]$.

name	γ_t	convex	condition	last/average iterate convergence	convergence rate w.r.t T	ϵ optimal iteration of f or $\|\nabla f\|$
SGD	$\frac{R}{B\sqrt{T}}$	True	B	Avg	$\mathcal{O}(\frac{1}{\sqrt{T}})$	$\mathcal{O}(\frac{1}{\epsilon^2})$
SGD	$\frac{1}{2\mu}$	True	B, μ	Last	$\mathcal{O}(\frac{1}{T})$	$\mathcal{O}(\frac{1}{\epsilon})$
SGD	$\min\left\{\frac{1}{L}, \frac{\gamma}{\sigma\sqrt{T}}\right\}$	False	L, $\mathbf{var}[\nabla f_\xi] \leq \sigma$	Last	$\mathcal{O}(\frac{1}{T^{1/4}})$	$\mathcal{O}(\frac{1}{\epsilon^4})$

Summary for stochastic funciton optimization of finite sum $F = \sum_i f_i/n$.

name	γ_t	convex	condition	last/average iterate convergence	convergence rate w.r.t T	ϵ optimal iteration	computation per-iteration	ϵ optimal computation cost
Full GD	$\frac{1}{\overline{L_i}}$	True	$\overline{L_i}, \mu$	Last	$\mathcal{O}(\left(1-\frac{\mu}{\overline{L_i}}\right)^T)$	$\mathcal{O}(\kappa \ln(\frac{1}{\varepsilon}))$	$\mathcal{O}(n)$	$\mathcal{O}(n\kappa \ln(\frac{1}{\varepsilon}))$
SAG/SAGA	$1/16L$	True	L_i	Last	$\mathcal{O}((1-O(\frac{1}{n})-O(\kappa)^T))$	$\mathcal{O}((n+\kappa)\ln(\frac{1}{\epsilon}))$	$\mathcal{O}(1)$	$\mathcal{O}((n+\kappa)\ln(\frac{1}{\epsilon}))$
SVRG	$< 1/2L$	True	L_i	Last	$\mathcal{O}(\rho^T))$	$\mathcal{O}((n+\kappa)\ln(\frac{1}{\epsilon}))$	$\mathcal{O}(2)$	$\mathcal{O}((n+\kappa)\ln(\frac{1}{\epsilon}))$
SPIDER	$\eta < 1/2L$	False	Avg-L, $\mathbf{var}[\nabla f_\xi] \leq \sigma$	Avg	$\mathcal{O}(\frac{1}{\sqrt{T}})$	$\mathcal{O}(\frac{1}{\epsilon^2})$	$\mathcal{O}(\frac{\sigma}{\epsilon})$	$\mathcal{O}(\frac{\sigma}{\epsilon^3})$