Expected Values

Let $\mu(x)$ be the probability that $X < x$, then:

$$\mathbb{E}(X) := \int x \, d\mu = \int x \, \mu(x) \, dx$$

$\mathbb{E}(X)$ gives the *Expected Value* when we draw a random variable X from a distribution whose "weightedness" is described by cumulative distribution function: $\mu(x)$.

Note that $\mathbb{E}$ is a linear operator:

- $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$ where α is a constant
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

WWST: $\quad \mathbb{E}\left(\|x^{(k+1)} - x^*\|_2^2\right) \leq (1 - 2m\alpha_k)\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) + \alpha_k^2 G^2.$

Strong Convexity Equivalence

Let f be strongly convex:

$$\nabla^2 f \succeq mI$$

$$(x-y)^T \left(\nabla^2 f - mI\right)(x-y) \geq 0$$

$$(x-y)^T(\nabla f(x) - \nabla f(y)) - m(x-y)^T I(x-y) \geq 0$$

$$(x-y)^T(\nabla f(x) - \nabla f(y)) - m\|x-y\|_2^2 \geq 0$$

$$(x-y)^T(\nabla f(x) - \nabla f(y)) \geq m\|x-y\|_2^2$$

Substitute $y \leftarrow x^*$:

$$(x-x^*)^T(\nabla f(x) - \nabla f(x^*)) \geq m\|x-x^*\|_2^2$$

$$(x-x^*)^T \nabla f(x) \geq m\|x-x^*\|_2^2$$

Convergence Proof

$$(1): x^{(k+1)} = x^{(k)} - \alpha_k g^{(k)}, \quad (2): \mathbb{E}\|g^{(k)}\|_2 = \nabla f\left(x^{(k)}\right) \leq G$$

(3): Since f is strongly convex: $(x - x^*)^T \nabla f(x) \leq m\|x - x^*\|_2^2$.

$$\begin{aligned}
\mathbb{E}\left(\|x^{(k+1)} - x^*\|_2^2\right) &= \mathbb{E}\left(\|(x^{(k)} - \alpha_k g^{(k)}) - x^*\|_2^2\right) \qquad \text{by (1)} \\
&= \mathbb{E}\left(\|(x^{(k)} - x^*) - \alpha_k g^{(k)}\|_2^2\right) \\
&= \mathbb{E}\left(\|x^{(k)} - x^*\|_2^2 - 2\alpha_k (x^{(k)} - x^*)^T g^{(k)} + \alpha_k^2 \|g^{(k)}\|_2^2\right) \\
&= \mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) - 2\alpha_k \mathbb{E}\left((x^{(k)} - x^*)^T g^{(k)}\right) + \alpha_k^2 \mathbb{E}\left(\|g^{(k)}\|_2^2\right) \\
&\leq \mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) - 2\alpha_k \mathbb{E}\left((x^{(k)} - x^*)^T \nabla f(x^{(k)})\right) + \alpha_k^2 G^2 \quad \text{by (2)} \\
&\leq \mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) - 2m\alpha_k \mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) + \alpha_k^2 G^2 \quad \text{by (3)} \\
&= (1 - 2m\alpha_k)\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) + \alpha_k^2 G^2. \quad \square
\end{aligned}$$

Diminishing (Square Summable but not Summable) Step Size

Take a diminishing step size: $\alpha_k = \alpha / k$ where $\alpha > \frac{1}{2m}$. Then:

$$\begin{aligned}\mathbb{E}\left(\|x^{(k+1)} - x^*\|_2^2\right) &\leq (1 - 2m\alpha_k)\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) + \alpha_k^2 G^2 \\ &= (1 - 2m\alpha/k)\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) + \alpha^2 G^2 / k^2.\end{aligned}$$

WWS by induction that:

$$\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) \leq Q/k$$

where

$$Q = \max\{\|x_1 - x^*\|_2^2, \alpha^2 G^2 / (2m\alpha - 1)\}.$$

Diminishing Step Size Proof

Base Case: Clearly, $\|x_1 - x^*\|_2^2 \leq \|x_1 - x^*\|_2^2/1$.

Inductive Step: Assume: $\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) \leq Q/k$.

$$\begin{aligned}
\mathbb{E}\left(\|x^{(k+1)} - x^*\|_2^2\right) &\leq \frac{\left(1 - \frac{2m\alpha}{k}\right)\frac{\alpha^2 G^2}{(2m\alpha - 1)}}{k} + \frac{\alpha^2 G^2}{k^2} \\
&= \frac{(\alpha^2 G^2)(k - 2m\alpha)}{k^2(2m\alpha - 1)} + \frac{\alpha^2 G^2(2m\alpha - 1)}{k^2(2m\alpha - 1)} \\
&= \frac{(\alpha^2 G^2)(k - 2m\alpha) + \alpha^2 G^2(2m\alpha - 1)}{k^2(2m\alpha - 1)} \\
&= \frac{(\alpha^2 G^2)(k - 2m\alpha + 2m\alpha - 1)}{k^2(2m\alpha - 1)} \\
&= \left((\alpha^2 G^2)/(2m\alpha - 1)\right)(k - 1)/(k^2) \\
&< (Q)(k - 1)/(k^2 - 1) \\
&= Q/(k + 1). \quad \square
\end{aligned}$$

Diminishing Step Size Convergence Rate

After k iterations,

$$\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) \leq Q/k$$

where Q is a problem-dependent constant.

So,

$$\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) \approx \mathcal{O}(1/k)$$

$$\Rightarrow \mathbb{E}\left(\|x^{(k)} - x^*\|_2\right) \approx \mathcal{O}(1/\sqrt{k}).$$

If ∇f is Lipschitz continuous, then there exists L such that:

$$L\|x^{(k)} - x^*\|_2 \geq \|\nabla f(x^{(k)}) - \nabla f(x^*)\|_2$$

$$\Rightarrow \int L\|x^{(k)} - x^*\|_2 dx \geq \int \|\nabla f(x^{(k)}) - \nabla f(x^*)\|_2 dx$$

$$\Rightarrow \frac{L}{2}\|x^{(k)} - x^*\|_2^2 \geq \|f(x^{(k)}) - f(x^*)\|_2$$

So

$$\mathbb{E}\left(\|f(x^{(k)}) - f(x^*)\|_2\right) \leq \frac{L}{2}\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) \leq \frac{LQ/2}{k} \approx \mathcal{O}(1/k).$$

Another Formulation for Convergence Rate

$$(1)\text{: } x^{(k+1)} = x^{(k)} - \alpha_k g^{(k)}, \quad (2)\text{: } \mathbb{E}\|g^{(k)}\|_2 = \nabla f\left(x^{(k)}\right) \leq G$$

(3): Now suppose f is only convex: $f(x) - f(y) \geq (x-y)^T \nabla f(x)$.

$$\begin{aligned}
\mathbb{E}\left(\|x^{(k+1)} - x^*\|_2^2\right) &\leq \mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) - 2\alpha_k \mathbb{E}\left((x^{(k)} - x^*)^T g^{(k)}\right) + \alpha_k^2 \mathbb{E}\left(\|g^{(k)}\|_2^2\right) \\
&\leq \mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) - 2\alpha_k \mathbb{E}\left(f(x^{(k)}) - f(x^*)\right) + \alpha_k^2 G^2 \text{ by (2) and (3)}
\end{aligned}$$

$$\Rightarrow$$

$$\begin{aligned}
\sum_{k=1}^N \mathbb{E}\left(\|x^{(k+1)} - x^*\|_2^2\right) &\leq \sum_{k=1}^N \left(\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) - 2\alpha_k \mathbb{E}\left(f(x^{(k)}) - f(x^*)\right) + \alpha_k^2 G^2\right) \\
&\leq \sum_{k=1}^N \mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) - 2\left(f(x^{best}) - f(x^*)\right) \sum_{k=1}^N \alpha_k + \sum_{k=1}^N \alpha_k^2 G^2
\end{aligned}$$

$$\Longleftrightarrow$$

$$\begin{aligned}
f(x^{best}) - f(x^*) &\leq \frac{\sum_{k=1}^N \left(\mathbb{E}\left(\|x^{(k)} - x^*\|_2^2\right) - \mathbb{E}\left(\|x^{(k+1)} - x^*\|_2^2\right)\right) + \sum_{k=1}^N \alpha_k^2 G^2}{2\sum_{k=1}^N \alpha_k} \\
&= \frac{\|x^{(1)} - x^*\|_2^2 + \sum_{k=1}^N \alpha_k^2 G^2}{2\sum_{k=1}^N \alpha_k} = \frac{D_X^2 + \sum_{k=1}^N \alpha_k^2 G^2}{2\sum_{k=1}^N \alpha_k}
\end{aligned}$$

Constant Step Size

Let the step size α_k be a constant α:

$$f(x^{best}) - f(x^*) \leq \frac{D_X^2 + N\alpha^2 G^2}{2N\alpha}$$

Find optmial α^* by minimize RHS WRT α:

$$\alpha^* = \frac{D_X}{G\sqrt{N}}$$

Then we converge to the neighborhood

$$f(x^{(N)}) - f(x^*) \leq G^2\alpha/2$$

in N iterations with a maximum error of

$$\frac{D_X M}{\sqrt{N}} \approx \mathcal{O}\left(\frac{1}{\sqrt{N}}\right).$$

Review

If f is strongly convex:

Take diminishing steps (square summable but not summable) and in k iterations converge to the *expected* error:

$$\mathbb{E}\left(\|f(x^{(k+1)}) - f(x^*)\|_2\right) \approx \mathcal{O}\left(\frac{1}{k}\right)$$

If f is only convex:

Pick N iterations ahead of time such that $f(x^{(N)}) - f(x^*) \leq \frac{G^2\alpha}{2}$ is an acceptable error. Take constant steps of magnitude $\frac{D_X}{G\sqrt{N}}$ and converge to the neighborhood with error:

$$\mathbb{E}\left(\|f(x^{(k+1)}) - f(x^*)\|_2\right) \leq \frac{G^2\alpha}{2} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$$

Footnote:
In practice, the constants G and D_X are not known in advance, nor is the desired error. A better strategy is to take constant steps of some magnitude until convergence stalls, then halve the step size and continue.