Gradient Descent.

Convex, continuous function: $f(x)$.

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y - x)$.

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y - x)$.

L-Lipshitz assumption:

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y - x)$.

L-Lipshitz assumption:
$f(y) \leq f(x) + \frac{L}{2}\|y - x\|^2$.

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

L-Lipshitz assumption:
$f(y) \leq f(x) + \frac{L}{2}\|y-x\|^2$.
$(\nabla f(y) - \nabla f(x)) \leq L(y-x)$.

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

L-Lipshitz assumption:
$f(y) \leq f(x) + \frac{L}{2}\|y-x\|^2$.
$(\nabla f(y) - \nabla f(x)) \leq L(y-x)$.

$x_{t+1} = x_t - \frac{1}{L}\nabla f(x_t)$.

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y - x)$.

L-Lipshitz assumption:
$f(y) \leq f(x) + \frac{L}{2}\|y - x\|^2$.
$(\nabla f(y) - \nabla f(x)) \leq L(y - x)$.

$x_{t+1} = x_t - \frac{1}{L}\nabla f(x_t)$.

Average gradient $\geq \nabla f(x_t)/2$

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

L-Lipshitz assumption:
$f(y) \leq f(x) + \frac{L}{2}\|y-x\|^2$.
$(\nabla f(y) - \nabla f(x)) \leq L(y-x)$.

$x_{t+1} = x_t - \frac{1}{L}\nabla f(x_t)$.

Average gradient $\geq \nabla f(x_t)/2$
thus reduce function value by at least $\|\nabla(f(x_t))\|^2/2$.

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

L-Lipshitz assumption:
$f(y) \leq f(x) + \frac{L}{2}\|y-x\|^2$.
$(\nabla f(y) - \nabla f(x)) \leq L(y-x)$.

$x_{t+1} = x_t - \frac{1}{L}\nabla f(x_t)$.

Average gradient $\geq \nabla f(x_t)/2$
thus reduce function value by at least $\|\nabla(f(x_t))\|^2/2$.

Closest to x^*?

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

L-Lipshitz assumption:
$f(y) \leq f(x) + \frac{L}{2}\|y-x\|^2$.
$(\nabla f(y) - \nabla f(x)) \leq L(y-x)$.

$x_{t+1} = x_t - \frac{1}{L}\nabla f(x_t)$.

Average gradient $\geq \nabla f(x_t)/2$
thus reduce function value by at least $\|\nabla(f(x_t))\|^2/2$.

Closest to x^*?

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y - x)$.

L-Lipshitz assumption:
$f(y) \leq f(x) + \frac{L}{2}\|y - x\|^2$.
$(\nabla f(y) - \nabla f(x)) \leq L(y - x)$.

$x_{t+1} = x_t - \frac{1}{L}\nabla f(x_t)$.

Average gradient $\geq \nabla f(x_t)/2$
thus reduce function value by at least $\|\nabla(f(x_t))\|^2/2$.

Closest to x^*?
Functional distance: $f(x_t) - f(x^*)$.

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

L-Lipshitz assumption:
$f(y) \leq f(x) + \frac{L}{2}\|y-x\|^2$.
$(\nabla f(y) - \nabla f(x)) \leq L(y-x)$.

$x_{t+1} = x_t - \frac{1}{L}\nabla f(x_t)$.

Average gradient $\geq \nabla f(x_t)/2$
thus reduce function value by at least $\|\nabla(f(x_t))\|^2/2$.

Closest to x^*?
Functional distance: $f(x_t) - f(x^*)$.
Distance to solution: $\|x_t - x^*\|$.

Gradient Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

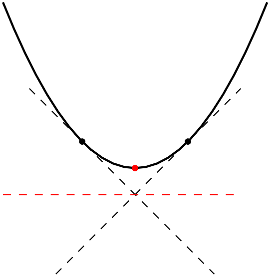
L-Lipshitz assumption:
$f(y) \leq f(x) + \frac{L}{2}\|y-x\|^2$.
$(\nabla f(y) - \nabla f(x)) \leq L(y-x)$.

$x_{t+1} = x_t - \frac{1}{L}\nabla f(x_t)$.

Average gradient $\geq \nabla f(x_t)/2$
thus reduce function value by at least $\|\nabla(f(x_t))\|^2/2$.

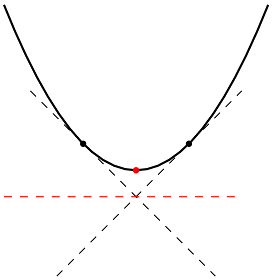
Closest to x^*?
Functional distance: $f(x_t) - f(x^*)$.
Distance to solution: $\|x_t - x^*\|$.
Translate using L: $f(x_t) - f(x^*) \leq \frac{L}{2}\|x_t - x^*\|^2$.

Mirror Descent



- Each point gives a linear lower bound.
- Average of the lower bounds becomes flatter.
- Add the point with current worst regret.
- Output average of queried points.
- $x = \alpha x_1 + (1 - \alpha) x_2$

Mirror Descent



- Each point gives a linear lower bound.
- Average of the lower bounds becomes flatter.
- Add the point with current worst regret.
- Output average of queried points.
- $x = \alpha x_1 + (1-\alpha) x_2$
  $\implies f(x) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$.

Analysis doesn't require L-Lipschitz.

Mirror Descent: Regret Minimization

- Average **Regret** with loss vector ξ_i's

$$R_k(u) = \frac{1}{k}\sum_{i=0}^{k-1}\langle \xi_i, z_i - u\rangle$$

Why care about average regret? Bounds gap to OPT:

With $\xi_i = \nabla f(z_i)$, $\bar{z} = \frac{1}{k}\sum_{i=0}^{k-1} z_i$,

$$f(\bar{z}) - f(u) \leq \frac{1}{k}\sum_{i=0}^{k-1} f(z_i) - f(u) \leq \frac{1}{k}\sum_{i=0}^{k-1}\langle \nabla f(z_i), z_i - u\rangle = R_k(u)$$

$$f(\bar{z}) - \mathrm{OPT} \leq \max_u R_k(u)$$

Mirror Descent: Regret Minimization

- Regularized average regret

$$\tilde{R}_k(u) = \frac{1}{\alpha k}\left(-w(u) + \alpha \sum_{i=0}^{k-1} \langle \xi_i, z_i - u \rangle\right)$$
$$= R_k(u) - \frac{w(u)}{\alpha k}$$

Distance Generating Function

$$w : \mathbb{R}^n \to \mathbb{R}$$

1-strongly convex for norm $\|\cdot\|$:

$$w(y) \geq w(x) + \langle \nabla w(x), y - x \rangle + \frac{1}{2}\|x - y\|^2$$

Mirror Descent: Regret Minimization

- Regularized average regret

$$\tilde{R}_k(u) = \frac{1}{\alpha k}\left(-w(u) + \alpha \sum_{i=0}^{k-1} \langle \xi_i, z_i - u \rangle\right)$$
$$= R_k(u) - \frac{w(u)}{\alpha k}$$

Distance Generating Function

$$w : \mathbb{R}^n \to \mathbb{R}$$

1-strongly convex for norm $\|\cdot\|$:

$$w(y) \geq w(x) + \langle \nabla w(x), y - x \rangle + \frac{1}{2}\|x - y\|^2$$

For ℓ_2-norm, simply $w(x) = \frac{1}{2}\|x\|_2^2$.

Mirror Descent: Regret Minimization

- Regularized average regret

$$\tilde{R}_k(u) = \frac{1}{\alpha k}(-w(u) + \alpha \sum_{i=0}^{k-1} \langle \xi_i, z_i - u \rangle)$$
$$= R_k(u) - \frac{w(u)}{\alpha k}$$

Distance Generating Function

$$w : \mathbb{R}^n \to \mathbb{R}$$

1-strongly convex for norm $\|\cdot\|$:

$$w(y) \geq w(x) + \langle \nabla w(x), y - x \rangle + \frac{1}{2}\|x - y\|^2$$

For ℓ_2-norm, simply $w(x) = \frac{1}{2}\|x\|_2^2$.
(For distributions: $w(x) = -\sum_i x_i \log x_i$.)

Bregman divergence

$$V_x(y) = w(y) - \langle \nabla w(x), y - x \rangle - w(x) \geq \frac{1}{2}\|x - y\|^2$$

Standard three point property of Bregman divergence:

$$\forall x, y \geq 0 \quad \langle -\nabla V_x(y), y - u \rangle = V_x(u) - V_y(u) - V_x(y),$$

Bregman divergence

$$V_x(y) = w(y) - \langle \nabla w(x), y - x \rangle - w(x) \geq \frac{1}{2}\|x - y\|^2$$

Standard three point property of Bregman divergence:

$$\forall x, y \geq 0 \quad \langle -\nabla V_x(y), y - u \rangle = V_x(u) - V_y(u) - V_x(y),$$

For ℓ_2-norm, $V_x(y) = \frac{1}{2}\|x - y\|_2^2$, $\nabla V_x(y) = (x - y)$

Bregman divergence

$$V_x(y) = w(y) - \langle \nabla w(x), y - x \rangle - w(x) \geq \frac{1}{2}\|x - y\|^2$$

Standard three point property of Bregman divergence:

$$\forall x, y \geq 0 \quad \langle -\nabla V_x(y), y - u \rangle = V_x(u) - V_y(u) - V_x(y),$$

For ℓ_2-norm, $V_x(y) = \frac{1}{2}\|x - y\|_2^2$, $\nabla V_x(y) = (x - y)$

Three point property $\leftrightarrow$ Law of cosines

Bregman divergence

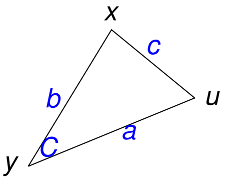
$$V_x(y) = w(y) - \langle \nabla w(x), y - x \rangle - w(x) \geq \frac{1}{2}\|x - y\|^2$$

Standard three point property of Bregman divergence:

$$\forall x, y \geq 0 \quad \langle -\nabla V_x(y), y - u \rangle = V_x(u) - V_y(u) - V_x(y),$$

For ℓ_2-norm, $V_x(y) = \frac{1}{2}\|x - y\|_2^2$, $\nabla V_x(y) = (x - y)$

Three point property $\leftrightarrow$ Law of cosines



$$c^2 = a^2 + b^2 - 2ab\cos(C)$$

Bregman divergence

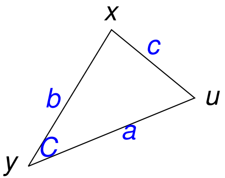
$$V_x(y) = w(y) - \langle \nabla w(x), y - x \rangle - w(x) \geq \frac{1}{2}\|x - y\|^2$$

Standard three point property of Bregman divergence:

$$\forall x, y \geq 0 \quad \langle -\nabla V_x(y), y - u \rangle = V_x(u) - V_y(u) - V_x(y),$$

For ℓ_2-norm, $V_x(y) = \frac{1}{2}\|x - y\|_2^2$, $\nabla V_x(y) = (x - y)$

Three point property $\leftrightarrow$ Law of cosines



$$c^2 = a^2 + b^2 - 2ab\cos(C) \text{ or } 2ab\cos(C) = c^2 - a^2 - b^2$$

Bregman divergence

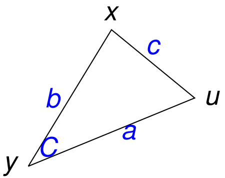
$$V_x(y) = w(y) - \langle \nabla w(x), y - x \rangle - w(x) \geq \frac{1}{2}\|x - y\|^2$$

Standard three point property of Bregman divergence:

$$\forall x, y \geq 0 \quad \langle -\nabla V_x(y), y - u \rangle = V_x(u) - V_y(u) - V_x(y),$$

For ℓ_2-norm, $V_x(y) = \frac{1}{2}\|x - y\|_2^2$, $\nabla V_x(y) = (x - y)$

Three point property $\leftrightarrow$ Law of cosines



$$c^2 = a^2 + b^2 - 2ab\cos(C) \text{ or } 2ab\cos(C) = c^2 - a^2 - b^2$$

$$a^2 = V_y(u), \; b^2 = V_x(u), \quad c^2 = V_x(u), \; 2ab\cos(C) = -(x - y)\cdot(y - u)$$

Mirror Descent

$$z_{k+1} = \mathrm{Mirr}(z_k, \alpha\xi_k) = \underset{z\in Q}{\mathrm{argmin}}\{V_{z_k}(z) + \alpha\langle\xi_k, z - z_k\rangle\}$$

Mirror Descent

$$z_{k+1} = \text{Mirr}(z_k, \alpha\xi_k) = \underset{z\in Q}{\text{argmin}}\{V_{z_k}(z) + \alpha\langle\xi_k, z - z_k\rangle\}$$

Equivalent to regret minimization when $Q = \mathbb{R}^n$:

- Optimality condition of MD step:

$$\begin{aligned}\nabla V_{z_k}(z_{k+1}) &= -\alpha\xi_k \\ z_{k+1} - z_k &= -\alpha\xi_k \\ z_{k+1} &= z_0 - \sum_i \alpha\xi_i\end{aligned}$$

- Regret Minimization:

$$z_{k+1} = \underset{z}{\text{argmax}}\{-w(z) + \alpha\sum_{i=0}^{k}\langle\xi_i, z_i - z\rangle\}$$

  Optimality condition:

$$z_{k+1} = -\sum_i \alpha\xi_i$$

Mirror Descent

$$z_{k+1} = \mathrm{Mirr}(z_k, \alpha\xi_k) = \operatorname*{argmin}_{z\in Q}\{V_{z_k}(z) + \alpha\langle \xi_k, z - z_k\rangle\}$$

Recall: $\nabla V_{z_k}(z_{k+1}) = -\alpha\xi_k \quad z_{k+1} - z_k = -\alpha\xi_k$

Lemma:

$$\begin{aligned}\alpha\langle\xi_k, z_k - u\rangle &\le \alpha\langle\xi_k, z_k - z_{k+1}\rangle + V_{z_k}(u) - V_{z_{k+1}}(u) - V_{z_k}(z_{k+1}) \\ &\le \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u) \qquad \forall u \in Q\end{aligned}$$

$$\begin{aligned}&= \alpha\langle\xi_k, z_k - z_{k+1}\rangle - \langle\nabla V_{z_k}(z_{k+1}), z_{k+1} - u\rangle \\ &= \alpha\langle\xi_k, z_k - z_{k+1}\rangle + V_{z_k}(u) - V_{z_{k+1}}(u) - V_{z_k}(z_{k+1})\end{aligned}$$

Mirror Descent

$$z_{k+1} = \mathrm{Mirr}(z_k, \alpha\xi_k) = \operatorname*{argmin}_{z \in Q}\{V_{z_k}(z) + \alpha\langle \xi_k, z - z_k\rangle\}$$

Recall: $\nabla V_{z_k}(z_{k+1}) = -\alpha\xi_k \quad z_{k+1} - z_k = -\alpha\xi_k$

Lemma:

$$\begin{aligned}\alpha\langle \xi_k, z_k - u\rangle &\le \alpha\langle \xi_k, z_k - z_{k+1}\rangle + V_{z_k}(u) - V_{z_{k+1}}(u) - V_{z_k}(z_{k+1}) \\ &\le \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u) \qquad \forall u \in Q\end{aligned}$$

Proof:

$$\alpha\langle \xi_k, z_k - u\rangle = \alpha\langle \xi_k, z_k - z_{k+1}\rangle + \alpha\langle \xi_k, z_{k+1} - u\rangle$$

□

Mirror Descent

$$z_{k+1} = \mathrm{Mirr}(z_k, \alpha\xi_k) = \operatorname*{argmin}_{z\in Q}\{V_{z_k}(z) + \alpha\langle \xi_k, z - z_k\rangle\}$$

Recall: $\nabla V_{z_k}(z_{k+1}) = -\alpha\xi_k \quad z_{k+1} - z_k = -\alpha\xi_k$

Lemma:

$$\begin{aligned}\alpha\langle \xi_k, z_k - u\rangle &\le \alpha\langle \xi_k, z_k - z_{k+1}\rangle + V_{z_k}(u) - V_{z_{k+1}}(u) - V_{z_k}(z_{k+1}) \\ &\le \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u) \qquad \forall u \in Q\end{aligned}$$

Proof:

$$\begin{aligned}\alpha\langle \xi_k, z_k - u\rangle &= \alpha\langle \xi_k, z_k - z_{k+1}\rangle + \alpha\langle \xi_k, z_{k+1} - u\rangle \\ &= \alpha\langle \xi_k, z_k - z_{k+1}\rangle - \langle \nabla V_{z_k}(z_{k+1}), z_{k+1} - u\rangle \\ &= \alpha\langle \xi_k, z_k - z_{k+1}\rangle + V_{z_k}(u) - V_{z_{k+1}}(u) - V_{z_k}(z_{k+1}) \\ &\le \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u)\end{aligned}$$

□

Mirror Descent

- $\alpha\langle \xi_k, z_k - u\rangle \le \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u)$

  Telescoping T iterations, and **width** $\|\xi_k\|_*^2 \le \rho^2$

Mirror Descent

- $\alpha\langle\xi_k, z_k - u\rangle \leq \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u)$

  Telescoping T iterations, and **width** $\|\xi_k\|_*^2 \leq \rho^2$

$$1 \cdot \frac{\alpha^2\rho^2}{2} + V_{z_0}(u) - V_{z_1}(u)$$

Mirror Descent

- $\alpha\langle \xi_k, z_k - u\rangle \leq \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u)$

  Telescoping T iterations, and **width** $\|\xi_k\|_*^2 \leq \rho^2$

$$2 \cdot \frac{\alpha^2\rho^2}{2} + V_{z_0}(u) - \cancel{V_{z_1}(u)} + \cancel{V_{z_1}(u)} - V_{z_2}(u)$$

Mirror Descent

- $\alpha\langle \xi_k, z_k - u\rangle \le \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u)$

  Telescoping T iterations, and **width** $\|\xi_k\|_*^2 \le \rho^2$

  $$3 \cdot \frac{\alpha^2\rho^2}{2} + V_{z_0}(u) - \cancel{V_{z_1}(u)} + \cancel{V_{z_1}(u)} - \cancel{V_{z_2}(u)} + \cancel{V_{z_2}(u)} - V_{z_3}(u) \ldots$$

Mirror Descent

- $\alpha\langle\xi_k, z_k - u\rangle \leq \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u)$

  Telescoping T iterations, and **width** $\|\xi_k\|_*^2 \leq \rho^2$

$$\alpha \sum_{i=0}^{T-1} \langle\xi_i, z_i - u\rangle \leq \frac{\alpha^2\rho^2 T}{2} + V_{z_0}(u) - V_{z_T}(u)$$

Mirror Descent

- $\alpha\langle \xi_k, z_k - u\rangle \leq \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u)$

  Telescoping T iterations, and **width** $\|\xi_k\|_*^2 \leq \rho^2$

$$\alpha \sum_{i=0}^{T-1} \langle \xi_i, z_i - u\rangle \leq \frac{\alpha^2 \rho^2 T}{2} + V_{z_0}(u) - V_{z_T}(u)$$

- $\alpha = \frac{\varepsilon}{\rho^2}$, **diameter** $V_{z_0}(u) \leq \Theta$, in $T = \frac{2\rho^2\Theta}{\varepsilon^2}$ iterations

$$\forall u, f(\bar{z}) - f(u) \leq \frac{\alpha\rho^2}{2} + \frac{V_{z_0}(u)}{\alpha T} \leq \varepsilon$$

Mirror Descent

- $\alpha\langle\xi_k, z_k - u\rangle \leq \frac{\alpha^2}{2}\|\xi_k\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u)$

  Telescoping T iterations, and **width** $\|\xi_k\|_*^2 \leq \rho^2$

$$\alpha \sum_{i=0}^{T-1} \langle\xi_i, z_i - u\rangle \leq \frac{\alpha^2\rho^2 T}{2} + V_{z_0}(u) - V_{z_T}(u)$$

- $\alpha = \frac{\varepsilon}{\rho^2}$, **diameter** $V_{z_0}(u) \leq \Theta$, in $T = \frac{2\rho^2\Theta}{\varepsilon^2}$ iterations

$$\forall u, f(\bar{z}) - f(u) \leq \frac{\alpha\rho^2}{2} + \frac{V_{z_0}(u)}{\alpha T} \leq \varepsilon$$

- Regret terms $\frac{\alpha^2}{2}\|\xi_k\|_*^2$ accumulate, bound step size α. Bregman divergence terms telescope.

Mirror Descent.

Convex, continuous function: $f(x)$.

Mirror Descent.

Convex, continuous function: $f(x)$.

$$f(y) \geq f(x) + \nabla f(x)(y - x).$$

Mirror Descent.

Convex, continuous function: $f(x)$.

$$f(y) \geq f(x) + \nabla f(x)(y - x).$$

Mirror Descent.

Convex, continuous function: $f(x)$.

$$f(y) \geq f(x) + \nabla f(x)(y - x).$$

No lipshitz condition necessary.

Mirror Descent.

Convex, continuous function: $f(x)$.

$$f(y) \geq f(x) + \nabla f(x)(y - x).$$

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Mirror Descent.

Convex, continuous function: $f(x)$.

$$f(y) \geq f(x) + \nabla f(x)(y - x).$$

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Introduces:

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y - x)$.

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Introduces:

$w(x)$ strongly convex function with $L = 1$

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y - x)$.

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Introduces:

$w(x)$ strongly convex function with $L = 1$

"Divergence" function $V_x(y) = w(y) - (w(x) + \nabla(w(x)) \cdot (y - x))$.

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y - x)$.

No lipshitz condition necessary.
Gradient need not be continuous: absolute value.

Introduces:
$w(x)$ strongly convex function with $L = 1$
"Divergence" function $V_x(y) = w(y) - (w(x) + \nabla(w(x)) \cdot (y - x))$.

For sequence of "gradients": ψ_i.

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y - x)$.

No lipshitz condition necessary.
Gradient need not be continuous: absolute value.

Introduces:
$w(x)$ strongly convex function with $L = 1$
"Divergence" function $V_x(y) = w(y) - (w(x) + \nabla(w(x)) \cdot (y - x))$.

For sequence of "gradients": ψ_i.
For $w(x) = \frac{1}{2}\|x - y\|^2$,
$z_k = z_0 - \alpha \|\psi_i\|^2$.
Bound "regret".

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Introduces:

$w(x)$ strongly convex function with $L = 1$

"Divergence" function $V_x(y) = w(y) - (w(x) + \nabla(w(x)) \cdot (y-x))$.

For sequence of "gradients": ψ_i.

For $w(x) = \frac{1}{2}\|x-y\|^2$,

$z_k = z_0 - \alpha\|\psi_i\|^2$.

Bound "regret".

$\sum_i^k \psi_i \cdot (z_i - u) \leq \alpha^2 \sum_i^k \|\psi_i\|^2 + V_{z_k}(u) - V_{z_0}(u)$.

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Introduces:

$w(x)$ strongly convex function with $L = 1$

"Divergence" function $V_x(y) = w(y) - (w(x) + \nabla(w(x)) \cdot (y-x))$.

For sequence of "gradients": ψ_i.

For $w(x) = \frac{1}{2}\|x-y\|^2$,

$z_k = z_0 - \alpha\|\psi_i\|^2$.

Bound "regret".

$\sum_i^k \psi_i \cdot (z_i - u) \leq \alpha^2 \sum_i^k \|\psi_i\|^2 + V_{z_k}(u) - V_{z_0}(u)$.

"Expert's" type analysis.

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Introduces:

$w(x)$ strongly convex function with $L=1$

"Divergence" function $V_x(y) = w(y) - (w(x) + \nabla(w(x)) \cdot (y-x))$.

For sequence of "gradients": ψ_i.

For $w(x) = \frac{1}{2}\|x-y\|^2$,

$z_k = z_0 - \alpha\|\psi_i\|^2$.

Bound "regret".

$\sum_i^k \psi_i \cdot (z_i - u) \leq \alpha^2 \sum_i^k \|\psi_i\|^2 + V_{z_k}(u) - V_{z_0}(u)$.

"Expert's" type analysis.

Best expert is u against psi_i.

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y - x)$.

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Introduces:

$w(x)$ strongly convex function with $L = 1$

"Divergence" function $V_x(y) = w(y) - (w(x) + \nabla(w(x)) \cdot (y - x))$.

For sequence of "gradients": ψ_i.

For $w(x) = \frac{1}{2}\|x - y\|^2$,

$z_k = z_0 - \alpha\|\psi_i\|^2$.

Bound "regret".

$\sum_i^k \psi_i \cdot (z_i - u) \leq \alpha^2 \sum_i^k \|\psi_i\|^2 + V_{z_k}(u) - V_{z_0}(u)$.

"Expert's" type analysis.

Best expert is u against psi_i.

$\psi_i = \nabla(f(x_i))$, $\bar{z} = \frac{1}{T}\sum_i z_i$.

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Introduces:

$w(x)$ strongly convex function with $L = 1$

"Divergence" function $V_x(y) = w(y) - (w(x) + \nabla(w(x)) \cdot (y-x))$.

For sequence of "gradients": ψ_i.

For $w(x) = \frac{1}{2}\|x-y\|^2$,

$z_k = z_0 - \alpha\|\psi_i\|^2$.

Bound "regret".

$\sum_i^k \psi_i \cdot (z_i - u) \leq \alpha^2 \sum_i^k \|\psi_i\|^2 + V_{z_k}(u) - V_{z_0}(u)$.

"Expert's" type analysis.

Best expert is u against psi_i.

$\psi_i = \nabla(f(x_i))$, $\bar{z} = \frac{1}{T}\sum_i z_i$.

$f(\bar{z}) \leq f(u) + \alpha^2 \sum_i^k \|\nabla f(z_i)\|^2 + \max_{x \in Q} w(x)$.

Mirror Descent.

Convex, continuous function: $f(x)$.

$f(y) \geq f(x) + \nabla f(x)(y-x)$.

No lipshitz condition necessary.

Gradient need not be continuous: absolute value.

Introduces:

$w(x)$ strongly convex function with $L = 1$

"Divergence" function $V_x(y) = w(y) - (w(x) + \nabla(w(x)) \cdot (y-x))$.

For sequence of "gradients": ψ_i.

For $w(x) = \frac{1}{2}\|x-y\|^2$,

$z_k = z_0 - \alpha\|\psi_i\|^2$.

Bound "regret".

$\sum_i^k \psi_i \cdot (z_i - u) \leq \alpha^2 \sum_i^k \|\psi_i\|^2 + V_{z_k}(u) - V_{z_0}(u)$.

"Expert's" type analysis.

Best expert is u against psi_i.

$\psi_i = \nabla(f(x_i))$, $\bar{z} = \frac{1}{T}\sum_i z_i$.

$f(\bar{z}) \leq f(u) + \alpha^2 \sum_i^k \|\nabla f(z_i)\|^2 + \max_{x \in Q} w(x)$.

Loss is compare to linear lower bound on function value at u.

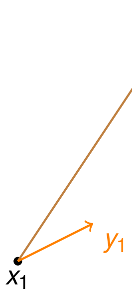
Linear Coupling

Intuition: If $\|\nabla f(x_k)\|_*^2$ large

- GD can make large primal progress $\frac{1}{2L}\|\nabla f(x_k)\|_*^2$
- MD suffers large regret $\frac{\alpha^2}{2}\|\nabla f(x)\|_*^2$
- Use primal progress to cover regret.
- Regret terms no longer accumulates,
  telescope as the primal progress.

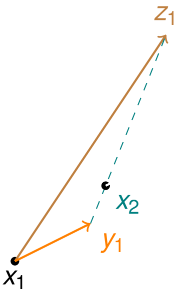
Linear Coupling

- $x_0 = y_0 = z_0$.
- Coupling: $x_{k+1} = \tau z_k + (1 - \tau) y_k$.
- MD: $z_{k+1} = \text{Mirr}(z_k, \alpha \nabla f(x_{k+1}))$
- GD: $y_{k+1} = \text{Grad}(x_{k+1})$.



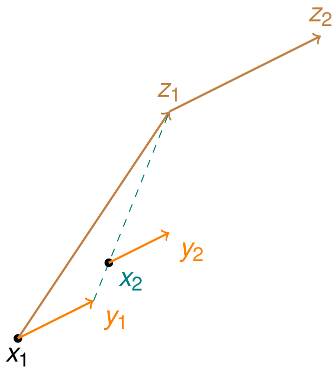
Linear Coupling

- $x_0 = y_0 = z_0$.
- Coupling: $x_{k+1} = \tau z_k + (1 - \tau) y_k$.
- MD: $z_{k+1} = \mathrm{Mirr}(z_k, \alpha \nabla f(x_{k+1}))$
- GD: $y_{k+1} = \mathrm{Grad}(x_{k+1})$.



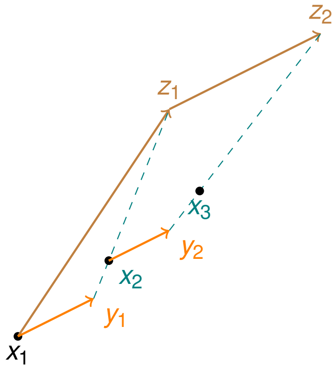
Linear Coupling

- $x_0 = y_0 = z_0$.
- Coupling: $x_{k+1} = \tau z_k + (1-\tau) y_k$.
- MD: $z_{k+1} = \mathrm{Mirr}(z_k, \alpha \nabla f(x_{k+1}))$
- GD: $y_{k+1} = \mathrm{Grad}(x_{k+1})$.



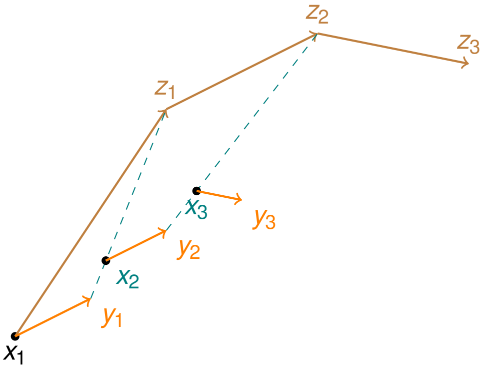
Linear Coupling

- $x_0 = y_0 = z_0$.
- Coupling: $x_{k+1} = \tau z_k + (1 - \tau) y_k$.
- MD: $z_{k+1} = \mathrm{Mirr}(z_k, \alpha \nabla f(x_{k+1}))$
- GD: $y_{k+1} = \mathrm{Grad}(x_{k+1})$.



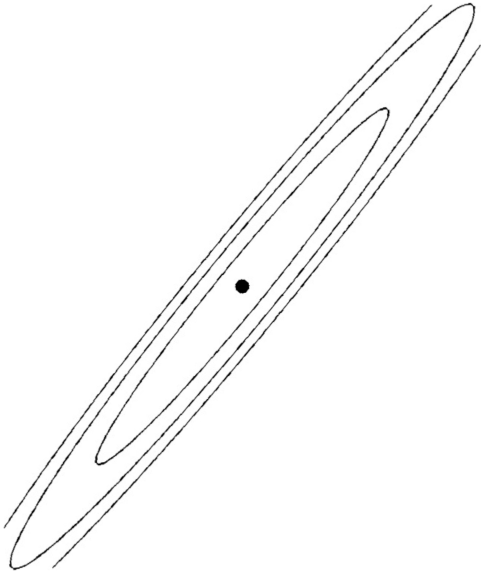
Linear Coupling

- $x_0 = y_0 = z_0$.
- Coupling: $x_{k+1} = \tau z_k + (1-\tau) y_k$.
- MD: $z_{k+1} = \text{Mirr}(z_k, \alpha \nabla f(x_{k+1}))$
- GD: $y_{k+1} = \text{Grad}(x_{k+1})$.



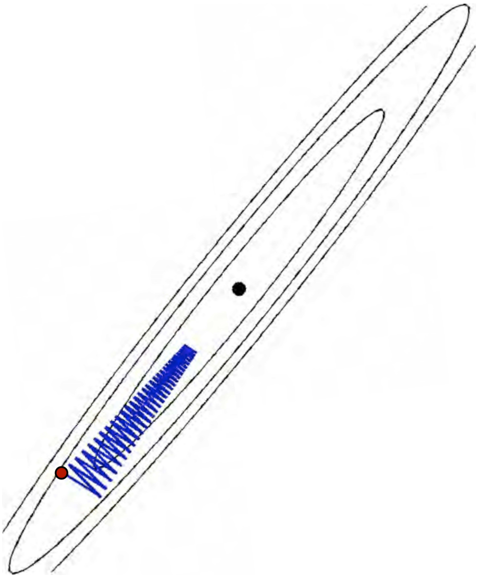
Linear Coupling

Momentum View:



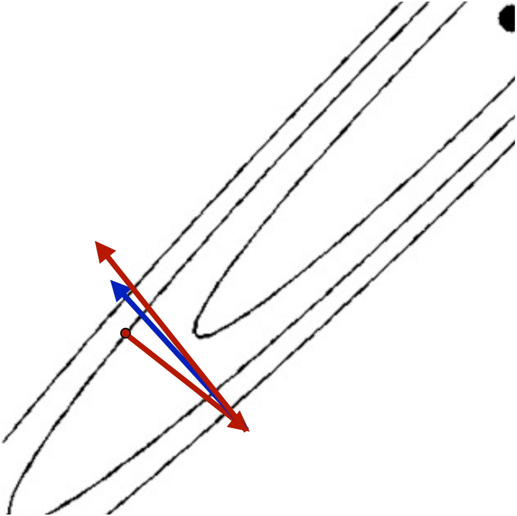
Linear Coupling

Momentum View:



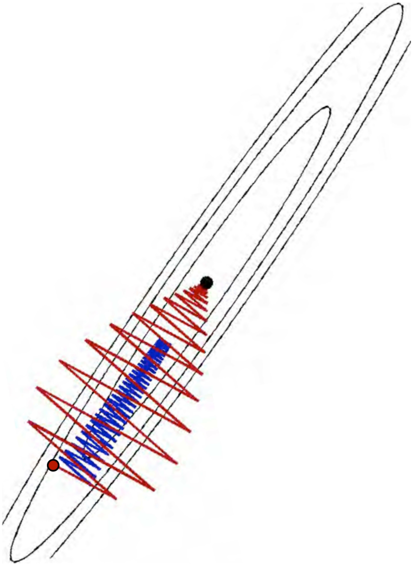
Linear Coupling

Momentum View:

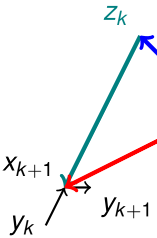


Linear Coupling

Momentum View:

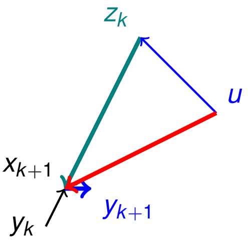


Bound $\alpha(f(x_{k+1}) - f(u)) \leq \alpha\langle \nabla f(x_{k+1}), x_{k+1} - u\rangle$



$$\begin{aligned}&\alpha\langle \nabla f(x_{k+1}), x_{k+1} - u\rangle\\ =&\alpha\langle \nabla f(x_{k+1}), z_k - u\rangle\\ &+\alpha\langle \nabla f(x_{k+1}), x_{k+1} - z_k\rangle\end{aligned}$$

Bound $\alpha(f(x_{k+1}) - f(u)) \leq \alpha\langle \nabla f(x_{k+1}), x_{k+1} - u\rangle$

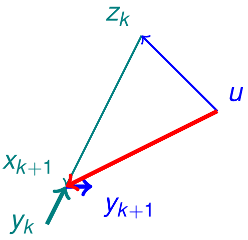


$$
\begin{aligned}
&\alpha\langle \nabla f(x_{k+1}), x_{k+1} - u\rangle \\
=&\alpha\langle \nabla f(x_{k+1}), z_k - u\rangle \\
&+ \alpha\langle \nabla f(x_{k+1}), x_{k+1} - z_k\rangle \\
\leq&\alpha^2 L(f(x_{k+1}) - f(y_{k+1})) \\
&+ V_{z_k}(u) - V_{z_{k+1}}(u)
\end{aligned}
$$

MD: $\alpha\langle \nabla f(x_{k+1}), z_k - u\rangle \leq \frac{\alpha^2}{2}\|\nabla f(x_{k+1})\|_*^2 + V_{z_k}(u) - V_{z_{k+1}}(u)$

GD: $f(x_{k+1}) - f(y_{k+1}) \geq \frac{1}{2L}\|\nabla f(x_{k+1})\|_*^2$

Bound $\alpha(f(x_{k+1}) - f(u)) \leq \alpha\langle\nabla f(x_{k+1}), x_{k+1} - u\rangle$

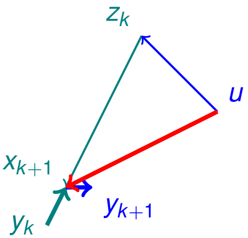


$$
\begin{aligned}
&\alpha\langle\nabla f(x_{k+1}), x_{k+1} - u\rangle \\
=&\alpha\langle\nabla f(x_{k+1}), z_k - u\rangle \\
&+ \alpha\langle\nabla f(x_{k+1}), x_{k+1} - z_k\rangle \\
\leq&\alpha^2 L(f(x_{k+1}) - f(y_{k+1})) \\
&+ V_{z_k}(u) - V_{z_{k+1}}(u) \\
&+ \alpha\langle\nabla f(x_{k+1}), \frac{1-\tau}{\tau}(y_k - x_{k+1})\rangle
\end{aligned}
$$

Coupling:

$x_{k+1} = \tau z_k + (1-\tau)y_k \quad \rightarrow \quad \tau(x_{k+1} - z_k) = (1-\tau)(y_k - x_{k+1})$

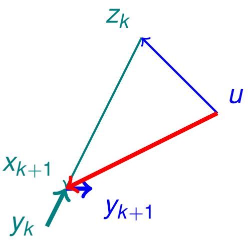
Bound $\alpha(f(x_{k+1}) - f(u)) \leq \alpha\langle\nabla f(x_{k+1}), x_{k+1} - u\rangle$



$$
\begin{aligned}
&\alpha\langle\nabla f(x_{k+1}), x_{k+1} - u\rangle \\
=&\alpha\langle\nabla f(x_{k+1}), z_k - u\rangle \\
&+ \alpha\langle\nabla f(x_{k+1}), x_{k+1} - z_k\rangle \\
\leq&\alpha^2 L(f(x_{k+1}) - f(y_{k+1})) \\
&+ V_{z_k}(u) - V_{z_{k+1}}(u) \\
&+ \frac{1-\tau}{\tau}\alpha(f(y_k) - f(x_{k+1}))
\end{aligned}
$$

$$
\begin{aligned}
\alpha\langle\nabla f(x_{k+1}), \tfrac{1-\tau}{\tau}(y_k - x_{k+1})\rangle &= \alpha\tfrac{1-\tau}{\tau}\langle\nabla f(x_{k+1}), y_k - x_{k+1}\rangle) \\
&\leq \tfrac{1-\tau}{\tau}\alpha(f(y_k) - f(x_{k+1}))
\end{aligned}
$$

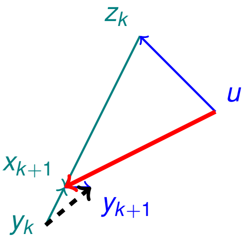
Bound $\alpha(f(x_{k+1}) - f(u)) \leq \alpha\langle \nabla f(x_{k+1}), x_{k+1} - u\rangle$



$$
\begin{aligned}
&\alpha\langle \nabla f(x_{k+1}), x_{k+1} - u\rangle \\
=&\alpha\langle \nabla f(x_{k+1}), z_k - u\rangle \\
&+ \alpha\langle \nabla f(x_{k+1}), x_{k+1} - z_k\rangle \\
\leq&\alpha^2 L(f(x_{k+1}) - f(y_{k+1})) \\
&+ V_{z_k}(u) - V_{z_{k+1}}(u) \\
&+ \alpha^2 L(f(y_k) - f(x_{k+1}))
\end{aligned}
$$

Let $\alpha^2 L = \dfrac{1-\tau}{\tau}\alpha$

Bound $\alpha(f(x_{k+1}) - f(u)) \le \alpha\langle \nabla f(x_{k+1}), x_{k+1} - u\rangle$



$$
\begin{aligned}
&\alpha\langle \nabla f(x_{k+1}), x_{k+1} - u\rangle \\
=&\alpha\langle \nabla f(x_{k+1}), z_k - u\rangle \\
&+ \alpha\langle \nabla f(x_{k+1}), x_{k+1} - z_k\rangle \\
\le&\alpha^2 L(\cancel{f(x_{k+1})} - f(y_{k+1})) \\
&+ V_{z_k}(u) - V_{z_{k+1}}(u) \\
&+ \alpha^2 L(f(y_k) - \cancel{f(x_{k+1})})
\end{aligned}
$$

Let $\alpha^2 L = \dfrac{1-\tau}{\tau}\alpha$

Both components telescope!

Linear Coupling

- Summing over $0, \ldots, T-1$, with $\bar{x} = \frac{1}{T}\sum_i x_i$

$$f(\bar{x}) - f(u) \leq \frac{\alpha L}{T}(f(y_0) - f(y_T)) + \frac{V_{z_0}(u)}{\alpha T}$$

Linear Coupling

- Summing over $0, \ldots, T-1$, with $\bar{x} = \frac{1}{T}\sum_i x_i$

$$f(\bar{x}) - f(u) \leq \frac{\alpha L}{T}(f(y_0) - f(y_T)) + \frac{V_{z_0}(u)}{\alpha T}$$

- If $f(y_0) - \text{OPT} \leq d$, diameter $V_{z_0}(u) \leq \Theta$, $\alpha = \sqrt{\frac{\Theta}{Ld}}, T = 4\sqrt{\frac{L\Theta}{d}}$

$$f(\bar{x}) - f(u) \leq \frac{\alpha L d + \Theta/\alpha}{T} \leq \frac{d}{2}$$

Linear Coupling

- Summing over $0,\ldots,T-1$, with $\bar{x} = \frac{1}{T}\sum_i x_i$

$$f(\bar{x}) - f(u) \le \frac{\alpha L}{T}(f(y_0) - f(y_T)) + \frac{V_{z_0}(u)}{\alpha T}$$

- If $f(y_0) - \mathrm{OPT} \le d$, diameter $V_{z_0}(u) \le \Theta$,
$\alpha = \sqrt{\frac{\Theta}{Ld}}, T = 4\sqrt{\frac{L\Theta}{d}}$

$$f(\bar{x}) - f(u) \le \frac{\alpha L d + \Theta/\alpha}{T} \le \frac{d}{2}$$

- In $T = 4\sqrt{\frac{L\Theta}{d}}$ iterations,

$$f(x_0) - \mathrm{OPT} \le d \quad \rightarrow \quad f(\bar{x}) - \mathrm{OPT} \le \frac{d}{2}$$

To get ε-approximation:

$$T = O(\sqrt{\frac{L\Theta}{\varepsilon}} + \sqrt{\frac{L\Theta}{2\varepsilon}} + \ldots) = O(\sqrt{\frac{L\Theta}{\varepsilon}})$$

Linear Coupling

- With $\alpha_k = \frac{k+1}{2L}$, can remove phases, and have $f(y_T) - f(u) \leq \varepsilon$ after $T = O(\sqrt{\frac{L\Theta}{\varepsilon}})$ iterations.
  Almost the same as Nesterov's.
- GD: $O(\frac{LR^2}{2})$ v.s. MD: $O(\frac{\rho^2\Theta}{\varepsilon^2})$ v.s. AGD: $O(\sqrt{\frac{L\Theta}{\varepsilon}})$

Spectral Theorem.

For real symmetric matrix, A, there exists a set of unit vectors $u_1, \ldots, u_n$, and real numbers λ_i such $v_i \perp v_j$ and $Au_i = \lambda_i u_i$.

Spectral Theorem.

For real symmetric matrix, A, there exists a set of unit vectors $u_1, \ldots, u_n$, and real numbers λ_i such $v_i \perp v_j$ and $Au_i = \lambda_i u_i$.

One idea in proof:

Spectral Theorem.

For real symmetric matrix, A, there exists a set of unit vectors $u_1, \ldots, u_n$, and real numbers λ_i such $v_i \perp v_j$ and $Au_i = \lambda_i u_i$.

One idea in proof:

$\min x^T Ax$ s.t $\|x\| = 1$.

Spectral Theorem.

For real symmetric matrix, A, there exists a set of unit vectors $u_1, \ldots, u_n$, and real numbers λ_i such $v_i \perp v_j$ and $Au_i = \lambda_i u_i$.

One idea in proof:

$$\min x^T Ax \text{ s.t } \|x\| = 1.$$

How?
Constrained optimization.

Spectral Theorem.

For real symmetric matrix, A, there exists a set of unit vectors $u_1, \ldots, u_n$, and real numbers λ_i such $v_i \perp v_j$ and $Au_i = \lambda_i u_i$.

One idea in proof:

$\min x^T Ax$ s.t $\|x\| = 1$.

How?

Constrained optimization.

$x^T Ax - \lambda(\|x\|^2 - 1)$

Spectral Theorem.

For real symmetric matrix, A, there exists a set of unit vectors $u_1, \ldots, u_n$, and real numbers λ_i such $v_i \perp v_j$ and $Au_i = \lambda_i u_i$.

One idea in proof:

$$\min x^T Ax \text{ s.t } \|x\| = 1.$$

How?
Constrained optimization.

$$x^T Ax - \lambda(\|x\|^2 - 1)$$

Fix λ.

Spectral Theorem.

For real symmetric matrix, A, there exists a set of unit vectors $u_1, \ldots, u_n$, and real numbers λ_i such $v_i \perp v_j$ and $Au_i = \lambda_i u_i$.

One idea in proof:

$$\min x^T A x \text{ s.t } \|x\| = 1.$$

How?

Constrained optimization.

$$x^T A x - \lambda(\|x\|^2 - 1)$$

Fix λ. Minimize for x.

Spectral Theorem.

For real symmetric matrix, A, there exists a set of unit vectors $u_1, \ldots, u_n$, and real numbers λ_i such $v_i \perp v_j$ and $Au_i = \lambda_i u_i$.

One idea in proof:

$\min x^T Ax$ s.t $\|x\| = 1$.

How?

Constrained optimization.

$x^T Ax - \lambda(\|x\|^2 - 1)$

Fix λ. Minimize for x.

$2Ax - 2\lambda x = 0$

Spectral Theorem.

For real symmetric matrix, A, there exists a set of unit vectors $u_1, \ldots, u_n$, and real numbers λ_i such $v_i \perp v_j$ and $Au_i = \lambda_i u_i$.

One idea in proof:

$\min x^T Ax$ s.t $\|x\| = 1$.

How?

Constrained optimization.

$x^T Ax - \lambda(\|x\|^2 - 1)$

Fix λ. Minimize for x.

$2Ax - 2\lambda x = 0$ $Ax = \lambda x$.

Spectral Theorem.

For real symmetric matrix, A, there exists a set of unit vectors $u_1, \ldots, u_n$, and real numbers λ_i such $v_i \perp v_j$ and $Au_i = \lambda_i u_i$.

One idea in proof:

$$\min x^T Ax \text{ s.t } \|x\| = 1.$$

How?

Constrained optimization.

$$x^T Ax - \lambda(\|x\|^2 - 1)$$

Fix λ. Minimize for x.

$2Ax - 2\lambda x = 0$ $Ax = \lambda x$.

Minimizer is “eigenvector”.

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1)...(u_n, \lambda_n)$.

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1) \ldots (u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1) ... (u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$
$u_i \perp u_j$

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1)...(u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$
$u_i \perp u_j$
$\lambda_1 > \lambda_2 ... > \lambda_n$

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1)...(u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$
$u_i \perp u_j$
$\lambda_1 > \lambda_2 ... > \lambda_n$

Take "random" x_0.

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1)...(u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$
$u_i \perp u_j$
$\lambda_1 > \lambda_2 ... > \lambda_n$

Take "random" x_0.

$x_0 = a_1 u_1 + \ldots a_2 u_2 + \cdots + a_n u_n$.

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1)...(u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$
$u_i \perp u_j$
$\lambda_1 > \lambda_2 ... > \lambda_n$

Take "random" x_0.

$x_0 = a_1 u_1 + \ldots a_2 u_2 + \cdots + a_n u_n$.

$x_{t+1} = Ax_t = A^t x_0$

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1)...(u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$
$u_i \perp u_j$
$\lambda_1 > \lambda_2 ... > \lambda_n$

Take "random" x_0.

$x_0 = a_1 u_1 + \ldots a_2 u_2 + \cdots + a_n u_n$.

$x_{t+1} = Ax_t = A^t x_0$

$x_{t+1} = \lambda_1^t a_1 u_0 + \cdots \lambda_n^t a_n u_n$

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1)...(u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$
$u_i \perp u_j$
$\lambda_1 > \lambda_2 ... > \lambda_n$

Take "random" x_0.

$x_0 = a_1 u_1 + \dots a_2 u_2 + \cdots + a_n u_n$.

$x_{t+1} = Ax_t = A^t x_0$

$x_{t+1} = \lambda_1^t a_1 u_0 + \cdots \lambda_n^t a_n u_n$

Since $\lambda_1 > \lambda 2 > \dots$.

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1)...(u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$
$u_i \perp u_j$
$\lambda_1 > \lambda_2 ... > \lambda_n$

Take "random" x_0.

$x_0 = a_1 u_1 + \ldots a_2 u_2 + \cdots + a_n u_n$.

$x_{t+1} = Ax_t = A^t x_0$

$x_{t+1} = \lambda_1^t a_1 u_0 + \cdots \lambda_n^t a_n u_n$

Since $\lambda_1 > \lambda 2 > \ldots$.

$x_t / |x_t|$ converges to u_1.

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1)...(u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$
$u_i \perp u_j$
$\lambda_1 > \lambda_2 ... > \lambda_n$

Take “random” x_0.

$x_0 = a_1 u_1 + \ldots a_2 u_2 + \cdots + a_n u_n$.

$x_{t+1} = Ax_t = A^t x_0$

$x_{t+1} = \lambda_1^t a_1 u_0 + \cdots \lambda_n^t a_n u_n$

Since $\lambda_1 > \lambda 2 > \ldots$.

$x_t / |x_t|$ converges to u_1.

Get rest?

Finding an eigenvector: power method

A has eigenpairs: $(u_1, \lambda_1)...(u_n, \lambda_n)$.

$Au_i = \lambda_i u_i$
$u_i \perp u_j$
$\lambda_1 > \lambda_2 ... > \lambda_n$

Take "random" x_0.

$x_0 = a_1 u_1 + \ldots a_2 u_2 + \cdots + a_n u_n$.

$x_{t+1} = Ax_t = A^t x_0$

$x_{t+1} = \lambda_1^t a_1 u_0 + \cdots \lambda_n^t a_n u_n$

Since $\lambda_1 > \lambda 2 > \ldots$.

$x_t / |x_t|$ converges to u_1.

Get rest? Orthoganalize and induction.

Iterative solution of linear system.

$Ax = b.$

Iterative solution of linear system.

$Ax = b$.

$x_0 = b$.

Iterative solution of linear system.

$Ax = b.$

$x_0 = b.$

$r_t = b - Ax_t$

Iterative solution of linear system.

$Ax = b.$

$x_0 = b.$

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t.$

Iterative solution of linear system.

$Ax = b$.

$x_0 = b$.

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t$.

Iterative solution of linear system.

$Ax = b$.

$x_0 = b$.

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t$.

Analysis:

Iterative solution of linear system.

$Ax = b$.

$x_0 = b$.

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t$.

Analysis:
$r_t = b - Ax_t = Ax^* - Ax_t = A(x^* - x_t)$

Iterative solution of linear system.

$Ax = b$.

$x_0 = b$.

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t$.

Analysis:

$r_t = b - Ax_t = Ax^* - Ax_t = A(x^* - x_t)$
$r_t = r_t - \alpha A r_t = (I - \alpha A) r_i$

Iterative solution of linear system.

$Ax = b$.

$x_0 = b$.

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t$.

Analysis:
$r_t = b - Ax_t = Ax^* - Ax_t = A(x^* - x_t)$
$r_t = r_t - \alpha A r_t = (I - \alpha A) r_i$

Iterative solution of linear system.

$Ax = b$.

$x_0 = b$.

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t$.

Analysis:
$r_t = b - Ax_t = Ax^* - Ax_t = A(x^* - x_t)$
$r_t = r_t - \alpha A r_t = (I - \alpha A) r_i$

If $\lambda_1 > ... > \lambda_n$.

Iterative solution of linear system.

$Ax = b$.

$x_0 = b$.

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t$.

Analysis:
$r_t = b - Ax_t = Ax^* - Ax_t = A(x^* - x_t)$
$r_t = r_t - \alpha A r_t = (I - \alpha A) r_i$

If $\lambda_1 > ... > \lambda_n$.
$r_t = (1 - \alpha\lambda_1)^t a_0 u_0 + \cdots (1 - \alpha\lambda_n)^t a_n u_n$.

Iterative solution of linear system.

$Ax = b.$

$x_0 = b.$

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t.$

Analysis:
$r_t = b - Ax_t = Ax^* - Ax_t = A(x^* - x_t)$
$r_t = r_t - \alpha A r_t = (I - \alpha A) r_i$

If $\lambda_1 > ... > \lambda_n$.
$r_t = (1 - \alpha\lambda_1)^t a_0 u_0 + \cdots (1 - \alpha\lambda_n)^t a_n u_n.$

Set $\alpha < \lambda_n / \lambda_1$, and $(1 - \alpha) < 1$.

Iterative solution of linear system.

$Ax = b$.

$x_0 = b$.

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t$.

Analysis:
$r_t = b - Ax_t = Ax^* - Ax_t = A(x^* - x_t)$
$r_t = r_t - \alpha A r_t = (I - \alpha A) r_i$

If $\lambda_1 > ... > \lambda_n$.
$r_t = (1 - \alpha\lambda_1)^t a_0 u_0 + \cdots (1 - \alpha\lambda_n)^t a_n u_n$.

Set $\alpha < \lambda_n/\lambda_1$, and $(1 - \alpha) < 1$.

Converges at rate λ_1/λ_n.

Iterative solution of linear system.

$Ax = b$.

$x_0 = b$.

$r_t = b - Ax_t$
$x_{t+1} = x_t + \alpha r_t$.

Analysis:
$r_t = b - Ax_t = Ax^* - Ax_t = A(x^* - x_t)$
$r_t = r_t - \alpha A r_t = (I - \alpha A) r_i$

If $\lambda_1 > ... > \lambda_n$.
$r_t = (1 - \alpha\lambda_1)^t a_0 u_0 + \cdots (1 - \alpha\lambda_n)^t a_n u_n$.

Set $\alpha < \lambda_n/\lambda_1$, and $(1 - \alpha) < 1$.

Converges at rate λ_1/λ_n.

Assuming positive λ's.