$$X_i \in \mathbb{R}^n \ \& \ A \in \mathbb{R}^{n \times n}$$

<u>Objective</u>: $\underset{w_t \in \mathbb{R}^n, \|w_t\|_2 = 1}{\text{Min}} - \left(\frac{w_t^T A w_t}{w_t^T w_t} \right)$, where $A = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T$

$$\underset{w_t \in \mathbb{R}^n, \|w_t\|_2 = 1}{\text{Min}} \quad - (w_t^T A w_t)$$

Let $f(w_t) = - w_t^T A w_t = E_\xi(\overbrace{f(w_t; \xi)}^{\text{stochastic loss function}})$

then

$$\nabla f(w_t) \Big|_{\|w_t\|_2 = 1} = -2 (A - (w_t^T A w_t) I) w_t$$

$$\Rightarrow g_t = \underbrace{\nabla f(w_t, \xi_t)}_{\text{stochastic gradient}} = -2 (xx^T - (w_t^T (xx^T) w_t) I) w_t \quad \text{--- (1)}$$

where xx^T = stochastic approx. of A

<u>Goal</u>:

<u>Update rule</u> [after naively applying adaptive term]

$$XX^T = x_{i_t} x_{i_t}^T$$ [where $i_t \in [n]$, but we are omitting subscripts here]

$$w'_{t+1} = (XX^T) w_t + \eta_t \frac{\hat{m}_t}{(\hat{V}_t)^p}$$

or

$$w'_{t+1} = (XX^T) w_t + \eta_t \hat{V}_t^{-p} \hat{m}_t$$

where

$$\hat{V}_t^{-p} = \mathrm{diag}(\hat{v}_{t,1}^{-p}, \hat{v}_{t,2}^{-p} \ldots \hat{v}_{t,n}^{-p})$$

and

$$\hat{v}_t = \max(\hat{v}_{t-1}, v_t)$$

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \; ; \; \hat{m}_t = \frac{m_t}{(1-\beta_1^t)}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2$$

Finally, normalization

$$w_{t+1} = \frac{w'_{t+1}}{\|w'_{t+1}\|_2}$$

Scanned by CamScanner

$\langle \cdot, \cdot \rangle \rightarrow$ denotes inner product

ALGORITHM

<u>Input:</u> w_0, step-size $\{\eta_t\}_{t=1}^{T}$, β_1, β_2, p.

1. $m_0 \leftarrow 0$, $\hat{v}_0 \leftarrow 0$, $v_0 \leftarrow 0$

2. for $t=1$ to T do:

$g_t = \nabla f(w_t, \xi_t)$ [see eq^n. 1]

pick $i_t \in [n]$ uniformly at random

$$w_t' = (x_{i_t} x_{i_t}^T) w_{t-1} + \eta_t \hat{V}_t^{-p} \hat{m}_t$$

$$w_t = \frac{w_t'}{\|w_t'\|}$$

3. end for
4. <u>output:</u> w_T

<u>Goal:</u> To bound $E(1 - \langle w_T, U_1 \rangle^2)$

where

Eigendecomposition of $A = \sum_{i=1}^{n} \lambda_i U_i U_i^T$

s.t. $U_i^T U_j = 0 \ \forall \ i \neq j$ &

$\|U_i\|_2 = 1$

☐ We start by bounding

$$E\left(1 - \langle w_{t+1}, u_1 \rangle^2\right)$$

Consider the analysis in [Shamir 2015]

Since u_i's span the entire $\mathbb{R}^n$

$\Rightarrow$

$$w'_{t+1} = \sum_{i=1}^{n} c_i u_i$$

$$\Rightarrow \quad \|w'_{t+1}\|_2^2 = \sum_{i=1}^{n} c_i^2 = \sum_{i=1}^{n} \langle w'_{t+1}, u_i \rangle^2$$

Also, consider

$$\langle w'_{t+1}, u_i \rangle = \underbrace{u_i^T (XX^T) w_t}_{a_i} + \underbrace{\eta_t \langle \hat{V}_t \hat{m}_t, u_i \rangle}_{z_i}$$

$$\langle w'_{t+1}, u_i \rangle = a_i + z_i$$

So

$$E\left(\langle w_{t+1}, U_1\rangle^2\right) = E\left(\frac{\langle w'_{t+1}, U_1\rangle^2}{\|w'_{t+1}\|_2^2}\right)$$

$$= E\left(\frac{\langle w'_{t+1}, U_1\rangle^2}{\sum_{i=1}^{n}\langle w'_{t+1}, U_i\rangle^2}\right)$$

$$= E\left(\frac{(a_1 + z_1)^2}{\sum_{i=1}^{n}(a_i + z_i)^2}\right)$$

① Now, the problem is to bound the last term using a second-order Taylor approx. based analysis done in (Shamir 2015)

② Once we get a bound, the next step is to try to prove something similar to lemma 1 of Shamir 2015.

③ Next step would be to use reoccurence on $E(1 - \langle w_{t+1}, U_1\rangle^2)$ from $t=1 \ldots$ to T to get a bound over $E(1 - \langle w_T, U_1\rangle^2)$.