Convex Optimization 10-725, Lecture 15: Sum of Square: The Ultimate form of Convex Optimization

Yuanzhi Li

Assistant Professor, Carnegie Mellon University
Principle Researcher, Microsoft Research

Today

So Far

- We have been mainly talking about convex optimization.
- Although all the algorithms we have learned can be applied to general settings (non-convex).
- For example, people use SGD + M / Adam / Projected stochastic gradient descent in deep learning.
- However, we only have convergence guarantees of these algorithms in the convex setting.
- Question: Given a general optimization problem, can we make it convex?

"Convexification"

- Question: Given a general optimization problem, can we make it convex?
- Meaning that given a general set $\mathcal{D}$ and a general function f where we want to solve

$$\min_{x \in \mathcal{D}} f(x)$$

- Can we write an equivalent optimization problem

$$\min_{y \in \mathcal{D}'} g(y)$$

- Such that $\mathcal{D}'$ is a convex set, g is a convex function, and there is a KNOWN function r such that $r(y) = x$.
- After finding y^*, we can apply r and recover x^*.

"Convexification"

- Example: non convex optimization over a non-convex constraint set:

$$\min_{x_1^4+x_2^4 \le 1, x_1^2 - x_2^2 \ge 0} x_1^2 - 3x_2^2$$

- This is equivalent to a convex optimization

$$\min_{y_1^2+y_2^2 \le 1, y_1 - y_2 \ge 0, y_1 \ge 0, y_2 \ge 0} y_1 - 3y_2$$

- Such that $x_1^2 = y_1, x_2^2 = y_2$.
- Question: How do we do it in general?
- For example if we have:

$$\min_{x_1^4+x_2^4+x_1-x_2 \le 1, x_1^2 - x_1x_2 - x_2^2 \ge 0} x_1^2 - 3x_2^2 + 4x_1$$

“Convexification”

- We will talk about the Sum of Squares approach, which can automatically turn any (analytic) optimization problem into a convex optimization problem.
- Sum of Squares approach is very simple, once we understand its spirit.

Sum of Squares

- Motivation: Let us first consider un-constraint optimization: $\min f(x)$ for general function f.
- Example of a very non-convex function:

$$f(x) = x_1^4 - 2x_1x_2 + x_2^4 - x_2^3 + 2x_2 - 3x_1^2$$

- How do we minimize this thing?
- How do we "convexifiy" this thing?

Sum of Squares

- Let us start with an analytic function f, those functions can be written as:

$$f(x) = \sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C_{\vec{\alpha}} \prod_{i \in [d]} x_i^{\vec{\alpha}_i}$$

- Starting point: turn $\min f$ into a linear optimization problem by
- Creating a new variable $X_{\vec{\alpha}}$ for each monomial $\prod_{i \in [d]} x_i^{\vec{\alpha}_i}$.
- Let $X = \{X_{\vec{\alpha}}\}_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d}$, then we can write $f(x)$ as:

$$G(X) = \sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C_{\vec{\alpha}} X_{\vec{\alpha}}$$

Sum of Squares

- Let $X = \{X_{\vec{\alpha}}\}_{\vec{\alpha}\text{:non-negative vectors in } \mathbb{Z}^d}$, then we can write $f(x)$ as:

$$G(X) = \sum_{\vec{\alpha}\text{:non-negative vectors in } \mathbb{Z}^d} C_{\vec{\alpha}} X_{\vec{\alpha}}$$

- Example: $f(x) = x_1^4 - 2x_1x_2 + x_2^2$. Then we have: $X = (X_{(4,0)}, X_{(1,1)}, X_{(0,2)})$ and

$$G(X) = X_{(4,0)} - 2X_{(1,1)} + X_{(0,2)}$$

- Observation: No matter how complicated (non-convex) f is, G is always linear (convex).

Sum of Squares

- Example: $f(x) = x_1^4 - 2x_1x_2 + x_2^2$. Then we have: $X = (X_{(4,0)}, X_{(1,1)}, X_{(0,2)})$ and

$$G(X) = X_{(4,0)} - 2X_{(1,1)} + X_{(0,2)}$$

- Observation: No matter how complicated (non-convex) f is, G is always linear (convex).
- But clearly, minimizing G w.r.t. X can not recover the minimizer of f, because the variables $X_{\vec{\alpha}}$ are constraint and dependent.
- For example, $x_2^2 \geq 0$, so clearly we need to add constraint $X_{(0,2)} \geq 0$
- But more importantly,

$$x_1^4 + 1/4 \geq x_1^2, \quad x_1^2 + 2x_1x_2 + x_2^2 \geq 0$$

- So we should also add constraints like

$$X_{(4,0)} + 1/4 \geq X_{(2,0)}, \quad X_{(2,0)} + 2X_{(1,1)} + X_{(0,2)} \geq 0$$

Sum of Squares

- But more importantly,

$$x_1^4 + 1/4 \geq x_1^2, \quad x_1^2 + 2x_1x_2 + x_2^2 \geq 0$$

- So we should also add constraints like

$$X_{(4,0)} + 1/4 \geq X_{(2,0)}, \quad X_{(2,0)} + 2X_{(1,1)} + X_{(0,2)} \geq 0$$

- Main Question: What are the sets of constraints we need to add?

Sum of Squares

- Main observation: For every polynomial $h(x)$, we must have

$$h(x)^2 \geq 0$$

.

- Let $C(h)$ be the coefficient of h in monomial basis, so

$$h(x) = \sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C(h)_{\vec{\alpha}} \prod_{i \in [d]} x_i^{\vec{\alpha}_i}$$

- So we need to add the constraint:

$$\sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C(h^2)_{\vec{\alpha}} X_{\vec{\alpha}} \geq 0$$

- For every polynomial h.
- Example: $(x_1 + 1)^2 \geq 0$, so we need to add $X_{(2,0)} + 2X_{(1,0)} + 1 \geq 0$.

Sum of Squares

- To minimize the general function

$$f(x) = \sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C_{\vec{\alpha}} \prod_{i \in [d]} x_i^{\vec{\alpha}_i}$$

- We want to alternatively minimize

$$G(X) = \sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C_{\vec{\alpha}} X_{\vec{\alpha}}$$

- With constraints:

$$\sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C(h^2)_{\vec{\alpha}} X_{\vec{\alpha}} \geq 0$$

- For every polynomial h.
- Main observation: This is convex optimization over convex constraint set (polytope)!

Sum of Squares

Theorem

The new minimization problem is equivalent to the original problem, in the sense that

- *(1).* $\min_{X \text{ satisfies all constraints}} G(X) = \min_x f(x)$.
- *(2). As long as f has a unique minimizer x^*, then G has a unique minimizer X^* and*

$$x^* = (X^*_{(1,0,0,...,0)}, X^*_{(0,1,0,...,0)}, ..., X^*_{(0,0,0,...,0,1)})$$

Sum of Squares

- The new minimization problem is convex, and it is equivalent to the original problem.
- Here, the new variables $X_{\vec{\alpha}}$ is called the Pseudo-Expectation.
- For notation simplicity, for a polynomial g, we use $\tilde{\mathbb{E}}_X g$ to denote

$$\tilde{\mathbb{E}}_X g = \sum_{\vec{\alpha}} C(g)_{\vec{\alpha}} X_{\vec{\alpha}}$$

- We can alternatively write the minimization as:

$$\min \tilde{\mathbb{E}}_X f, \text{s.t. for all polynomial } h : \tilde{\mathbb{E}}_X h^2 \geq 0$$

Sum of Squares

- Now we need to worry about the constraints:

$$\sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C(h^2)_{\vec{\alpha}} X_{\vec{\alpha}} \geq 0$$

- For every polynomial h.
- Clearly, this is still inefficient, since there are infinitely mainly polynomials h.
- Idea 1: Restrict the degree of h^2 to be no more than D: $\deg(h^2) \leq D$. Meaning that we only consider h such that for every $\vec{\alpha}$ with $\|\vec{\alpha}\|_1 > D$, $C(h^2)_{\vec{\alpha}} = 0$.

- So we need to add the constraint:

$$\sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C(h^2)_{\vec{\alpha}} X_{\vec{\alpha}} \geq 0$$

- For every polynomial h.
- Clearly, this is inefficient, since there are infinitely mainly polynomials h.
- Idea 1: Restrict the degree of h^2 to be no more than D: $\deg(h^2) \leq D$. Meaning that we only consider h such that for every $\vec{\alpha}$ with $\|\vec{\alpha}\|_1 > D$, $C(h^2)_{\vec{\alpha}} = 0$.

Sum of Squares

- Idea 1: Restrict the degree of h^2 to be no more than D: $\deg(h^2) \leq D$. Meaning that we only consider h such that for every $\vec{\alpha}$ with $\|\vec{\alpha}\|_1 > d$, $C(h^2)_{\vec{\alpha}} = 0$.
- So we only add constraints

$$\sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C(h^2)_{\vec{\alpha}} X_{\vec{\alpha}} \geq 0$$

- For every polynomial h with $\deg(h^2) \leq D$.
- D is some hyper-parameter you will choose in practice.

Sum of Squares

- So we only add constraints

$$\sum_{\vec{\alpha}\text{:non-negative vectors in } \mathbb{Z}^d} C(h^2)_{\vec{\alpha}} X_{\vec{\alpha}} \geq 0$$

- For every polynomial h with $\deg(h^2) \leq D$.
- But still, there are infinitely many polynomials h with $\deg(h^2) \leq D$.

Sum of Squares

- So we only add constraints

$$\sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C(h^2)_{\vec{\alpha}} X_{\vec{\alpha}} \geq 0$$

- For every polynomial h with $\deg(h^2) \leq D$.
- Main observation: Let $M_D(X)$ be the matrix such that

$$[M_D(X)]_{\vec{\alpha},\vec{\alpha}'} = X_{\vec{\alpha}+\vec{\alpha}'}$$

  for all $\vec{\alpha}, \vec{\alpha}'$: non-negative vectors in $\mathbb{Z}^d$ with $\|\vec{\alpha}\|_1 \leq D/2, \|\vec{\alpha}'\|_1 \leq D/2$
- Then the original set of constraints is equivalent to the following constraint:

$$M_D(X) \succeq 0$$

- This is because $\sum_{\vec{\alpha}} C(h^2)_{\vec{\alpha}} X_{\vec{\alpha}} = c^\top M_D(X) c$ for $c = (C(h)_{\vec{\alpha}})_{\vec{\alpha}}$.

Degree D Sum of Squares

- Degree D Sum of Squares for un-constraint optimization: To minimize $\min f(x)$:
- It is equivalent to minimizing the new problem:

$$G(X) = \sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C_{\vec{\alpha}}(f) X_{\vec{\alpha}}$$

- With constraints that form a convex set:

$$M_D(X) \succeq 0$$

- Where for all $\vec{\alpha}, \vec{\alpha'}$: non-negative vectors in $\mathbb{Z}^d$ with $\|\vec{\alpha}\|_1 \leq D/2, \|\vec{\alpha}'\|_1 \leq D/2$ ":

$$[M_D(X)]_{\vec{\alpha}, \vec{\alpha'}} = X_{\vec{\alpha}+\vec{\alpha'}}$$

- Then $x^* = (X_{(1,0,\ldots,0)}, X_{(0,1,\ldots,0)}, X_{(0,0,\ldots,1)})$.
- This is the Sum of Squares program, where $X_{\vec{\alpha}}$ are called the "pseudo-expectations".
- We denote $\tilde{E}(f) := \sum_{\vec{\alpha}} C_{\vec{\alpha}}(f) X_{\vec{\alpha}}$

Sum of Squares

- To solve $\min_{X: M_D(X) \geq 0} G(X)$, we need to use the interior point method, the self-concordant barrier function for this PSD cone is

$$R(X) = -\log\det(M_D(X))$$

Sum of Squares

- For constraint minimization $\min f(x)$ such that $\{h_i(x) \geq 0\}_{i \in [m]}$:
- Choice 1: Use duality to turn it into unconstraint optimization over x.
- Choice 2: We still minimize

$$G(X) = \sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C_{\vec{\alpha}} X_{\vec{\alpha}}$$

- Now we need to add new constraints:

$$\sum_{\vec{\alpha}:\text{non-negative vectors in } \mathbb{Z}^d} C(h_i h^2)_{\vec{\alpha}} X_{\vec{\alpha}} \geq 0$$

- For every $i \in [m]$ and every polynomial h.
- Theorem: With these new constraints, the optimization is also equivalent to $\min f(x)$ such that $\{h_i(x) \geq 0\}_{i \in [m]}$ (not going to prove here)

Sum of Squares

- Main Application: Robust linear regression.
- Suppose we are given N data points $(x^{(i)}, y^{(i)})$, $1-\tau$ fraction of them satisfies $(y - \langle w^*, x \rangle)^2 \le \varepsilon^2$, and τ fraction of the data is arbitrarily corrupted (both x and y).
- Solving linear regression

$$\min_w \sum_{i \in [N]} \left(\langle x^{(i)}, w \rangle - y^{(i)} \right)^2$$

will not work.

- We want to solve the Robust Linear Regression

$$\min_{\alpha_i} \min_w \sum_{i \in [N]} \alpha_i \left(\langle x^{(i)}, w \rangle - y^{(i)} \right)^2$$

Such that $\alpha_i^2 = \alpha_i$, $\frac{1}{N} \sum_i \alpha_i \ge 1 - \tau$.

Sum of Squares

- Main Application: Robust linear regression.
- We want to solve the Robust Linear Regression

$$\min_{\alpha_i} \min_{w} \sum_{i \in [N]} \alpha_i \left(\langle x^{(i)}, w \rangle - y^{(i)} \right)^2$$

  Such that $\alpha_i^2 = \alpha_i$, $\frac{1}{N} \sum_i \alpha_i \geq 1 - \tau$.
- This is non-convex optimization.
- Main Claim: We can use degree 8 sum of squares (over both α and w) to solve this.

Sum of Squares

- We want to solve the Robust Linear Regression

$$\min_{\alpha_i} \min_{w} \sum_{i \in [N]} \alpha_i \left(\langle x^{(i)}, w \rangle - y^{(i)} \right)^2$$

  Such that $\alpha_i^2 = \alpha_i$, $\frac{1}{N} \sum_i \alpha_i \geq 1 - \tau$.

- Let $\tilde{x}^{(i)}, \tilde{y}^{(i)}$ be the uncorrupted data, such that for every i,

$$(\tilde{y}^{(i)} - \langle w^*, \tilde{x}^{(i)} \rangle)^2 \leq \varepsilon^2$$

- Let $\tilde{\alpha}_i = 0$ if i is corrupted, and $\tilde{\alpha}_i = 1$ otherwise (we do not know $\tilde{\alpha}_i = 0$, they are just for proof purpose).
- Then

$$1 = \alpha_i \tilde{\alpha}_i + \alpha_i (1 - \tilde{\alpha}_i) + (1 - \alpha_i)$$

Sum of Squares

- Since

$$1 = \alpha_i \tilde{\alpha}_i + \alpha_i (1 - \tilde{\alpha}_i) + (1 - \alpha_i)$$

- So we have:

$$\sum_{i \in [N]} \left(\langle \tilde{x}^{(i)}, w \rangle - \tilde{y}^{(i)} \right)^2$$

$$= \sum_{i \in [N]} \alpha_i \tilde{\alpha}_i \left(\langle \tilde{x}^{(i)}, w \rangle - \tilde{y}^{(i)} \right)^2$$

$$+ \sum_{i \in [N]} \alpha_i (1 - \tilde{\alpha}_i) \left(\langle \tilde{x}^{(i)}, w \rangle - \tilde{y}^{(i)} \right)^2$$

$$+ \sum_{i \in [N]} (1 - \alpha_i) \left(\langle \tilde{x}^{(i)}, w \rangle - \tilde{y}^{(i)} \right)^2$$

Sum of Squares

- So we have:

$$\sum_{i\in[N]} \left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^2$$

$$= \sum_{i\in[N]} \alpha_i \tilde{\alpha}_i \left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^2 + \sum_{i\in[N]} \alpha_i (1-\tilde{\alpha}_i) \left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^2$$

$$+ \sum_{i\in[N]} (1-\alpha_i) \left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^2$$

- This means: (since $f = g$ implies $\tilde{E}(f) = \tilde{E}(g)$)

$$\sum_{i\in[N]} \tilde{E} \left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^2$$

$$= \sum_{i\in[N]} \tilde{E} \alpha_i \tilde{\alpha}_i \left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^2 + \sum_{i\in[N]} \tilde{E} \alpha_i (1-\tilde{\alpha}_i) \left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^2$$

$$+ \sum \tilde{E}(1-\alpha_i) \left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^2$$

- So we have:
- First term:

$$\sum_{i\in[N]} \tilde{E}\alpha_i\tilde{\alpha}_i\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^2$$

$$=\sum_{i\in[N]} \tilde{E}\alpha_i\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^2$$

$$-\sum_{i\in[N]} \tilde{E}\alpha_i(1-\tilde{\alpha}_i)\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^2$$

- The last term is a (negative) sum of squares, so we know that

$$\sum_{i\in[N]} \tilde{E}\alpha_i\tilde{\alpha}_i\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^2$$

$$\leq\sum_{i\in[N]} \tilde{E}\alpha_i\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^2$$

Which is the objective value of the Sum of Squares program.

- Second term:

$$\sum_{i\in[N]} \tilde{E}\alpha_i(1-\tilde{\alpha}_i)\tilde{\alpha}_i\left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^2$$

- By Cauchy-Shwartz: $\tilde{E}^2(fg) \le \tilde{E}f^2\tilde{E}g^2)$, this is because of $\tilde{E}(f\tilde{E}(fg) - g\tilde{E}f^2)^2 \ge 0$, so $\tilde{E}f^2\tilde{E}^2(fg) - 2\tilde{E}^2(fg)\tilde{E}f^2 + (\tilde{E}f^2)^2\tilde{E}g^2 \ge 0$. So

$$\left(\frac{1}{N}\sum_{i\in[N]} \tilde{E}\alpha_i(1-\tilde{\alpha}_i)\left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^2\right)^2$$

$$\le \left(\frac{1}{N}\sum_{i\in[N]} \tilde{E}(1-\tilde{\alpha}_i)^2\right)\left(\frac{1}{N}\tilde{E}\sum_{i\in[N]} \alpha_i^2\left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^4\right)^2$$

$$\le \tau\left(\frac{1}{N}\sum_{i\in[N]} \tilde{E}\alpha_i^2\left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^4\right) \le \tau\left(\frac{1}{N}\sum_{i\in[N]} \tilde{E}\left(\langle \tilde{x}^{(i)}, w\rangle - \tilde{y}^{(i)}\right)^4\right)$$

- Recall $\tilde{\alpha}_i = 0$ if i is corrupted, and $\tilde{\alpha}_i = 1$ otherwise

Sum of Squares

- $$\tau\left(\frac{1}{N}\sum_{i\in[N]}\tilde{E}\alpha_i^2\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^4\right)\le\tau\left(\frac{1}{N}\sum_{i\in[N]}\tilde{E}\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^4\right)$$
- This is because ($\alpha_i^2=\alpha_i$ is in the Sum of Square constraint)

$$\tilde{E}\alpha_i^2 f^2-\tilde{E}f^2=\tilde{E}(\alpha_i^2-1)f^2$$

$$=\tilde{E}(-\alpha_i^2+2\alpha_i-1)f^2$$

$$=-\tilde{E}(\alpha_i-1)^2f^2\le 0$$

Sum of Squares

- Third term can be done similarly, so we get:

$$\frac{1}{N}\sum_{i\in[N]}\tilde{E}\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^2\leq\frac{1}{N}\text{Sum of Square Objective}$$

$$+O(\tau)^{1/2}\left(\frac{1}{N}\sum_{i\in[N]}\tilde{E}\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^4\right)^{1/2}$$

- If $\tilde{x}^{(i)}$'s are Gaussian, then by hyper-contractivity, we have

$$\left(\frac{1}{N}\sum_{i\in[N]}\tilde{E}\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^4\right)^{1/2}\leq O(1)\frac{1}{N}\sum_{i\in[N]}\tilde{E}\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^2$$

Sum of Squares

- Therefore,

$$\frac{1}{N}\sum_{i\in[N]}\tilde{E}\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^2\le\frac{1}{N}\text{Sum of Square Objective}$$

$$+O(\tau)^{1/2}\frac{1}{N}\sum_{i\in[N]}\tilde{E}\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^2$$

- Which means

$$\frac{1}{N}\sum_{i\in[N]}\tilde{E}\left(\langle\tilde{x}^{(i)},w\rangle-\tilde{y}^{(i)}\right)^2\le\frac{1+O(1/\tau^{1/2})}{N}\text{Sum of Square Objective}$$

- Which means

$$\frac{1}{N}\sum_{i\in[N]}\left(\langle\tilde{x}^{(i)},\tilde{E}[w]\rangle-\tilde{y}^{(i)}\right)^2\le\frac{1+O(1/\tau^{1/2})}{N}\text{Sum of Square Objective}$$

Sum of Squares

- Last inequality we uses:

$$\tilde{E}\left(g(w)\right)^2 - \left(g(\tilde{E}[w])\right)^2$$

$$= \tilde{E}(g(w) - g(\tilde{E}[w]))^2 \geq 0$$

- Which means

$$\frac{1}{N}\sum_{i\in[N]}\left(\langle \tilde{x}^{(i)}, \tilde{E}[w]\rangle - \tilde{y}^{(i)}\right)^2 \leq \frac{1 + O(1/\tau^{1/2})}{N}\text{Sum of Square Objective}$$

$$\leq (1 + O(1/\tau^{1/2}))\varepsilon^2$$