Online Appendix
Estimation Contracts for Outlier-Robust Geometric Perception

Proof 1 — Proof of Proposition 6: A Posteriori Contract for (TLS)

We start by restating the theorem for the reader's convenience.

Proposition 21 (Restatement of Proposition 6)**.** *Consider Problem 1 with measurements* $(\boldsymbol{y}_i, \boldsymbol{A}_i)$*,* $i \in [n]$*, and denote with* γ° *the squared residual error of the ground truth* $\boldsymbol{x}^\circ$ *over the set of inliers* $\mathcal{I}$*,* i.e., $\gamma^\circ \triangleq \sum_{i \in \mathcal{I}} \left\| \boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x}^\circ \right\|_2^2$*. Moreover, assume the measurement set contains at least* $\frac{n+\bar{d}}{2} + \frac{\gamma^\circ}{\bar{c}^2}$ *inliers, where* $\bar{d}$ *is the size of a minimal set, and that every subset of* $\bar{d}$ *inliers is nondegenerate. Then, for any integer* $d_\mathcal{J}$ *such that* $\bar{d} \leq d_\mathcal{J} \leq (2\alpha - 1)n - \frac{\gamma^\circ}{\bar{c}^2}$*, an optimal solution* $\boldsymbol{x}_\mathsf{TLS}$ *of* (TLS) *satisfies*

$$\left\| \boldsymbol{x}_\mathsf{TLS} - \boldsymbol{x}^\circ \right\|_2 \leq \frac{2\sqrt{d_\mathcal{J}}\ \bar{c}}{\min_{\mathcal{J} \subset \mathcal{I}_\mathsf{TLS}, |\mathcal{J}| = d_\mathcal{J}} \sigma_{\min(\boldsymbol{A}_\mathcal{J})}}, \tag{57}$$

where $\mathcal{I}_\mathsf{TLS}$ *is the set of inliers selected by* (TLS)*,* $\boldsymbol{A}_\mathcal{J}$ *is the matrix obtained by horizontally stacking all submatrices* $\boldsymbol{A}_i$ *for all* $i \in \mathcal{J}$*, and* $\sigma_{\min}(\cdot)$ *denotes the smallest singular value of a matrix. Moreover, if the inliers are noiseless,* i.e., $\boldsymbol{\epsilon} = \boldsymbol{0}$ *in eq.* (17)*, and for a sufficiently small* $\bar{c} > 0$*,* $\boldsymbol{x}_\mathsf{TLS} = \boldsymbol{x}^\circ$*.*

Proof. Call $\mathcal{I}(\boldsymbol{x}_\mathsf{TLS})$ the inlier set corresponding to an estimate $\boldsymbol{x}_\mathsf{TLS}$ (*i.e.*, $\mathcal{I}(\boldsymbol{x}_\mathsf{TLS}) \triangleq \{i \in [n] : \left\| \boldsymbol{y}_i - \boldsymbol{A}^\mathsf{T} \boldsymbol{x}_\mathsf{TLS} \right\|_2 \leq \bar{c}\}$). Moreover, define the TLS cost at $\boldsymbol{x}_\mathsf{TLS}$ as:

$$f(\boldsymbol{x}_\mathsf{TLS}) = \sum_{i \in \mathcal{I}(\boldsymbol{x}_\mathsf{TLS})} \left\| \boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x}_\mathsf{TLS} \right\|_2^2 + \bar{c}^2 (n - |\mathcal{I}(\boldsymbol{x}_\mathsf{TLS})|). \tag{58}$$

We first prove that $|\mathcal{I}(\boldsymbol{x}_\mathsf{TLS})| \geq \alpha\, n - \frac{\gamma^\circ}{\bar{c}^2}$. Towards this goal, we observe that the cost evaluated at the ground truth $\boldsymbol{x}^\circ$ is:

$$f(\boldsymbol{x}^\circ) = \gamma^\circ + \bar{c}^2 (n - |\mathcal{I}|) = \gamma^\circ + \bar{c}^2 (n - \alpha\, n) \tag{59}$$

which follows from the assumption that the inliers have squared residual error γ° and there are $\alpha\, n$ of them. Now assume by contradiction that there exists an $\boldsymbol{x}_\mathsf{TLS}$ that solves (TLS) and is such that $|\mathcal{I}(\boldsymbol{x}_\mathsf{TLS})| < \alpha\, n - \frac{\gamma^\circ}{\bar{c}^2}$. Such an estimate would achieve a cost:

$$\begin{aligned} f(\boldsymbol{x}_\mathsf{TLS}) &= \sum_{i \in \mathcal{I}(\boldsymbol{x}_\mathsf{TLS})} \left\| \boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x}_\mathsf{TLS} \right\|_2^2 + \bar{c}^2 (n - |\mathcal{I}(\boldsymbol{x}_\mathsf{TLS})|) && (60) \\ &> \sum_{i \in \mathcal{I}(\boldsymbol{x}_\mathsf{TLS})} \left\| \boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x}_\mathsf{TLS} \right\|_2^2 + \bar{c}^2 \left(n - \alpha\, n + \frac{\gamma^\circ}{\bar{c}^2} \right) && (61) \\ &\geq \bar{c}^2 \left(n - \alpha\, n + \frac{\gamma^\circ}{\bar{c}^2} \right) = \gamma^\circ + \bar{c}^2 \left(n - \alpha\, n \right), && (62) \end{aligned}$$

which is larger than $f(\boldsymbol{x}^\circ)$, hence contradicting optimality of $\boldsymbol{x}_\mathsf{TLS}$, and implying $|\mathcal{I}(\boldsymbol{x}_\mathsf{TLS})| \geq \alpha\, n - \frac{\gamma^\circ}{\bar{c}^2}$.

Since $|\mathcal{I}(\boldsymbol{x}^\circ)| = \alpha\, n$ and $|\mathcal{I}(\boldsymbol{x}_\mathsf{TLS})| \geq \alpha\, n - \frac{\gamma^\circ}{\bar{c}^2}$ then:

$$|\mathcal{I}(\boldsymbol{x}^\circ) \cap \mathcal{I}(\boldsymbol{x}_\mathsf{TLS})| \overbrace{\geq}^{\text{sets overlap in } [n]} 2\alpha\, n - n - \frac{\gamma^\circ}{\bar{c}^2} \overbrace{\geq}^{\text{using } \alpha\, n \geq \frac{n+\bar{d}}{2} + \frac{\gamma^\circ}{\bar{c}^2}} \bar{d}, \tag{63}$$

The subset of measurements $|\mathcal{I}(\boldsymbol{x}^\circ) \cap \mathcal{I}(\boldsymbol{x}_{\mathsf{TLS}})|$ are simultaneously solved by $\boldsymbol{x}_{\mathsf{TLS}}$ and $\boldsymbol{x}^\circ$ (*i.e.*, are such that $\left\|\boldsymbol{y}_i - \boldsymbol{A}^\mathsf{T}\boldsymbol{x}\right\|_2 \leq \bar{c}$ for both $\boldsymbol{x} = \boldsymbol{x}^\circ$ and $\boldsymbol{x} = \boldsymbol{x}_{\mathsf{TLS}}$). Therefore, we can follow the same line of thoughts as in the proof of Proposition 5, and prove the first claim.

In the case of noiseless inliers, $\gamma^\circ = 0$ (or, equivalently, $\boldsymbol{y}_i - \boldsymbol{A}^\mathsf{T}\boldsymbol{x}^\circ = \boldsymbol{0}$, for all $i \in \mathcal{I}$) and we can always choose $\bar{c} > 0$ small enough such that the corresponding estimate $\boldsymbol{x}_{\mathsf{TLS}}$ satisfies the selected measurements exactly, *i.e.*, $\boldsymbol{y}_i - \boldsymbol{A}^\mathsf{T}\boldsymbol{x}_{\mathsf{TLS}} = \boldsymbol{0}$. Therefore, we can follow the same line of the proof of Proposition 5 (for the case of noiseless inliers) to conclude $\boldsymbol{x}_{\mathsf{TLS}} = \boldsymbol{x}^\circ$. ■

Proof 2 — Proof of Theorem 12: Contract for Relaxation of (LTS1)

We start by restating the theorem for the reader's convenience.

Theorem 22 (Restatement of Theorem 12)**.** *Consider Problem 1 with measurements $(\boldsymbol{y}_i, \boldsymbol{A}_i)$, $i \in [n]$, and known outlier rate β. Call $\dddot{\mathcal{I}}$ the set of uncorrupted measurements $(\boldsymbol{y}_i^\star, \boldsymbol{A}_i^\star)$, $i \in [n]$, where the outliers are replaced by inliers and assume that the set of matrices $\boldsymbol{A}_i^\star$, $i \in \dddot{\mathcal{I}}$, is k-certifiably C-hypercontractive with $k \geq 4$. Moreover, assume $\beta < \beta_{\max} = \sqrt[\frac{k}{2}-1]{1/(C(k/2)^{\frac{k}{2}}\, 2^{3k-1})}$. Then, Algorithm 1 with relaxation order $r \geq k$ outputs an estimate $\boldsymbol{x}_{\mathsf{lts-sdp1}}$ (not necessarily in $\mathbb{X}$) such that:*

$$
\mathsf{err}_{\dddot{\mathcal{I}}}(\boldsymbol{x}_{\mathsf{lts-sdp1}}) \leq (1 + C_1(k,\beta)^{\frac{2}{k}})\ \mathsf{opt}_{\dddot{\mathcal{I}}} + C_2(k,\beta)^{\frac{2}{k}} \left(\frac{1}{n} \sum_{i=1}^{n} \left\| \boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T} \boldsymbol{x}^\star \right\|_2^k \right)^{\frac{2}{k}}, \tag{64}
$$

where $C_1(k,\beta)$ and $C_2(k,\beta)$ are given functions, $\mathsf{err}_{\dddot{\mathcal{I}}}(\boldsymbol{x}) \triangleq \frac{1}{n}\sum_{i=1}^n \left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^2$ is the residual error of an estimate $\boldsymbol{x}$ with respect to the inliers $\dddot{\mathcal{I}}$, $\boldsymbol{x}^\star \triangleq \arg\min_{\boldsymbol{x}\in\mathbb{X}} \frac{1}{n}\sum_{i=1}^n \left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^2$ is the best estimate from an oracle estimator that has access to all the inliers, and $\mathsf{opt}_{\dddot{\mathcal{I}}} \triangleq \mathsf{err}_{\dddot{\mathcal{I}}}(\boldsymbol{x}^\star)$ is the corresponding residual error with respect to the inliers $\dddot{\mathcal{I}}$.

The proof is an adaptation of Lemma 5.3 and Lemma 5.6 in [15] to the case of vector-valued measurements. Let us start by clarifying all relevant notation:

$$
\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}} \triangleq \left\{ \begin{array}{c} \omega_i^2 = \omega_i, \quad i \in [n] \\ \sum_{i=1}^n \omega_i = \alpha n \\ \omega_i \cdot (\bar{\boldsymbol{y}}_i - \boldsymbol{y}_i) = \boldsymbol{0} \quad i \in [n] \\ \omega_i \cdot (\bar{\boldsymbol{A}}_i - \boldsymbol{A}_i) = \boldsymbol{0} \quad i \in [n] \\ \boldsymbol{x} \in \mathbb{X} \end{array} \right\} \qquad \text{(constraints in (LTS1))} \tag{65}
$$

$$
\begin{aligned}
\{\boldsymbol{y}_i, \boldsymbol{A}_i\}_{i\in[n]} &\qquad \text{(given measurements)} && (66)\\
\{\boldsymbol{y}_i^\star \boldsymbol{A}_i^\star\}_{i\in[n]}, &\qquad \text{(uncorrupted measurements with outliers replaced by inliers)} && (67)\\
\boldsymbol{V} \triangleq \{\bar{\boldsymbol{y}}_i \bar{\boldsymbol{A}}_i\}_{i\in[n]}, &\qquad \text{(auxiliary variables in (LTS1))} && (68)
\end{aligned}
$$

$$\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) \triangleq \frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^2 \quad \text{(error of } \boldsymbol{x} \text{ w.r.t. uncorrupted measurements)} \tag{69}$$

$$\mathsf{err}(\boldsymbol{x}, \boldsymbol{V}) \triangleq \frac{1}{n}\sum_{i=1}^{n}\left\|\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{A}}_i^\mathsf{T}\boldsymbol{x}\right\|_2^2 \quad \text{(cost in (LTS1) without exponent } k/2\text{)}. \tag{70}$$

$$\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x}, \boldsymbol{V}) \triangleq \frac{1}{n}\sum_{i=1}^{n}\omega_i'\left\|\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{A}}_i^\mathsf{T}\boldsymbol{x}\right\|_2^2 \quad \text{(error w.r.t. set of measurements specified by } \boldsymbol{\omega}'\text{)} \tag{71}$$

Proof. The proof is quite involved and proceeds in two steps. First, we derive an sos proof that states that the inliers picked up by any feasible solution for (LTS1) must also satisfy a desired error bound. Then, we move to pseudo-expectations and conclude that the result of the moment relaxation must satisfy the same bound, which can be manipulated into eq. (64).

Sos proof of robust certifiability (adapted from Lemma 5.6 in [15]). For a given tuple $(\boldsymbol{\omega}, \boldsymbol{x}, \boldsymbol{V})$ that satisfies $\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}$, let $\boldsymbol{\omega}'$ be such that $\omega_i' = \omega_i$ iff i is an inlier and $\omega_i' = 0$ otherwise (intuitively, $\boldsymbol{\omega}'$ is the indicator for the subset of the selected measurements $\boldsymbol{\omega}$ that are inliers).[35] Then call the corresponding error $\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x}, \boldsymbol{V})$ as in (71). We note that:

$$\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x}, \boldsymbol{V}) = \frac{1}{n}\sum_{i=1}^{n}\omega_i'\left\|\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{A}}_i^\mathsf{T}\boldsymbol{x}\right\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}\omega_i'\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^2. \tag{72}$$

The previous equality follows from the fact that (i) by definition, $\omega_i' = 1$ implies $\omega_i = 1$ and since the tuple $(\boldsymbol{\omega}, \boldsymbol{x}, \boldsymbol{V})$ satisfies $\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}$, then whenever $\omega_i = 1$ we must have $\bar{\boldsymbol{A}}_i = \boldsymbol{A}_i$ and $\bar{\boldsymbol{y}}_i = \boldsymbol{y}_i$, and (ii) ω_i' can only be 1 for inliers, for which $\boldsymbol{A}_i = \boldsymbol{A}_i^\star$ and $\boldsymbol{y}_i = \boldsymbol{y}_i^\star$. Therefore, $\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x}, \boldsymbol{V})$ is essentially the error attained by $\boldsymbol{x}$, but restricted to the true inliers in the set of measurements selected by $\boldsymbol{\omega}$.

We now show that any set of variables $(\boldsymbol{\omega}, \boldsymbol{x}, \boldsymbol{V})$ that are feasible for (LTS1) (*i.e.,* that satisfy the constraint set $\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}$), must also satisfy the following bound

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}} \left|\frac{k}{\boldsymbol{x}}\right. (\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x}, \boldsymbol{V}))^{\frac{k}{2}} \leq C_1(k,\beta)(\mathsf{err}(\boldsymbol{x}, \boldsymbol{V}))^{\frac{k}{2}} + C_2(k,\beta)\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k. \tag{73}$$

We start by noting that $\sum_{i=1}^{n}\omega_i' \geq (1-2\beta)n$: this follows from the fact that the two sets, the selected measurements $\{i : \omega_i = 1\}$ and the set of true inliers, have each size $(1-\beta)n$, hence their intersection must contain at least $(1-2\beta)n$ measurements. Therefore:

$$\frac{1}{n}\sum_{i=1}^{n}(1-\omega_i')^2 \overbrace{=}^{\text{binary variable}} \frac{1}{n}\sum_{i=1}^{n}(1-\omega_i') = 1 - \frac{1}{n}\sum_{i=1}^{n}\omega_i' \leq 1-(1-2\beta) \text{ hence:}$$

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}} \overbrace{\left|\frac{2}{\boldsymbol{\omega}'}\right.}^{\textit{Fact A13}} \left\{\frac{1}{n}\sum_{i=1}^{n}(1-\omega_i')^2 \leq 2\beta\right\}. \tag{74}$$

Now we note that from definition (69), we can expand $\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x})$ as:

$$\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}\omega_i'\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^2 + \frac{1}{n}\sum_{i=1}^{n}(1-\omega_i')\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^2 \tag{75}$$

[35]Note that $\boldsymbol{\omega}'$ is not required to (and typically does not) satisfy $\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}$.

Combining (75) and (72), and observing that the result is a sum of squares, we get:

$$\Big|\frac{4}{\boldsymbol{\omega},\boldsymbol{x}}\ \mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V}) = \overbrace{\frac{1}{n}\sum_{i=1}^{n}(1-\omega_i')\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^2}^{\text{error for uncorrupted measurements not selected by }\boldsymbol{\omega}'} \geq 0 \tag{76}$$

Elevating both members of (76) to $k/2$, and then using the sos version of Hölder's inequality (Fact A23), we get:

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}\Big|\frac{k}{\boldsymbol{\omega},\boldsymbol{x},\boldsymbol{V}}\ (\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V}))^{\frac{k}{2}} \overbrace{=}^{\textit{using } (76)} \left(\frac{1}{n}\sum_{i=1}^{n}(1-\omega_i')\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^2\right)^{\frac{k}{2}} \tag{77}$$

$$\leq \overbrace{\left(\frac{1}{n}\sum_{i=1}^{n}(1-\omega_i')^2\right)^{\frac{k}{2}-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^k\right)}^{\text{using (A67) in Fact A23}} \leq \overbrace{(2\beta)^{\frac{k}{2}-1}}^{\text{using (74)}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^k\right). \tag{78}$$

Note that Fact A23 requires the exponent to be ≥ 1 and a power of 2, which in turns implies $\frac{k}{2} \geq 2$ or $k \geq 4$, as required by the statement of the theorem. Now we observe that:

$$\Big|\frac{k}{\boldsymbol{x}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^k\right) \tag{79}$$

$$\overbrace{=}^{\text{adding/subtracting } (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x} + (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k\right) = \tag{80}$$

$$\overbrace{\leq}^{\text{using Fact A20}} 2^k\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k + 2^k\frac{1}{n}\sum_{i=1}^{n}\left\|(\boldsymbol{A}_i^\star)^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\star)\right\|_2^k. \tag{81}$$

By certifiable hypercontractivity of $\boldsymbol{A}_i^\star$, $i \in [n]$:

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}\Big|\frac{k}{\boldsymbol{x}}\ \frac{1}{n}\sum_{i=1}^{n}\left\|(\boldsymbol{A}_i^\star)^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\star)\right\|_2^k \leq C(k/2)^{\frac{k}{2}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|(\boldsymbol{A}_i^\star)^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\star)\right\|_2^2\right)^{\frac{k}{2}}. \tag{82}$$

We can further bound the term above as follows:

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}\Big|\frac{k}{\boldsymbol{x}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|(\boldsymbol{A}_i^\star)^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\star)\right\|_2^2\right)^{\frac{k}{2}} \tag{83}$$

$$\overbrace{=}^{\text{adding/subtracting } \boldsymbol{y}_i^\star}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|-\boldsymbol{y}_i^\star + (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x} + \boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^2\right)^{\frac{k}{2}} \tag{84}$$

$$\overbrace{\leq}^{\text{using Fact A19}}\left(\frac{1}{n}\sum_{i=1}^{n}\left(2\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}\right\|_2^2 + 2\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^2\right)\right)^{\frac{k}{2}} \tag{85}$$

$$\underbrace{=}^{\text{rearranging}} \left(2\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}\right\|_2^2+2\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^2\right)^{\frac{k}{2}} \tag{86}$$

$$\underbrace{\leq}^{\text{using Fact A15}} 2^{\frac{k}{2}}\left(2\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}\right\|_2^2\right)^{\frac{k}{2}}+2^{\frac{k}{2}}\left(2\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^2\right)^{\frac{k}{2}} \tag{87}$$

$$\underbrace{=}^{\text{rearranging}} 2^k\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}\right\|_2^2\right)^{\frac{k}{2}}+2^k\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^2\right)^{\frac{k}{2}}. \tag{88}$$

Finally, using again the sos version of Hölder's inequality

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}\Big|\frac{k}{\boldsymbol{x}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^2\right)^{\frac{k}{2}}\leq\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^k. \tag{89}$$

Combining the above:

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}\Big|\frac{k}{\boldsymbol{x}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}\right\|_2^k\right) \tag{90}$$

$$\underbrace{\leq}^{(81)} 2^k\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^k+2^k\frac{1}{n}\sum_{i=1}^{n}\left\|(\boldsymbol{A}_i^\star)^\top(\boldsymbol{x}-\boldsymbol{x}^\star)\right\|_2^k \tag{91}$$

$$\underbrace{\leq}^{(82)} 2^k\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^k+2^kC(k/2)^{\frac{k}{2}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|(\boldsymbol{A}_i^\star)^\top(\boldsymbol{x}-\boldsymbol{x}^\star)\right\|_2^2\right)^{\frac{k}{2}} \tag{92}$$

$$\underbrace{\leq}^{(88)} 2^k\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^k \tag{93}$$

$$+2^kC(k/2)^{\frac{k}{2}}\left(2^k\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}\right\|_2^2\right)^{\frac{k}{2}}+2^k\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^2\right)^{\frac{k}{2}}\right) \tag{94}$$

$$\underbrace{\leq}^{(89)} 2^k\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^k \tag{95}$$

$$+2^kC(k/2)^{\frac{k}{2}}\left(2^k\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}\right\|_2^2\right)^{\frac{k}{2}}+2^k\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^k\right) \tag{96}$$

$$\underbrace{=}^{\text{rearranging}} C(k/2)^{\frac{k}{2}}\,2^{2k}\Bigg(\overbrace{\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}\right\|_2^2}^{\mathsf{err}_{\mathcal{I}^\star}(\boldsymbol{x})}\Bigg)^{\frac{k}{2}} \tag{97}$$

$$+\left(2^k+C(k/2)^{\frac{k}{2}}\,2^{2k}\right)\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star-(\boldsymbol{A}_i^\star)^\top\boldsymbol{x}^\star\right\|_2^k. \tag{98}$$

Hence, together with (78):

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}} \Big|\frac{k}{\boldsymbol{x}} \left(\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right)^{\frac{k}{2}} \leq (2\beta)^{\frac{k}{2}-1} C(k/2)^{\frac{k}{2}}\, 2^{2k} \mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x})^{\frac{k}{2}} \tag{99}$$

$$+(2\beta)^{\frac{k}{2}-1}\left(2^k + C(k/2)^{\frac{k}{2}}\, 2^{2k}\right)\frac{1}{n}\sum_{i=1}^n \left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k. \tag{100}$$

Applying Fact A17 to the right-hand-side with $a = \mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x})$, $b = \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})$, $\delta^{\frac{k}{2}} = (2\beta)^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}\, 2^{2k}$, and exponent $k/2$:

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}} \Big|\frac{k}{\boldsymbol{x}} \left(\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right)^{\frac{k}{2}} \leq \overbrace{2^{\frac{k}{2}}(2\beta)^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}\, 2^{2k}}^{=\beta^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}2^{3k-1}}\left(\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right)^{\frac{k}{2}} \tag{101}$$

$$+2^{\frac{k}{2}}(2\beta)^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}\, 2^{2k}\left(\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right)^{\frac{k}{2}} \tag{102}$$

$$+(2\beta)^{\frac{k}{2}-1}\left(2^k + C(k/2)^{\frac{k}{2}}\, 2^{2k}\right)\frac{1}{n}\sum_{i=1}^n \left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k. \tag{103}$$

Rearranging the terms:

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}} \Big|\frac{k}{\boldsymbol{x}} (1-\beta^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}\, 2^{3k-1})\left(\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right)^{\frac{k}{2}} \leq \beta^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}\, 2^{3k-1}\left(\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right)^{\frac{k}{2}}$$
$$+(2\beta)^{\frac{k}{2}-1}\left(2^k + C(k/2)^{\frac{k}{2}}\, 2^{2k}\right)\frac{1}{n}\sum_{i=1}^n \left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k. \tag{104}$$

Noting that choosing $\beta < \sqrt[\frac{k}{2}-1]{\frac{1}{C(k/2)^{\frac{k}{2}}\, 2^{3k-1}}}$ makes the constant $1-\beta^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}\, 2^{3k-1}$ positive, we can divive both members of the inequality (104) by such constant and obtain:

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}} \Big|\frac{k}{\boldsymbol{x}} \left(\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right)^{\frac{k}{2}} \leq \frac{\beta^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}\, 2^{3k-1}}{1-\beta^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}\, 2^{3k-1}}\left(\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right)^{\frac{k}{2}} \tag{105}$$

$$+\frac{(2\beta)^{\frac{k}{2}-1}\left(2^k + C(k/2)^{\frac{k}{2}}\, 2^{2k}\right)}{1-\beta^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}\, 2^{3k-1}}\frac{1}{n}\sum_{i=1}^n \left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k, \tag{106}$$

Now defining $C_1(k,\beta) \triangleq \frac{\beta^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}2^{3k-1}}{1-\beta^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}2^{3k-1}}$ and $C_2(k,\beta) \triangleq \frac{(2\beta)^{\frac{k}{2}-1}\left(2^k+C(k/2)^{\frac{k}{2}}2^{2k}\right)}{1-\beta^{\frac{k}{2}-1}C(k/2)^{\frac{k}{2}}2^{3k-1}}$, and noting that $\Big|\frac{k}{\boldsymbol{x}} \mathsf{err}(\boldsymbol{x},\boldsymbol{V}) \geq \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})$,[36] we finally get:

$$\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}} \Big|\frac{k}{\boldsymbol{x}} \left(\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right)^{\frac{k}{2}} \leq C_1(k,\beta)(\mathsf{err}(\boldsymbol{x},\boldsymbol{V}))^{\frac{k}{2}} + C_2(k,\beta)\frac{1}{n}\sum_{i=1}^n \left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k, \tag{107}$$

which matches our claim in (73).

[36]This follows from the definition of the two errors: $\mathsf{err}(\boldsymbol{x},\boldsymbol{V}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V}) = \frac{1}{n}\sum_{i=1}^n \left\|\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{A}}_i^\mathsf{T}\boldsymbol{x}\right\|_2^2 - \frac{1}{n}\sum_{i=1}^n \omega_i' \left\|\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{A}}_i^\mathsf{T}\boldsymbol{x}\right\|_2^2 = \frac{1}{n}\sum_{i=1}^n (1-\omega_i')\left\|\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{A}}_i^\mathsf{T}\boldsymbol{x}\right\|_2^2$, which is a sum of squares.

Completing the proof by moving to pseudo-distributions. Consider a pseudo-distribution $\tilde{\mu}$ that satisfies $\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}$. Using the sos proof in (73) and thanks to Fact A10, we conclude that if $\tilde{\mu}$ satisfies $\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}$ then it must also satisfy:

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[(\mathsf{err}_{\ddot{\mathcal{I}}(\boldsymbol{x})} - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V}))^{\frac{k}{2}}\right] \leq C_1(k,\beta) \overbrace{\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}(\boldsymbol{x},\boldsymbol{V})^{\frac{k}{2}}\right]}^{\text{by definition this is } \widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}}^{\frac{k}{2}}} + C_2(k,\beta)\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k\right). \tag{108}$$

Elevating to the power $\frac{2}{k}$ both sides and recalling that $(a+b)^q \leq a^q + b^q$ for any $0 < q < 1$:

$$\left(\tilde{\mathbb{E}}_{\tilde{\mu}}\left[(\mathsf{err}_{\ddot{\mathcal{I}}(\boldsymbol{x})} - \mathsf{err}(\boldsymbol{x},\boldsymbol{V}))^{\frac{k}{2}}\right]\right)^{\frac{2}{k}} \leq C_1(k,\beta)^{\frac{2}{k}}\ \widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}} + C_2(k,\beta)^{\frac{2}{k}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k\right)^{\frac{2}{k}}. \tag{109}$$

Now using the sos version of Hölder's inequality for pseudo-expectations (Fact A7, eq. (A17)):

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}_{\ddot{\mathcal{I}}(\boldsymbol{x})} - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right]^{\frac{k}{2}} \leq \tilde{\mathbb{E}}_{\tilde{\mu}}\left[(\mathsf{err}_{\ddot{\mathcal{I}}(\boldsymbol{x})} - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V}))^{\frac{k}{2}}\right], \tag{110}$$

and therefore (109) becomes:

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[(\mathsf{err}_{\ddot{\mathcal{I}}(\boldsymbol{x})} - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V}))\right] \leq C_1(k,\beta)^{\frac{2}{k}}\ \widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}} + C_2(k,\beta)^{\frac{2}{k}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k\right)^{\frac{2}{k}}. \tag{111}$$

By linearity of the pseudo-expectation and rearranging:

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}_{\ddot{\mathcal{I}}(\boldsymbol{x})}\right] \leq \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right] + C_1(k,\beta)^{\frac{2}{k}}\ \widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}} + C_2(k,\beta)^{\frac{2}{k}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k\right)^{\frac{2}{k}}. \tag{112}$$

Noting that $\Big|\frac{4}{\boldsymbol{\omega},\boldsymbol{x}}\ \{\mathsf{err}(\boldsymbol{x},\boldsymbol{V}) - \mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V}) \geq 0\}$,[37] and using Fact A10, we get $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right] \leq \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}(\boldsymbol{x},\boldsymbol{V})\right]$. Moreover, we observe

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}(\boldsymbol{x},\boldsymbol{V})\right] = \left(\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}(\boldsymbol{x},\boldsymbol{V})\right]^{\frac{k}{2}}\right)^{\frac{2}{k}} \overbrace{\leq}^{\text{Fact A7, eq. (A17)}} \left(\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}(\boldsymbol{x},\boldsymbol{V})^{\frac{k}{2}}\right]\right)^{\frac{2}{k}} = \widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}} \tag{113}$$

concluding that $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})\right] \leq \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}(\boldsymbol{x},\boldsymbol{V})\right] \leq \widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}}$. Using this inequality in (112):

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\mathsf{err}_{\ddot{\mathcal{I}}(\boldsymbol{x})}\right] \leq (1 + C_1(k,\beta)^{\frac{2}{k}})\ \widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}} + C_2(k,\beta)^{\frac{2}{k}}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k\right)^{\frac{2}{k}} \tag{114}$$

[37] This again follows from the definition of the errors $\mathsf{err}(\boldsymbol{x},\boldsymbol{V})$ and $\mathsf{err}_{\boldsymbol{\omega}'}(\boldsymbol{x},\boldsymbol{V})$, whose difference is a sum of squares.

Applying Hölder's inequality (Fact A7, eq. (A17)) one last time, we get $\mathsf{err}_{\ddot{\mathcal{I}}}(\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{x}]) \leq \tilde{\mathbb{E}}_{\tilde{\mu}}[\mathsf{err}_{\ddot{\mathcal{I}}}(\boldsymbol{x})]$, which leads to:

$$\mathsf{err}_{\ddot{\mathcal{I}}}(\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{x}]) \leq (1 + C_1(k,\beta)^{\frac{2}{k}})\, \widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}} + C_2(k,\beta)^{\frac{2}{k}} \left(\frac{1}{n}\sum_{i=1}^{n} \left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k\right)^{\frac{2}{k}}. \tag{115}$$

Finally, we need to prove that $\widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}} \leq \mathsf{opt}_{\ddot{\mathcal{I}}}$. Towards this goal, we observe that the (pseudo-)distribution supported on the point $(\boldsymbol{\omega}^\star, \boldsymbol{x}^\star, \boldsymbol{V})$ where $\omega_i^\star = 1$ for the true inliers and zero otherwise is feasible for $\mathcal{L}_{\boldsymbol{\omega},\boldsymbol{x}}$, hence by optimality $\widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}}^{\frac{k}{2}} \leq \left(\frac{1}{n}\sum_{i=1}^{n}\left\|\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{A}}_i^\mathsf{T}\boldsymbol{x}^\star\right\|_2^2\right)^{\frac{k}{2}}$, from which it follows:

$$\widehat{\mathsf{opt}}_{\mathsf{lts-sdp1}} \leq \frac{1}{n}\sum_{i=1}^{n}\left\|\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{A}}_i^\mathsf{T}\boldsymbol{x}^\star\right\|_2^2 \overbrace{\leq}^{\bar{\boldsymbol{y}}_i, \bar{\boldsymbol{A}}_i \text{ are inliers or zero}} \frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2 = \mathsf{opt}_{\ddot{\mathcal{I}}}. \tag{116}$$

Substituting (116) back into (115):

$$\mathsf{err}_{\ddot{\mathcal{I}}}(\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{x}]) \leq (1 + C_1(k,\beta)^{\frac{2}{k}})\, \mathsf{opt}_{\ddot{\mathcal{I}}} + C_2(k,\beta)^{\frac{2}{k}} \left(\frac{1}{n}\sum_{i=1}^{n} \left\|\boldsymbol{y}_i^\star - (\boldsymbol{A}_i^\star)^\mathsf{T}\boldsymbol{x}^\star\right\|_2^k\right)^{\frac{2}{k}}, \tag{117}$$

which proves the claim of Theorem 12.

■

Proof 3 — Proof of Proposition 13: Contract for Relaxation of (LTS2)

We start by restating the proposition for the reader's convenience.

Proposition 23 (Restatement of Proposition 13)**.** *Consider Problem 1 with measurements $(\boldsymbol{y}_i, \boldsymbol{A}_i)$, $i \in [n]$, and outlier rate $\beta < 0.5$ (or, equivalently, inlier rate $\alpha = 1-\beta > 0.5$). Call $\mathcal{I}$ the set of inliers and assume that the set of matrices $\boldsymbol{A}_i$, $i \in \mathcal{I}$, is k-certifiably $(\frac{\alpha^2\eta^2(1-2\bar{c})^2}{32\bar{c}}, 2\bar{c}, 2M_x)$-anti-concentrated for some $\eta > 0$. Then, Algorithm 2 with relaxation order $r \geq k/2$ outputs an estimate $\boldsymbol{x}_{\mathsf{lts-sdp2}}$ (not necessarily in $\mathbb{X}$) such that:*

$$\|\boldsymbol{x}_{\mathsf{lts-sdp2}} - \boldsymbol{x}^\circ\|_2 \leq M_x\left(\frac{\alpha\,\eta}{2} + 2\frac{1-\alpha}{\alpha}\right). \tag{118}$$

Towards proving the proposition we need to prove two technical lemmas (Lemmas 24 and 25 below). These lemmas extend results [37] to vector-valued and noisy measurements, and while in [37] they have been proposed to attack the high-outlier case (*i.e.,* for list-decodable regression), we show they are also useful to prove estimation contracts for the low-outlier case.

Note that the two lemmas below use a subset of constraints compared to the one in the constraint set of (LTS2) (*i.e.,* the set $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$ below does not contain the constraint $\sum_{i=1}^{n}\omega_i = \alpha n$): this will allow us to use them also to discuss the performance of (MC) and (TLS) later on.

Lemma 24 (Adapted from Lemma 4.1 in [37])**.** *Consider the following constraint set, for given measurements $(\boldsymbol{y}_i, \boldsymbol{A}_i)$, $i \in [n]$, a constant $\bar{c} \geq 0$, and where $\mathbb{X}$ is an explicitly bounded basic* *semi-algebraic set* (cf. *Assumption 2*):

$$\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}} \doteq \left\{ \begin{array}{c} \omega_i^2 = \omega_i, \quad i \in [n] \\ \omega_i \cdot \left\| \boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x} \right\|_2^2 \leq \bar{c}^2 \quad i \in [n] \\ \boldsymbol{x} \in \mathbb{X} \end{array} \right\}. \tag{119}$$

For any $t \geq k$ and set of n measurements with at least $\alpha\, n$ inliers, such that for the set of inliers $\mathcal{I}$, the set of matrices $\boldsymbol{A}_i$, $i \in \mathcal{I}$, is k-certifiably $(\frac{\alpha^2\eta^2(1-2\bar{c})^2}{32\bar{c}}, 2\bar{c}, 2M_x)$-anti-concentrated,[38]

$$\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}} \left|\frac{t}{\boldsymbol{\omega},\boldsymbol{x}}\right. \left\{ \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \omega_i \left\| \boldsymbol{x} - \boldsymbol{x}^\circ \right\|_2^2 \leq \frac{\alpha^2 \eta^2 M_x^2}{4} \right\}. \tag{120}$$

Proof. We follow the same logic as the proof of Lemma 4.1 in [37], but provide a slightly simpler derivation, based on our definition of certifiable anti-concentration. We first observe that for the inliers (*i.e.*, $i \in \mathcal{I}$) it holds:[39]

$$\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}} \left|\frac{t}{\boldsymbol{x}}\right. \omega_i \cdot \left\| \boldsymbol{A}_i^\mathsf{T} (\boldsymbol{x} - \boldsymbol{x}^\circ) \right\|_2^2 \overbrace{=}^{\text{adding/subtracting } \boldsymbol{y}_i} \omega_i \cdot \left\| (\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x}^\circ) - (\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x}) \right\|_2^2 \tag{121}$$

$$\overbrace{\leq}^{\textit{Fact A19}} \omega_i \cdot \left(2 \left\| \boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x}^\circ \right\|_2^2 + 2 \left\| \boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x} \right\|_2^2 \right) \tag{122}$$

$$\overbrace{\leq}^{\omega_i \leq 1 \text{ and Fact A13}} 2 \left\| \boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x}^\circ \right\|_2^2 + 2 \cdot \omega_i \cdot \left\| \boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x} \right\|_2^2 \tag{123}$$

(since $(\boldsymbol{\omega}, \boldsymbol{x})$ satisfy $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$, and inliers by definition satisfy $\left\| \boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T} \boldsymbol{x}^\circ \right\|_2^2 \leq \bar{c}^2$)

$$\leq 4\bar{c}^2. \tag{124}$$

Since the set of matrices $\boldsymbol{A}_i$, $i \in \mathcal{I}$, is k-certifiably $(\frac{\alpha^2\eta^2(1-2\bar{c})^2}{32\bar{c}}, 2\bar{c}, 2M_x)$-anti-concentrated, then there exists a univariate polynomial p such that for every $i \in \mathcal{I}$ and for every $t \geq k$:

$$\overbrace{\left\{ \omega_i \cdot \left\| \boldsymbol{A}_i^\mathsf{T} (\boldsymbol{x} - \boldsymbol{x}^\circ) \right\|_2^2 \leq 4\bar{c}^2 \right\}}^{\text{from (124)}} \overbrace{\left|\frac{t}{\boldsymbol{x}}\right. p^2 \left(\omega_i \cdot \left\| \boldsymbol{A}_i^\mathsf{T} (\boldsymbol{x} - \boldsymbol{x}^\circ) \right\|_2 \right) \geq (1 - 2\bar{c})^2}^{\text{using (32) with } \delta = 2\bar{c}} \tag{125}$$

and

$$\|\boldsymbol{x}\|_2^2 \leq M_x \overbrace{\left|\frac{t}{\boldsymbol{x}}\right. \|\boldsymbol{x} - \boldsymbol{x}^\circ\|_2^2 \leq 4M_x^2}^{\textit{Fact A19}} \tag{126}$$

$$\overbrace{\left|\frac{t}{\boldsymbol{x}}\right. \left\{ \|\boldsymbol{x} - \boldsymbol{x}^\circ\|_2^2 \cdot \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} p^2 \left(\left\| \boldsymbol{A}_i^\mathsf{T} (\boldsymbol{x} - \boldsymbol{x}^\circ) \right\|_2 \right) \leq \frac{\alpha^2 \eta^2 (1 - 2\bar{c})^2 M_x^2}{4} \right\}}^{\text{using certifiable anti-concentration in eq. (33)}}. \tag{127}$$

[38]The constant "32″ in the anti-concentration requirement is arbitrary (*i.e.*, it just amounts to a re-scaling of the parameter η) and has been chosen to keep the result consistent with the original statement in [37].

[39]Observe the analogy with the proof of Proposition 5.

Now we note that:

$$\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}} \left|\frac{t}{\boldsymbol{\omega},\boldsymbol{x}}\right. \omega_i \cdot p^2\left(\omega_i \cdot \left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right) \overbrace{=}^{\omega_i^2=\omega_i} \left(\omega_i \cdot p\left(\omega_i \cdot \left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right)\right)^2 \tag{128}$$

$$= \overbrace{\left(\omega_i \cdot \left(1 + h\left(\omega_i \cdot \left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right)\right)\right)^2}^{\text{calling } h \text{ the homogeneous part of } p \text{ and since } p(0)=1} = \left(\omega_i + \omega_i \cdot h\left(\omega_i \cdot \left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right)\right)^2 \tag{129}$$

$$\overbrace{=}^{\omega_i^2=\omega_i} \left(\omega_i + \omega_i \cdot h\left(\left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right)\right)^2 = \omega_i^2 \cdot \left(1 + h\left(\left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right)\right)^2 \tag{130}$$

$$= \omega_i^2 \cdot p^2\left(\left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right) \overbrace{\leq}^{\omega_i \leq 1} p^2\left(\left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right). \tag{131}$$

Combining the conclusions in (125), (127), and (131) we obtain:

$$\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}} \left|\frac{t}{\boldsymbol{\omega},\boldsymbol{x}}\right. \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \omega_i \left\|\boldsymbol{x}-\boldsymbol{x}^\circ\right\|_2^2 \overbrace{\leq \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \omega_i \left\|\boldsymbol{x}-\boldsymbol{x}^\circ\right\|_2^2 \cdot \frac{1}{(1-2\bar{c})^2} p^2\left(\omega_i \cdot \left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right)}^{\text{since } \frac{1}{(1-2\bar{c})^2} p^2\left(\omega_i \cdot \left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right) \geq 1 \text{ per } (125)} \tag{132}$$

$$\overbrace{\leq \frac{1}{(1-2\bar{c})^2}\left\|\boldsymbol{x}-\boldsymbol{x}^\circ\right\|_2^2 \cdot \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} p^2\left(\left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right)}^{\text{using } \omega_i \cdot p^2\left(\omega_i \cdot \left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right) \leq p^2\left(\left\|\boldsymbol{A}_i^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^\circ)\right\|_2\right) \text{ from } (131)} \overbrace{\leq \frac{\alpha^2\eta^2 M_x^2}{4}}^{\text{using } (127)}, \tag{133}$$

which concludes the proof. ∎

Lemma 25 (Adapted from Lemma 4.2 in [37])**.** *Under the same assumptions of Lemma 24, for any pseudo-distribution $\tilde{\mu}$ of level at least k satisfying $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$,*

$$\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\left\|\boldsymbol{v}_i-\boldsymbol{x}^\circ\right\|_2 \leq \frac{\alpha\eta M_x}{2}, \tag{134}$$

where the vectors $\boldsymbol{v}_i$ are extracted from the pseudo-moment matrix by setting $\boldsymbol{v}_i = \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i \boldsymbol{x}]}{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}$ if $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] > 0$, or $\boldsymbol{v}_i = \mathbf{0}$ otherwise, for $i \in [n]$.

Proof. By Lemma 24, we have $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}} \left|\frac{k}{\boldsymbol{\omega},\boldsymbol{x}}\right. \left\{\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \omega_i \left\|\boldsymbol{x}-\boldsymbol{x}^\circ\right\|_2^2 \leq \frac{\alpha^2\eta^2 M_x^2}{4}\right\}$. We also have: $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}} \left|\frac{2}{\boldsymbol{\omega}}\right. \left\{\omega_i^2 = \omega_i\right\}$ for any i. Therefore:

$$\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}} \left|\frac{k}{\boldsymbol{\omega},\boldsymbol{x}}\right. \left\{\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \left\|\omega_i\boldsymbol{x}-\omega_i\boldsymbol{x}^\circ\right\|_2^2 \leq \frac{\alpha^2\eta^2 M_x^2}{4}\right\}. \tag{135}$$

Since $\tilde{\mu}$ satisfies $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$, then it also satisfies:

$$\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\left\|\omega_i\boldsymbol{x}-\omega_i\boldsymbol{x}^\circ\right\|_2^2\right] \leq \frac{\alpha^2\eta^2 M_x^2}{4}. \tag{136}$$

Using the norm inequality for pseudo-distributions in Fact A8, we get $\left\|\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i \boldsymbol{x} - \omega_i \boldsymbol{x}^\circ\right]\right\|_2^2 \leq \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\|\omega_i \boldsymbol{x} - \omega_i \boldsymbol{x}^\circ\|_2^2\right]$; then observing that for any m-vector $\boldsymbol{z}$, $\|\boldsymbol{z}\|_1 \leq \sqrt{m}\,\|\boldsymbol{z}\|_2$ or, equivalently, $\|\boldsymbol{z}\|_1^2 \leq m\,\|\boldsymbol{z}\|_2^2$ (below we will apply this inequality to the vector of size $|\mathcal{I}|$ with entries $z_i = \frac{1}{|\mathcal{I}|}\left\|\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i \boldsymbol{x}\right] - \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \boldsymbol{x}^\circ\right\|_2$), and chaining the inequalities back to (136):

$$\overbrace{\left(\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\left\|\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i \boldsymbol{x}\right] - \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \boldsymbol{x}^\circ\right\|_2\right)^2}^{\|\boldsymbol{z}\|_1^2} \overbrace{\leq}^{\|\boldsymbol{z}\|_1^2 \leq m\|\boldsymbol{z}\|_2^2} \overbrace{\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\left\|\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i \boldsymbol{x}\right] - \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \boldsymbol{x}^\circ\right\|_2^2}^{m\|\boldsymbol{z}\|_2^2 \text{ with } m=|\mathcal{I}|} \tag{137}$$

$$\overbrace{\leq}^{\textit{Fact A8}} \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\|\omega_i \boldsymbol{x} - \omega_i \boldsymbol{x}^\circ\|_2^2\right] \overbrace{\leq}^{(136)} \frac{\alpha^2\eta^2 M_x^2}{4}. \tag{138}$$

Remembering that $\boldsymbol{v}_i = \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i \boldsymbol{x}]}{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}$ if $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] > 0$, or $\boldsymbol{v}_i = \mathbf{0}$ otherwise, and taking the square root of both members in (138):

$$\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I},\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]>0}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \overbrace{=}^{\text{by def. of } \boldsymbol{v}_i} \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\left\|\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i \boldsymbol{x}\right] - \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \boldsymbol{x}^\circ\right\|_2 \overbrace{\leq}^{(138)} \frac{\alpha\eta M_x}{2}, \tag{139}$$

concluding the proof of Lemma 25. ■

Proof of Proposition 13: First of all, we note that since $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$ in Algorithm 2 contains a superset of the constraints in $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$ defined in Lemma 24, the conclusions of Lemma 25 and Lemma 24 still hold if we replace $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$ with $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$. Therefore we have that any pseudo-distribution of level at least k (hence produced by a relaxation of order at least $k/2$) satisfying $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$ also satisfies:

$$\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \leq \frac{\alpha\eta M_x}{2} \overbrace{\iff}^{|\mathcal{I}|=\alpha\, n} \sum_{i\in\mathcal{I}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \leq \frac{\alpha^2\,\eta\, M_x\, n}{2}. \tag{140}$$

Let us define the set of outliers $\mathcal{O} \triangleq [n]\backslash\mathcal{I}$. We observe that since $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \leq 1$, then $\sum_{i\in\mathcal{O}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \leq (1-\alpha)n$. Moreover, using the triangle inequality $\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \leq 2M_x$, hence:

$$\sum_{i\in\mathcal{O}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \leq 2nM_x(1-\alpha). \tag{141}$$

Using Eq. (140) and Eq. (141):

$$\sum_{i=1}^{n}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 = \sum_{i\in\mathcal{I}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 + \sum_{i\in\mathcal{O}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \tag{142}$$

$$\leq \frac{\alpha^2\,\eta\,M_x\,n}{2} + 2nM_x(1-\alpha). \tag{143}$$

Now note that any pseudo-distribution $\tilde{\mu}$ satisfying $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$ is such that $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\sum_{i=1}^{n}\omega_i\right]=\alpha\, n$ (due to the constraint $\sum_{i=1}^{n}\omega_i=\alpha\, n$ in $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$), hence by linearity $\sum_{i=1}^{n}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]=\alpha\, n$. Dividing both members of (143) by $\sum_{j=1}^{n}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_j\right]$ (where we switched to using j as an index to avoid confusion):

$$\sum_{i=1}^{n}\frac{\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]}{\sum_{j=1}^{n}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_j\right]}\|\boldsymbol{v}_i-\boldsymbol{x}^{\circ}\|_2\leq\frac{1}{\sum_{j=1}^{n}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_j\right]}\left(\frac{\alpha^2\,\eta\,M_x\,n}{2}+2nM_x(1-\alpha)\right)\tag{144}$$

$$=\frac{1}{\alpha\, n}\left(\frac{\alpha^2\,\eta\,M_x\,n}{2}+2nM_x(1-\alpha)\right).\tag{145}$$

Using Jensen's inequality, we observe $\left\|\sum_{i=1}^{n}\frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n}\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]}\boldsymbol{v}_i-\boldsymbol{x}^{\circ}\right\|_2\leq\sum_{i=1}^{n}\frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n}\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]}\|\boldsymbol{v}_i-\boldsymbol{x}^{\circ}\|_2$, hence (165) becomes:

$$\left\|\sum_{i=1}^{n}\frac{\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]}{\sum_{j=1}^{n}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_j\right]}\boldsymbol{v}_i-\boldsymbol{x}^{\circ}\right\|_2\leq\frac{\alpha\,\eta\,M_x}{2}+2M_x\frac{1-\alpha}{\alpha},\tag{146}$$

which, recalling that $\boldsymbol{x}_{\mathsf{lts-sdp2}}=\sum_{i=1}^{n}\frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n}\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]}\boldsymbol{v}_i$, concludes the proof.

■

Proof 4 — Proof of Proposition 14: Contract for Relaxation of (MC1)

We start by restating the proposition for the reader's convenience.

Proposition 26 (Restatement of Proposition 14)**.** *Consider Problem 1 with measurements $(\boldsymbol{y}_i,\boldsymbol{A}_i)$, $i\in[n]$, and outlier rate $\beta<0.5$ (or, equivalently, inlier rate $\alpha=1-\beta>0.5$). Call $\mathcal{I}$ the set of inliers and assume that the set of matrices $\boldsymbol{A}_i$, $i\in\mathcal{I}$, is k-certifiably $\left(\frac{\alpha^2\eta^2(1-2\bar{c})^2}{32\bar{c}},2\bar{c},2M_x\right)$-anti-concentrated for some $\eta>0$. Then, Algorithm 3 with relaxation order $r\geq k/2$ outputs an estimate $\boldsymbol{x}_{\mathsf{mc-sdp}}$ (not necessarily in $\mathbb{X}$) such that:*

$$\|\boldsymbol{x}_{\mathsf{mc-sdp}}-\boldsymbol{x}^{\circ}\|_2\leq M_x\left(\frac{\alpha\,\eta}{2}+2\frac{1-\alpha}{\alpha}\right).\tag{147}$$

Proof. First of all, we note that the constraint set $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$ in (MC1) is the same as Lemma 24 and Lemma 25. Therefore we have that any pseudo-distribution $\tilde{\mu}$ of level at least k (hence produced by a relaxation of order at least $k/2$) satisfying $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$ also satisfies:

$$\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\|\boldsymbol{v}_i-\boldsymbol{x}^{\circ}\|_2\leq\frac{\alpha\eta M_x}{2}\overset{|\mathcal{I}|=\alpha\, n}{\iff}\sum_{i\in\mathcal{I}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\|\boldsymbol{v}_i-\boldsymbol{x}^{\circ}\|_2\leq\frac{\alpha^2\,\eta\,M_x\,n}{2}.\tag{148}$$

Let us define the set of outliers $\mathcal{O}\triangleq[n]\backslash\mathcal{I}$. We observe that since $\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\leq 1$, then $\sum_{i\in\mathcal{O}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\leq(1-\alpha)n$. Moreover, using the triangle inequality $\|\boldsymbol{v}_i-\boldsymbol{x}^{\circ}\|_2\leq 2M_x$, hence:

$$\sum_{i\in\mathcal{O}}\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]\|\boldsymbol{v}_i-\boldsymbol{x}^{\circ}\|_2\leq 2nM_x(1-\alpha).\tag{149}$$

Using Eq. (148) and Eq. (149):

$$\sum_{i=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 = \sum_{i\in\mathcal{I}} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 + \sum_{i\in\mathcal{O}} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \tag{150}$$

$$\leq \frac{\alpha^2\, \eta\, M_x\, n}{2} + 2nM_x(1-\alpha). \tag{151}$$

Let us call $\tilde{\mu}$ the pseudo-distribution that achieves the optimal solution in (MC1), and observe that the corresponding optimal objective $\sum_{i=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \geq \alpha\, n$: this follows from optimality of $\tilde{\mu}$ and from the fact that the pseudo-distribution supported on the single point $(\boldsymbol{x}^\circ, \boldsymbol{\omega}^\circ)$, where $\omega_i^\circ = 1$ if $i \in \mathcal{I}$ or zero otherwise, is feasible for (MC1) and achieves an objective $\alpha\, n$.

Now dividing both members of (159) by $\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_j\right]$:

$$\sum_{i=1}^{n} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_j\right]} \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \leq \frac{1}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_j\right]} \left(\frac{\alpha^2\, \eta\, M_x\, n}{2} + 2nM_x(1-\alpha)\right) \tag{152}$$

$$\leq \frac{1}{\alpha\, n} \left(\frac{\alpha^2\, \eta\, M_x\, n}{2} + 2nM_x(1-\alpha)\right). \tag{153}$$

Using Jensen's inequality $\left\| \sum_{i=1}^{n} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]} \boldsymbol{v}_i - \boldsymbol{x}^\circ \right\|_2 \leq \sum_{i=1}^{n} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]} \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2$ hence (153) becomes:

$$\left\| \sum_{i=1}^{n} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_j\right]} \boldsymbol{v}_i - \boldsymbol{x}^\circ \right\|_2 \leq \frac{\alpha\, \eta\, M_x}{2} + 2M_x \frac{1-\alpha}{\alpha}, \tag{154}$$

which, recalling that $\boldsymbol{x}_{\mathsf{mc-sdp}} = \sum_{i=1}^{n} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]} \boldsymbol{v}_i$, concludes the proof. ■

Proof 5 — Proof of Proposition 15: Contract for Relaxation of (TLS1)

We start by restating the proposition for the reader's convenience.

Proposition 27 (Restatement of Proposition 15)**.** *Consider Problem 1 with measurements $(\boldsymbol{y}_i, \boldsymbol{A}_i)$, $i \in [n]$, and outlier rate $\beta < 0.5$ (or, equivalently, inlier rate $\alpha = 1-\beta > 0.5$). Call $\mathcal{I}$ the set of inliers and assume that the set of matrices $\boldsymbol{A}_i$, $i \in \mathcal{I}$, is k-certifiably $(\frac{\alpha^2\eta^2(1-2\bar{c})^2}{32\bar{c}}, 2\bar{c}, 2M_x)$-anti-concentrated for some $\eta > 0$. Then, Algorithm 4 with relaxation order $r \geq k/2$ outputs an estimate $\boldsymbol{x}_{\mathsf{tls-sdp}}$ (not necessarily in $\mathbb{X}$) such that:*

$$\|\boldsymbol{x}_{\mathsf{tls-sdp}} - \boldsymbol{x}^\circ\|_2 \leq \frac{\alpha M_x n}{\alpha\, n - \frac{\gamma^\circ}{\bar{c}^2}} \left(\frac{\alpha\, \eta}{2} + 2\frac{1-\alpha}{\alpha}\right), \tag{155}$$

where $\gamma^\circ \triangleq \sum_{i\in\mathcal{I}} \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}^\circ\right\|_2^2$ is the squared residual error of the ground truth $\boldsymbol{x}^\circ$ over the inliers $\mathcal{I}$.

Proof. First of all, we note that the constraint set $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$ in (TLS1) is the same as Lemma 24 and Lemma 25. Therefore we have that any pseudo-distribution $\tilde{\mu}$ of level at least k (hence produced by a relaxation of order at least $k/2$) satisfying $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$ also satisfies:

$$\frac{1}{|\mathcal{I}|} \sum_{i\in\mathcal{I}} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \leq \frac{\alpha\eta M_x}{2} \overset{|\mathcal{I}|=\alpha\, n}{\iff} \sum_{i\in\mathcal{I}} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \leq \frac{\alpha^2\, \eta\, M_x\, n}{2}. \tag{156}$$

Let us define the set of outliers $\mathcal{O} \triangleq [n] \backslash \mathcal{I}$. We observe that since $\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i] \leq 1$, then $\sum_{i \in \mathcal{O}} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i] \leq (1-\alpha)n$. Moreover, using the triangle inequality $\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \leq 2M_x$, hence:

$$\sum_{i \in \mathcal{O}} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i] \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \leq 2nM_x(1-\alpha). \tag{157}$$

Using Eq. (156) and Eq. (157):

$$\begin{aligned} \sum_{i=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i] \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 &= \sum_{i \in \mathcal{I}} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i] \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 + \sum_{i \in \mathcal{O}} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i] \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 && (158) \\ &\leq \frac{\alpha^2\, \eta\, M_x\, n}{2} + 2nM_x(1-\alpha). && (159) \end{aligned}$$

Let us call $\tilde{\mu}$ the pseudo-distribution that achieves the optimal solution in (TLS1), and observe that $\tilde{\mu}$ achieves a cost:

$$\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\sum_{i=1}^{n} \omega_i \cdot \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}\right\|_2^2 + (1-\omega_i)\cdot \bar{c}^2\right] = \sum_{i=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i \cdot \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}\right\|_2^2\right] + \sum_{i=1}^{n} (1 - \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]) \cdot \bar{c}^2. \tag{160}$$

Now observe that the pseudo-distribution supported on the single point $(\boldsymbol{x}^\circ, \boldsymbol{\omega}^\circ)$, where $\omega_i^\circ = 1$ if $i \in \mathcal{I}$ or zero otherwise, is feasible for (TLS1) and achieves an objective $\sum_{i \in \mathcal{I}} \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}^\circ\right\|_2^2 + (1-\alpha)n\bar{c}^2$. Therefore, by using (160) and by optimality of $\tilde{\mu}$:

$$\sum_{i=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i \cdot \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}\right\|_2^2\right] + \overbrace{\sum_{i=1}^{n} (1 - \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]) \cdot \bar{c}^2}^{= n\bar{c}^2 - \sum_{i=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]\bar{c}^2} \leq \sum_{i \in \mathcal{I}} \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}^\circ\right\|_2^2 + \overbrace{(1-\alpha)n\bar{c}^2}^{= n\bar{c}^2 - \alpha\, n\, \bar{c}^2}. \tag{161}$$

Rearranging the terms in the previous inequality:

$$\begin{aligned} \sum_{i=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i] &\geq \frac{1}{\bar{c}^2}\left(\sum_{i=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i \cdot \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}\right\|_2^2\right] - \sum_{i \in \mathcal{I}} \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}^\circ\right\|_2^2 + \alpha\, n\, \bar{c}^2\right) && (162) \\ &\geq \frac{1}{\bar{c}^2}\left(\alpha\, n\, \bar{c}^2 - \sum_{i \in \mathcal{I}} \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}^\circ\right\|_2^2\right) = \alpha\, n - \frac{1}{\bar{c}^2}\sum_{i \in \mathcal{I}} \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}^\circ\right\|_2^2. && (163) \end{aligned}$$

Now dividing both members of (159) by $\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]$ and defining $\gamma^\circ \triangleq \sum_{i \in \mathcal{I}} \left\|\boldsymbol{y}_i - \boldsymbol{A}_i^\mathsf{T}\boldsymbol{x}^\circ\right\|_2^2$:

$$\begin{aligned} \sum_{i=1}^{n} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]} \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 &\leq \frac{1}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]}\left(\frac{\alpha^2\, \eta\, M_x\, n}{2} + 2nM_x(1-\alpha)\right) && (164) \\ &\overbrace{\leq}^{\text{using (163)}} \frac{1}{\alpha\, n - \frac{\gamma^\circ}{\bar{c}^2}}\left(\frac{\alpha^2\, \eta\, M_x\, n}{2} + 2nM_x(1-\alpha)\right). && (165) \end{aligned}$$

Using Jensen's inequality $\left\| \sum_{i=1}^{n} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]} \boldsymbol{v}_i - \boldsymbol{x}^\circ \right\|_2 \leq \sum_{i=1}^{n} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]} \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2$ hence (165) becomes:

$$\left\| \sum_{i=1}^{n} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]} \boldsymbol{v}_i - \boldsymbol{x}^\circ \right\|_2 \leq \frac{1}{\alpha\, n - \frac{\gamma^\circ}{\bar{c}^2}} \left(\frac{\alpha^2\, \eta\, M_x\, n}{2} + 2nM_x(1-\alpha) \right), \tag{166}$$

which, recalling that $\boldsymbol{x}_{\mathsf{tls-sdp}} = \sum_{i=1}^{n} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\sum_{j=1}^{n} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_j]} \boldsymbol{v}_i$, concludes the proof. ■

Proof 6 — Proof of Theorem 17: Contract for Relaxation of (LDR)

We start by restating the theorem for the reader's convenience.

Theorem 28 (Restatement of Proposition 17)**.** *Consider Problem 1 with measurements $(\boldsymbol{y}_i, \boldsymbol{A}_i)$, $i \in [n]$, and known outlier rate β (or, equivalently, known inlier rate $\alpha = 1 - \beta$), possibly with $\beta > 0.5$. Call $\mathcal{I}$ the set of inliers and assume that the set of matrices $\boldsymbol{A}_i$, $i \in \mathcal{I}$, is k-certifiably $(\frac{\alpha^2\eta^2(1-2\bar{c})^2}{32\bar{c}}, 2\bar{c}, 2M_x)$-anti-concentrated for some $\eta > 0$. Then, with probability at least $1 - \left(1 - \frac{\alpha}{2}\right)^{\frac{N}{\alpha}}$ (over the draw of the samples in the algorithms), where $N \geq 1$ is a user-defined parameter, Algorithm 5 with relaxation order $r \geq k/2$ outputs a list $\mathcal{L}$ of size N/α such that there is an estimate $\boldsymbol{x} \in \mathcal{L}$ (with $\boldsymbol{x}$ not necessarily in $\mathbb{X}$) such that*

$$\|\boldsymbol{x} - \boldsymbol{x}^\circ\|_2 \leq \eta M_x. \tag{167}$$

*Moreover, when $\alpha \geq 0.01$ (*i.e.*, at least 1% of the measurements are inliers) and $N = 10$, the relation $\|\boldsymbol{x} - \boldsymbol{x}^\circ\|_2 \leq \eta M_x$ holds with probability at least 0.99 over the draw of the samples.*

Proof. Note that Lemma 24 and Lemma 25 use a subset of constraints compared to the set $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$ in (LDR) (*i.e.*, the set $\mathcal{M}_{\boldsymbol{\omega},\boldsymbol{x}}$ in the lemmas does not contain the constraint $\sum_{i=1}^{n} \omega_i = \alpha n$, while $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$ does). Therefore, their conclusions will still hold in the context of (LDR). We start by proving the following lemma, which shows that the pseudo-distribution $\tilde{\mu}$ built by optimizing the moment relaxation of (LDR) "spreads" (*i.e.*, has enough support) across the inliers. The proof is an extension of Lemma 4.3 in [37] to the case of vector-valued measurements.

Lemma 29 (Adapted from Lemma 4.3 in [37])**.** *For any pseudo-distribution $\tilde{\mu}$ satisfying $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$ that minimizes $\left\|\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{\omega}]\right\|_2^2$, $\sum_{i\in\mathcal{I}} \tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i] \geq \alpha^2 n$.*

Proof. Let $\boldsymbol{u} = \frac{1}{\alpha n}\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{\omega}]$. Then, $\boldsymbol{u}$ is a non-negative vector satisfying $\sum_{i=1}^{n} u_i = 1$. Let $\mathrm{wt}(\mathcal{I}) = \sum_{i\in\mathcal{I}} u_i$ and let $\mathrm{wt}(\mathcal{O}) = \sum_{i\in\mathcal{O}} u_i$, where $\mathcal{O} \triangleq [n] \setminus \mathcal{I}$ is the set of outliers. Then, $\mathrm{wt}(\mathcal{I}) + \mathrm{wt}(\mathcal{O}) = 1$.

By contradiction, we show that if $\mathrm{wt}(\mathcal{I}) < \alpha$, then there exists a pseudo-distribution satisfying $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$ that achieves a lower value of $\left\|\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{\omega}]\right\|_2^2$, hence contradicting optimality of $\tilde{\mu}$. Towards this goal, we define a pseudo-distribution $\tilde{\mu}^\star$ which is supported on a single $(\boldsymbol{\omega}, \boldsymbol{x})$, the indicator vector $\mathbf{1}_{\mathcal{I}}$ and $\boldsymbol{x}^\circ$. Therefore, $\tilde{\mathbb{E}}_{\tilde{\mu}^\star}[\omega_i] = 1$ iff $i \in \mathcal{I}$ and zero otherwise. Clearly, $\tilde{\mu}^\star$ satisfies $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$. Therefore, any convex combination $\tilde{\mu}_\lambda = (1-\lambda)\tilde{\mu} + \lambda\tilde{\mu}^\star$ also satisfies $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$. We now show that whenever $\mathrm{wt}(\mathcal{I}) < \alpha$, then $\left\|\tilde{\mathbb{E}}_{\tilde{\mu}_\lambda}[\boldsymbol{\omega}]\right\|_2^2 < \left\|\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{\omega}]\right\|_2^2$ for some $\lambda > 0$, thus contradicting optimality of $\tilde{\mu}$. We observe that:

$$\boldsymbol{u}_\lambda = \frac{1}{\alpha\, n}\tilde{\mathbb{E}}_{\tilde{\mu}_\lambda}[\boldsymbol{\omega}] = \frac{1}{\alpha\, n}(1-\lambda)\tilde{\mathbb{E}}_{\tilde{\mu}}[\boldsymbol{\omega}] + \frac{1}{\alpha\, n}(\lambda)\tilde{\mathbb{E}}_{\tilde{\mu}^\star}[\boldsymbol{\omega}] = (1-\lambda)\boldsymbol{u} + \frac{\lambda}{\alpha\, n}\mathbf{1}_{\mathcal{I}}. \tag{168}$$

First, we compute the squared norm of $\boldsymbol{u}_\lambda$ using (168):

$$\|\boldsymbol{u}_\lambda\|_2^2 = \overbrace{(1-\lambda)^2\,\|\boldsymbol{u}\|_2^2 + 2\lambda(1-\lambda)\frac{\mathrm{wt}(\mathcal{I})}{\alpha\,n} + \frac{\lambda^2}{\alpha\,n}}^{\text{observing } \mathbf{1}_{\mathcal{I}}^\mathsf{T}\mathbf{1}_{\mathcal{I}} = \alpha\,n \text{ and } \mathbf{1}_{\mathcal{I}}^\mathsf{T}\boldsymbol{u} = \mathrm{wt}(\mathcal{I})}. \tag{169}$$

Next, we lower bound $\|\boldsymbol{u}\|_2^2$ in terms of $\mathrm{wt}(\mathcal{I})$ and $\mathrm{wt}(\mathcal{O})$. Observe that for any fixed values of $\mathrm{wt}(\mathcal{I})$ and $\mathrm{wt}(\mathcal{O})$, the minimum of $\|\boldsymbol{u}\|_2^2$ is attained by the vector $\boldsymbol{u}$ such that $u_i = \frac{1}{\alpha\,n}\mathrm{wt}(\mathcal{I})$ for each $i \in \mathcal{I}$ and $u_i = \frac{1}{(1-\alpha)n}\mathrm{wt}(\mathcal{O})$ otherwise. This gives:

$$\|\boldsymbol{u}\|_2^2 \geq \overbrace{\left(\frac{\mathrm{wt}(\mathcal{I})}{\alpha\,n}\right)^2 \alpha\,n}^{\text{sum of } u_i^2 \text{ for } i\in\mathcal{I}} + \overbrace{\left(\frac{1-\mathrm{wt}(\mathcal{I})}{(1-\alpha)n}\right)^2 (1-\alpha)n}^{\text{sum of } u_i^2 \text{ for } i\in\mathcal{O}} \tag{170}$$

$$= \frac{\mathrm{wt}(\mathcal{I})^2}{\alpha\,n} + \frac{(1-\mathrm{wt}(\mathcal{I}))^2}{(1-\alpha)n}$$

$$= \frac{1}{\alpha\,n}\cdot\left(\mathrm{wt}(\mathcal{I})^2 + (1-\mathrm{wt}(\mathcal{I}))^2\left(\frac{\alpha}{1-\alpha}\right)\right). \tag{171}$$

Combining (169) and (171):

$$\|\boldsymbol{u}_\lambda\|_2^2 - \|\boldsymbol{u}\|_2^2 = \overbrace{(-2\lambda+\lambda^2)\,\|\boldsymbol{u}\|_2^2}^{=(1-\lambda)^2\|\boldsymbol{u}\|_2^2 - \|\boldsymbol{u}\|_2^2} + 2\lambda(1-\lambda)\frac{\mathrm{wt}(\mathcal{I})}{\alpha\,n} + \frac{\lambda^2}{\alpha\,n} \tag{172}$$

$$\overbrace{\leq}^{\text{since } (-2\lambda+\lambda^2)\leq 0} \frac{-2\lambda+\lambda^2}{\alpha\,n}\left(\mathrm{wt}(\mathcal{I})^2 + (1-\mathrm{wt}(\mathcal{I}))^2\frac{\alpha}{1-\alpha}\right) + 2\lambda(1-\lambda)\frac{\mathrm{wt}(\mathcal{I})}{\alpha\,n} + \frac{\lambda^2}{\alpha\,n}. \tag{173}$$

Rearranging (note that this part slightly differs from [37], but with the same conclusion):

$$\|\boldsymbol{u}\|_2^2 - \|\boldsymbol{u}_\lambda\|_2^2 \geq \frac{\lambda}{\alpha\,n}\left((2-\lambda)\left(\mathrm{wt}(\mathcal{I})^2 + (1-\mathrm{wt}(\mathcal{I}))^2\frac{\alpha}{1-\alpha}\right) - 2(1-\lambda)\mathrm{wt}(\mathcal{I}) - \lambda\right) \tag{174}$$

$$= \frac{\lambda(2-\lambda)}{\alpha\,n}\left(\left(\mathrm{wt}(\mathcal{I})^2 + (1-\mathrm{wt}(\mathcal{I}))^2\frac{\alpha}{1-\alpha}\right) - \frac{2(1-\lambda)}{(2-\lambda)}\mathrm{wt}(\mathcal{I}) - \frac{\lambda}{(2-\lambda)}\right) \tag{175}$$

observing $\frac{2(1-\lambda)}{(2-\lambda)} = \frac{2(1-\lambda)}{2(1-\lambda)+\lambda} < 1$ (for $0 < \lambda \leq 1$)

and $\frac{1}{(2-\lambda)} \leq 1 < \frac{1-\mathrm{wt}(\mathcal{I})}{1-\alpha}$ (for $0 \leq \mathrm{wt}(\mathcal{I}) < \alpha \leq 1$)

$$> \frac{\lambda(2-\lambda)}{\alpha\,n}\left(\mathrm{wt}(\mathcal{I})^2 + (1-\mathrm{wt}(\mathcal{I}))^2\frac{\alpha}{1-\alpha} - \mathrm{wt}(\mathcal{I}) - \frac{1-\mathrm{wt}(\mathcal{I})}{1-\alpha}\lambda\right) \tag{176}$$

$$= \frac{\lambda(2-\lambda)}{\alpha\,n}\left(-\mathrm{wt}(\mathcal{I})(1-\mathrm{wt}(\mathcal{I})) + (1-\mathrm{wt}(\mathcal{I}))^2\frac{\alpha}{1-\alpha} - \frac{1-\mathrm{wt}(\mathcal{I})}{1-\alpha}\lambda\right) \tag{177}$$

$$= \frac{\lambda(2-\lambda)(1-\mathrm{wt}(\mathcal{I}))}{\alpha\,n(1-\alpha)}\left(-\mathrm{wt}(\mathcal{I})(1-\alpha) + (1-\mathrm{wt}(\mathcal{I}))\alpha - \lambda\right) \tag{178}$$

$$= \overbrace{\frac{\lambda(2-\lambda)(1-\mathrm{wt}(\mathcal{I}))}{\alpha\,n(1-\alpha)}}^{\geq 0}\left(\alpha - \mathrm{wt}(\mathcal{I}) - \lambda\right). \tag{179}$$

Now whenever $\text{wt}(\mathcal{I}) < \alpha$, $(\alpha - \text{wt}(\mathcal{I}) - \lambda) > 0$ for a sufficiently small λ. Thus we can choose a small enough $\lambda > 0$ such that $\|\boldsymbol{u}\|_2^2 - \|\boldsymbol{u}_\lambda\|_2^2 > 0$, which contradicts optimality of $\tilde{\mu}$. ■

Using Lemma 25 and Lemma 29 we can finally prove the correctness of Theorem 17. Let $\tilde{\mu}$ be a pseudo-distribution satisfying $\mathcal{T}_{\boldsymbol{\omega},\boldsymbol{x}}$ that minimizes $\left\|\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\boldsymbol{\omega}\right]\right\|_2^2$. Such a pseudo-distribution exists since the set contains at least the distribution with $\omega_i = 1$ iff $i \in \mathcal{I}$ and $\boldsymbol{x} = \boldsymbol{x}^\circ$.

From Lemma 25, we have $\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right] \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \le \frac{\alpha\eta M_x}{2}$. Let $Z \doteq \sum_{i\in\mathcal{I}} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{|\mathcal{I}|}$ (this is a normalization factor, such that $\frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{Z|\mathcal{I}|}$ is a valid pdf over the inliers, *i.e.*, sums up to 1). By a rescaling, we obtain:

$$\sum_{i\in\mathcal{I}} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]}{Z|\mathcal{I}|} \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \le \frac{1}{Z}\frac{\alpha\eta M_x}{2}. \tag{180}$$

Using Lemma 29, $Z \ge \alpha$. Therefore,

$$\sum_{i\in\mathcal{I}} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]}{Z|\mathcal{I}|} \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \le \frac{\eta M_x}{2}. \tag{181}$$

Let $i \in [n]$ be chosen with probability $\frac{\tilde{\mathbb{E}}_{\tilde{\mu}}[\omega_i]}{\alpha\, n}$. Then, we sample $i \in \mathcal{I}$ with probability $Z \ge \alpha$. By Markov's inequality:

$$\mathbb{P}\left(\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \le \eta M_x\right) = \mathbb{P}\left(\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \le \eta M_x | i \in \mathcal{I}\right) \cdot \overbrace{\mathbb{P}\left(i \in \mathcal{I}\right)}^{\ge\alpha} \tag{182}$$

$$\ge \alpha \cdot \mathbb{P}\left(\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \le \eta M_x | i \in \mathcal{I}\right) \tag{183}$$

$$\text{Markov's inequality: } \mathbb{P}\left(X \ge a\right) \le \frac{\mathbb{E}\left[X\right]}{a} \iff \mathbb{P}\left(X \le a\right) \ge 1 - \frac{\mathbb{E}\left[X\right]}{a}$$

$$\ge \alpha\left(1 - \frac{1}{\eta M_x}\mathbb{E}_{i\in\mathcal{I}}\left[\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2\right]\right) = \alpha\left(1 - \frac{1}{\eta M_x}\sum_{i\in\mathcal{I}} \frac{\tilde{\mathbb{E}}_{\tilde{\mu}}\left[\omega_i\right]}{Z|\mathcal{I}|} \|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2\right) \tag{184}$$

$$\overbrace{\ge}^{\text{using (181)}} \alpha\left(1 - \frac{1}{\eta M_x}\frac{\eta M_x}{2}\right) = \frac{\alpha}{2}. \tag{185}$$

So we concluded that $\mathbb{P}\left(\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \le \eta M_x\right) \ge \frac{\alpha}{2}$ (this is the probability that a single draw satisfies $\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \le \eta M_x$). Calling S (as in "success") the event that $\|\boldsymbol{v}_i - \boldsymbol{x}^\circ\|_2 \le \eta M_x$, we get that the probability of S after m draws is:

$$\mathbb{P}\left(S_m\right) = 1 - \overbrace{\left(1 - \mathbb{P}\left(S\right)\right)^m}^{\text{failing } m \text{ times}} \ge 1 - \left(1 - \frac{\alpha}{2}\right)^m \tag{186}$$

Finally, choosing the number of draws $m \ge \frac{N}{\alpha}$, we obtain

$$\mathbb{P}\left(S_m\right) \ge 1 - \left(1 - \frac{\alpha}{2}\right)^{\frac{N}{\alpha}} \tag{187}$$

which matches the first claim in Theorem 17.

Now the final claim (*i.e.,* the claim that (167) is satisfied with probability at least 0.99 for $\alpha \geq 0.01$ and $N = 10$) is just a particularization of (187) to the given choice of N. In particular, we first observe that the probability of success $1 - \left(1 - \frac{\alpha}{2}\right)^{\frac{N}{\alpha}}$ is a non-decreasing function of α. Then we note that the function $f(\alpha, N) \triangleq 1 - \left(1 - \frac{\alpha}{2}\right)^{\frac{N}{\alpha}}$ evaluated at $\alpha = 0.01$ and $N = 10$ is such that $f(0.01, 10) \geq 0.99$, which concludes the proof.

■