Supplementary Material for "The Solution Path Algorithm for Identity-Aware Multi-Object Tracking"

Shoou-I Yu[1], Deyu Meng[2], Wangmeng Zuo[3], Alexander Hauptmann[1]
[1] Carnegie Mellon University, [2] Xi'an Jiaotong University, [3] Harbin Institute of Technology
iyu@cs.cmu.edu, dymeng@mail.xjtu.edu.cn, wmzuo@hit.edu.cn, alex@cs.cmu.edu

Abstract

In this supplementary material, we provide the details of our Iterative Projection Method for solving the optimization problem (Eq. 4) in the main text, and present the theoretical substantiation for its convergence. Moreover, the complexity of our algorithm is analyzed. Finally, more experiment results such as the qualitative tracking results for the "nursing home long" data set and the analysis of how face recognition errors affect final tracking performance are shown.

1. Details of Our Iterative Projection Algorithm

We want to solve the following optimization problem:

$$\min_{\mathbf{G}} f(\mathbf{a}') = \left\|\mathbf{G} - \mathbf{a}'\right\|_2^2 \quad s.t. \quad \left\|\mathbf{G}\right\|_p \le 1, \tag{1}$$

where $\mathbf{a}', \mathbf{G} \in \mathbf{R}^c$, $\mathbf{a}' = (a_1, a_2, \cdots, a_c)^T$ is the known reference point and $\mathbf{G}$ is the queried projection point on ℓ_p ball $\left\|\mathbf{G}\right\|_p \le 1$.

The details of our iterative projection algorithm for solving (1) are as follows:

1: If $\left\|\mathbf{a}'\right\|_p \le 1$, then output $\mathbf{G} = \mathbf{a}'$, and terminate the entire procedure. Otherwise, record $\mathbf{s} = sign(\mathbf{a}')$ and reformulate $\mathbf{a}' = \mathbf{s} \odot \mathbf{a}'$ to make all its elements non-negative (i.e. to let $\mathbf{a}'$ located at the first quadrant). Initialize $\mathbf{G}^{(1)} = (g_1^{(1)}, g_2^{(1)}, \cdots, g_c^{(1)})^T$ as the intersection of line segment connecting $\mathbf{a}'$ and the origin and the boundary of the ℓ_p ball $\left\|\mathbf{G}\right\|_p = 1$. This intersection can be found efficiently by the binary search strategy. Set $l = 1$ and ε as a small threshold value.

2: **Repeat**:

3: Compute the tangent plane of the ℓ_p ball boundary curve $\left\|\mathbf{G}\right\|_p = 1$ at $\mathbf{G}^{(l)}$ as:

$$\pi^{(l)} = \{v | \mathbf{w}^{(l)T}(\mathbf{G} - \mathbf{G}^{(l)}) = 0\},$$

where

$$\begin{aligned} \mathbf{w}^{(l)} &= \left(\nabla \left\|\mathbf{G}\right\|_p\right)_{\mathbf{G}^{(l)}} \\ &= (p(g_1^{(l)})^{p-1}, p(g_2^{(l)})^{p-1}, \cdots, p(g_c^{(l)})^{p-1})^T, \end{aligned}$$

where $g_i^{(l)}$ is the i^{th} element of $\mathbf{G}^{(l)}$. Calculate the projection point of $\mathbf{a}'$ to $\pi^{(l)}$ as

$$\mathbf{x}^{(l)} = \mathbf{a}' - \frac{\mathbf{w}^{(l)T}\mathbf{a}' - \mathbf{w}^{(l)T}\mathbf{G}^{(l)}}{\left\|\mathbf{w}^{(l)}\right\|_2^2}\mathbf{w}^{(l)}.$$

Complexity of this step is $\mathcal{O}(c)$.

4. If $\mathbf{x}^{(l)}$ is located in the first quadrant (i.e., $\mathbf{x}^{(l)} \succeq 0$), then draw a line segment $\mathbf{z}(t)$ between $\mathbf{a}'$ and $\mathbf{x}^{(l)}$ as

$$\mathbf{z}(t) = (\mathbf{x}^{(l)} - \mathbf{a}')t + \mathbf{a}', 0 \le t \le 1,$$

and compute its intersection point $\mathbf{G}^{(l+1)}$ with the ℓ_p ball boundary curve $\left\|\mathbf{G}\right\|_p = 1$ using binary search. Then let $l = l + 1$, and go to the next iteration. Complexity of this step is $\mathcal{O}(c + \log(c))$, where binary search contributes to the $\log(c)$. The iteration count of binary search is a constant and not explicity written in the complexity formula.

5. If $\mathbf{x}^{(l)}$ is located outside the first quadrant, then calculate

$$t^* = \min_i(t_i^*), t_i^* = \frac{a_i}{a_i - x_i},$$

where a_i and x_i are the i^{th} elements of $\mathbf{a}'$ and $\mathbf{x}^{(l)}$, respectively. If $\mathbf{z}(t^*) = (\mathbf{x}^{(l)} - \mathbf{a}')t^* + \mathbf{a}'$ satisfies that $\left\|\mathbf{z}(t^*)\right\|_p \le 1$, then use the similar binary search strategy as step 4 to calculate the projection point $\mathbf{G}^{(l+1)}$. Then let $l = l + 1$, and go to the next iteration. Complexity of this step is $\mathcal{O}(c + \log(c))$.

6. If $\left\|\mathbf{z}(t^*)\right\|_p > 1$, this means that there is no intersection of the line segment $\mathbf{z}(t)$ ($0 \le t \le 1$) and the ℓ_p ball boundary curve $\left\|\mathbf{G}\right\|_p = 1$. Calculate the critical point $\mathbf{y}(s^*)$ where

$$\mathbf{y}(s) = (\mathbf{x}^{(l)} - \mathbf{G}^{(l)})s + \mathbf{G}^{(l)}, 0 \le s \le 1,$$

and

$$s^* = \min_i(s_i^*), s_i^* = \frac{g_i^{(1)}}{g_i^{(1)} - x_i}.$$

And then draw a line segment $\mathbf{z}(t)$ between $\mathbf{a}'$ and $\mathbf{y}(s^*)$ and compute its intersection point $\mathbf{G}^{(l+1)}$ with the ℓ_p ball boundary curve $\|\mathbf{G}\|_p = 1$ using the binary search strategy. Then let $l = l + 1$, and go to the next iteration. Complexity of this step is $\mathcal{O}(c + \log(c))$.

7. **End Repeat** when $\left\|\mathbf{G}^{(l)} - \mathbf{G}^{(l-1)}\right\| < \varepsilon$
8. Output the projection point $\mathbf{G} = \mathbf{s} \odot \mathbf{G}^{(l)}$.

In sum, the complexity of the above steps is $\mathcal{O}(c + \log(c)) \times \text{MaxIter} = \mathcal{O}(c + \log(c))$. Also, computing $\mathbf{a}'$ requires $\mathcal{O}((k + q)c)$, where k is the number of neighbors for the current observation (stored in the Laplacian matrices $\mathbf{K}$ and $\mathbf{L}$, and q is the number of constraints the current observation has. So the combined complexity of the iterative projection method used to solve Equation 6 in the main paper is $\mathcal{O}((k + q + \log(c))c)$.

2. Theoretical Principle of Our Algorithm

We use step-by-step remarks to explain the theoretical principle underlying our algorithm in detail.

2.1. Remarks for steps 1 and 8

Remark 1: When $\|\mathbf{a}'\|_p > 1$, it is easy to prove that its projection on the ℓ_p ball $\|\mathbf{G}\|_p \leq 1$ (the solution of (1)) is located on its boundary $\|\mathbf{G}\|_p = 1$. Furthermore, its projection lies on the same quadrant with $\mathbf{a}'$ [1].

Remark 2: Due to the symmetry property of the ℓ_2 objective and the ℓ_p constraint of (1), we can equivalently solve this optimization problem by getting the solution $\mathbf{G}$ for $f(|\mathbf{a}'|)$, where $|\mathbf{a}'| = \mathbf{s} \odot \mathbf{a}'$ and $\mathbf{s} = sign(\mathbf{a}')$ (step 1), and then transfer $\mathbf{G}$ (with all positive elements according to Remark 1) back to $\mathbf{G} = \mathbf{s} \odot \mathbf{G}$ (step 8). Here $\odot$ is the Hadamard product meaning the element-wise multiplication between two vectors.

Remark 3: When $\|\mathbf{a}'\|_p > 1$, since $\|\mathbf{0}\|_p < 1$, the intersection of line segment connecting $\mathbf{a}'$ and the origin $\mathbf{0}$ and the unit ℓ_p ball boundary $\|\mathbf{G}\|_p = 1$ can definitely be found.

2.2. Remark for step 3

Remark 4: In the first quadrant, it is evident that the unit ℓ_p ball boundary curve $\|\mathbf{G}\|_p = 1$ is convex. This means that the tangent plane of this curve at $\mathbf{G}^{(l)}$ is below it. See Figure 1 for better understanding.

2.3. Remark for step 4

Remark 5: Since $\|\mathbf{a}'\|_p > 1$ and $\left\|\mathbf{x}^{(l)}\right\|_p \leq 1$ (based on Remark 4), the intersection $\mathbf{G}^{(l+1)}$ of the line segment $\mathbf{z}(t)$ $(0 \leq t \leq 1)$ and $\|\mathbf{G}\|_p = 1$ definitely exists.

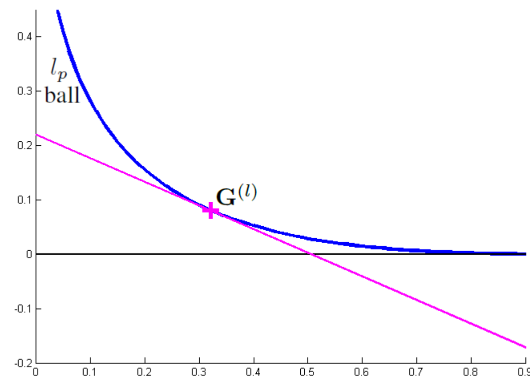

Figure 1: Principle illustration for Remark 4.

Since $\mathbf{x}^{(l)}$ is the projection of $\mathbf{a}'$ on $\pi^{(l)}$ and $\mathbf{G}^{(l)}$ is located on $\pi^{(l)}$, we have $\left\|\mathbf{x}^{(l)} - \mathbf{a}'\right\|_2^2 \leq \left\|\mathbf{G}^{(l)} - \mathbf{a}'\right\|_2^2$. Besides, since $\mathbf{G}^{(l+1)}$ is obtained at $\mathbf{z}(t')$ for certain $0 \leq t' \leq 1$ and $\mathbf{x}^{(\mathbf{l})}$ and $\mathbf{a}'$ are the two end points of $\mathbf{z}(t)$, we have $\left\|\mathbf{G}^{(l+1)} - \mathbf{a}'\right\|_2^2 \leq \left\|\mathbf{x}^{(l)} - \mathbf{a}'\right\|_2^2$. It thus holds that $\left\|\mathbf{G}^{(l+1)} - \mathbf{a}'\right\|_2^2 \leq \left\|\mathbf{G}^{(l)} - \mathbf{a}'\right\|_2^2$. See Figure 2 for better understanding.

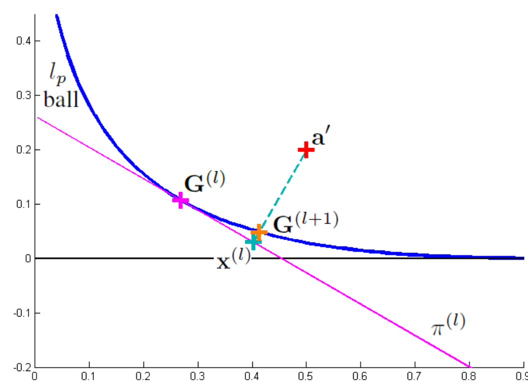

Figure 2: Principle illustration for Remark 5.

2.4. Remark for steps 5 and 6:

Remark 6: Along the line segment $\mathbf{z}(t)$ $(0 \leq t \leq 1)$ connecting $\mathbf{a}'$ and $\mathbf{x}^{(l)}$, the critical point of its i^{th} element varying from positive to negative can be calculated by:

$$(x_i - a_i)t_i^* + a_i = 0 \implies t_i^* = \frac{a_i}{a_i - x_i}.$$

Then it is evident that the critical point of $\mathbf{z}(t)$ varying out from the first quadrant at

$$t^* = \min_i(t_i^*).$$

We then have:

(i) When $\|\mathbf{z}(t^*)\|_p \leq 1$, since $\|\mathbf{a}'\|_p > 1$, the intersection of $\mathbf{z}(t)$ $(0 \leq t \leq 1)$ and $\|\mathbf{G}\|_p = 1$ exists in the first quadrant. Thus we can use binary search to find this intersection point. Based on the similar proof as Remark 5, we have $\left\|\mathbf{G}^{(l+1)}-\mathbf{a}'\right\|_2^2 \leq \left\|\mathbf{G}^{(l)}-\mathbf{a}'\right\|_2^2$. Please see Figure 3 for better understanding.

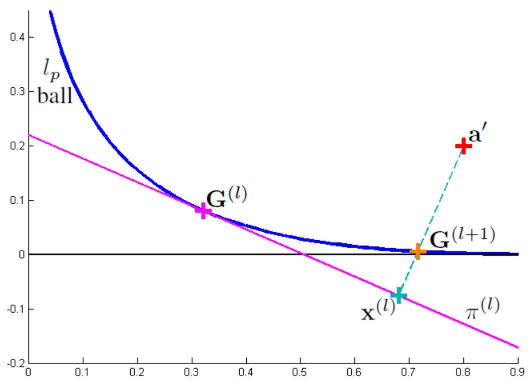

Figure 3: Principle illustration for Remark 6(i).

(ii) When $\|\mathbf{z}(t^*)\|_p > 1$, we know that $\mathbf{z}(t^*)$ is not inside the ℓ_p ball $\Omega = \{\mathbf{G}|\, \|\mathbf{G}\|_p \leq 1\}$. Since $\overline{\Omega}$ in the first quadrant is convex (equivalent to that its boundary curve $\|\mathbf{G}\|_p = 1$ in the first quadrant is convex) and $\mathbf{a}' \in \overline{\Omega}$, it holds that the entire line segment $\mathbf{z}(t)$ $(0 \leq t \leq 1)$ is in $\overline{\Omega}$ and has no intersection with the curve $\|\mathbf{G}\|_p = 1$ in the first quadrant. We thus utilize the following strategy to find the next iteration point.

By connecting the last iteration point $\mathbf{G}^{(l)}$ and the projection point $\mathbf{x}^{(l)}$, we can formulate a line segment $\mathbf{y}(t)$ $(0 \leq t \leq 1)$. Using the similar strategy like (i), we can find the critical point $\mathbf{y}(s^*)$ at which $\mathbf{y}(t)$ goes out from the first quadrant, where

$$s^* = \min_i(s_i^*), s_i^* = \frac{g_i^{(1)}}{g_i^{(l)} - x_i},$$

where $g_i^{(l)}$ and x_i are the i^{th} element of $\mathbf{G}^{(l)}$ and $\mathbf{x}^{(l)}$, respectively. Since both $\mathbf{y}(s^*)$ and $\mathbf{G}^{(l)}$ are on the tangent plane $\pi^{(l)}$, and $\mathbf{y}(s^*)$ is closer to the projection point $\mathbf{x}^{(l)}$ of $\mathbf{a}'$ than $\mathbf{G}^{(l)}$, we have that $\|\mathbf{y}(s^*)-\mathbf{a}'\|_2^2 \leq \left\|\mathbf{G}^{(l)}-\mathbf{a}'\right\|_2^2$.

Since $\pi^{(l)}$ is below the curve $\|\mathbf{G}\|_p = 1$ based on Remark 4, we know that $\|\mathbf{y}(s^*)\|_p \leq 1$. Then together with $\|\mathbf{a}'\|_p > 1$, it holds that the intersection $\mathbf{G}^{(l+1)}$ of the line segment connecting $\mathbf{a}'$ and $\mathbf{y}(s^*)$ and $\|\mathbf{G}\|_p = 1$ definitely exists in the first quadrant, and $\left\|\mathbf{G}^{(l+1)}-\mathbf{a}'\right\|_2^2 \leq \|\mathbf{y}(s^*)-\mathbf{a}'\|_2^2$. We thus have $\left\|\mathbf{G}^{(l+1)}-\mathbf{a}'\right\|_2^2 \leq \left\|\mathbf{G}^{(l)}-\mathbf{a}'\right\|_2^2$. The aforementioned can be easily understood by observing Figure 4.

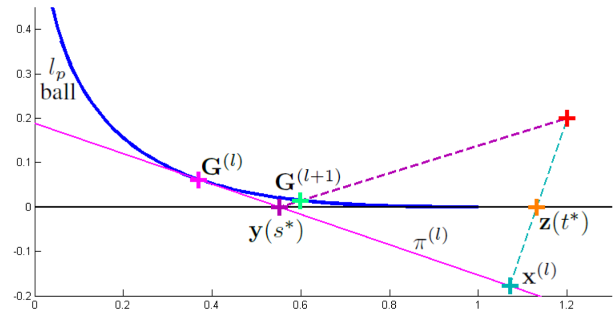

Figure 4: Principle illustration for Remark 6(ii).

Based on the aforementioned Remarks 4, 5 and 6, we know that during the iterative process of our algorithm, the objective $\left\|\mathbf{G}^{(l)}-\mathbf{a}'\right\|_2^2$ is monotonically decreasing with respect to the iteration number l under the constraint $\left\|\mathbf{G}^{(l)}\right\|_p \leq 1$. Our algorithm is thus convergent and expected to get a rational local minimum of the original problem.

3. Computational Complexity of Solution Path Optimization

For clarity, the following equations were copied from the main paper. The final loss function we are solving is as follows:

$$\begin{gathered} \min_{\mathbf{F}} Tr\left(\mathbf{F}^T(\mathbf{L}+\mathbf{K})\mathbf{F}\right) \\ s.t.\ \forall(a,b) \in \mathcal{Y}, F_{ab} = 1,\ \|\mathbf{F}_r\|_0 \leq 1,\ 1 \leq r \leq n \\ \forall(i,j) \in \mathcal{T}, \left\|\left[\mathbf{F}_{il} \quad \mathbf{F}_{jl}\right]\right\|_0 \leq 1, 1 \leq l \leq c. \end{gathered} \tag{2}$$

If we focus on the variables of a single observation, then Equation 2 becomes:

$$\begin{gathered} \min_{\mathbf{G}_i} \frac{1}{2}\|\mathbf{G}_i - \mathbf{a}\|_2^2\ \ s.t.\ \ \|\mathbf{G}_i\|_p \leq 1, \\ (i,\forall j) \in \mathcal{T}, \left\|\left[\mathbf{G}_{il}, \quad \mathbf{G}_{jl}\right]\right\|_p \leq 1, 1 \leq l \leq c, \end{gathered} \tag{3}$$

We optimize Equation 2 by updating $\mathbf{G}_i$ with Equation 3 for each of the n observations. As proved in Section 1 of the supplementary materials, the iterative projection algorithm requires $\mathcal{O}\left(c(k+q+\log(c))\right)$ per iteration, where q is the average number of constraints per observation, and k is the number of nearest neighbors of each observation. Thus, solving Equation 2 is $\mathcal{O}\left(nc(k+q+\log(c)) * \text{MaxIter}\right)$, which is efficient as it is approximately linear in the number of observations (n), classes (c), nearest neighbors (k), and constraints (q).

4. Qualitative Tracking Results

Due to space constraints, the qualitative tracking results for the *nursing home long* data set are shown in Figure 5.

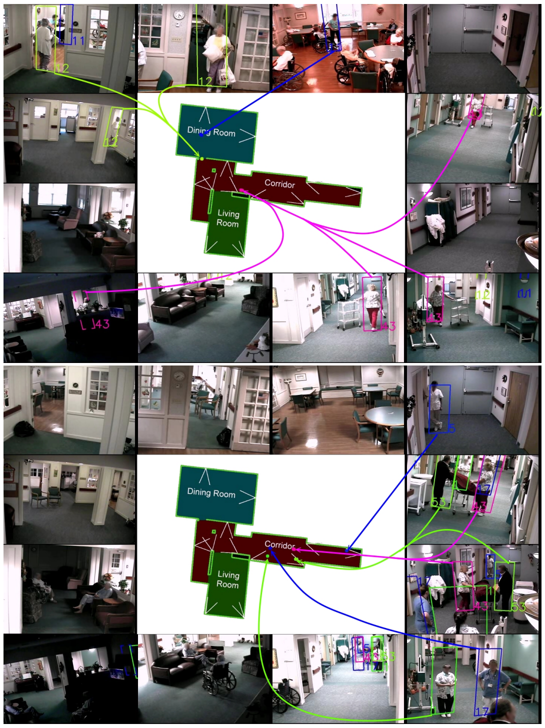

Figure 5: Snapshots of tracking results on *nursing home long* data set. Not all arrows are drawn for clarity.

5. Face Recognition Accuracy and Tracking Performance

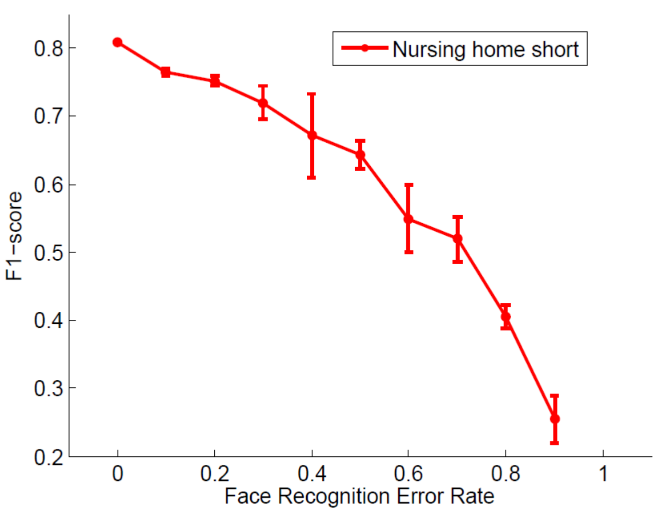

Figure 6: Effect of face errors on tracking performance for *nursing home short*.

Our tracker utilizes the output of the face recognizer to perform tracking, but face recognizers are not perfect. Therefore, we analyzed the effect of face recognition errors on tracking performance. We focused on *nursing home short* sequence, where manual verification shows around 2% face recognition error rate. The high accuracy is due to 1) limited number of people in the scene and 2) we manually mapped face clusters computed by the PittPatt toolkit into each individual. In other scenarios, we may not be able to achieve such high face recognition accuracy, thus experiments were performed where we randomly corrupted face recognition results by changing a face recognition result from one individual to the other. Results are shown in Figure 6, At each error rate, the experiment was repeated 3 times, and the 95% confidence interval is drawn. Results show that errors in face recognition significantly hurt performance, and when the face recognition error rate is less than 70%, a 25% accuracy drop in face recognition will cause around 10% drop in tracking F1-score.

References

[1] S. Bahmani and B. Raj. A unifying analysis of projected gradient descent for lp-constrained least squares. In *ArXiv:1107.4623v5*, 2012. 2