Cinematic Shot Generation using Convex Optimization

Haran Raajesh

May 2023

1 Abstract

This research focuses on improving the smoothness of bounding box motion in a reconstructed video of two dancers. The initial data set provided included the video and bounding boxes for each frame. The bounding boxes were represented by the X and Y coordinates of the center and half of the height of the boudning box.However, the reconstructed video exhibited significant jitters in the bounding box movement, resulting in an inconsistent representation of the dancers' motion. To address this issue, we framed it as a convex optimization challenge and utilized convex optimization methods. Our objective was to refine the video by minimizing jitters and achieving smoother bounding box motion. Through this project, we demonstrate the effectiveness of our approach in enhancing the overall quality of the visual tracking system by achieving smoother and more consistent bounding box movements.

2 Introduction

Convex optimization is a subset of mathematical optimization that deals with the minimization of convex functions over convex sets. At its core, convex optimization focuses on finding the global minimum of a function, which makes it uniquely beneficial for tackling a wide variety of problems in various domains. These domains range from machine learning and data analysis to control systems and image processing. In a convex optimization problem, the shape of the feasible region ensures that any local minimum is also a global minimum, making the solution easier to find compared to non-convex problems. A key advantage of using convex optimization is that it provides an efficient, accurate, and highly robust methodology for solving problems. It has been successfully applied in many fields, including ours - visual tracking in video processing.

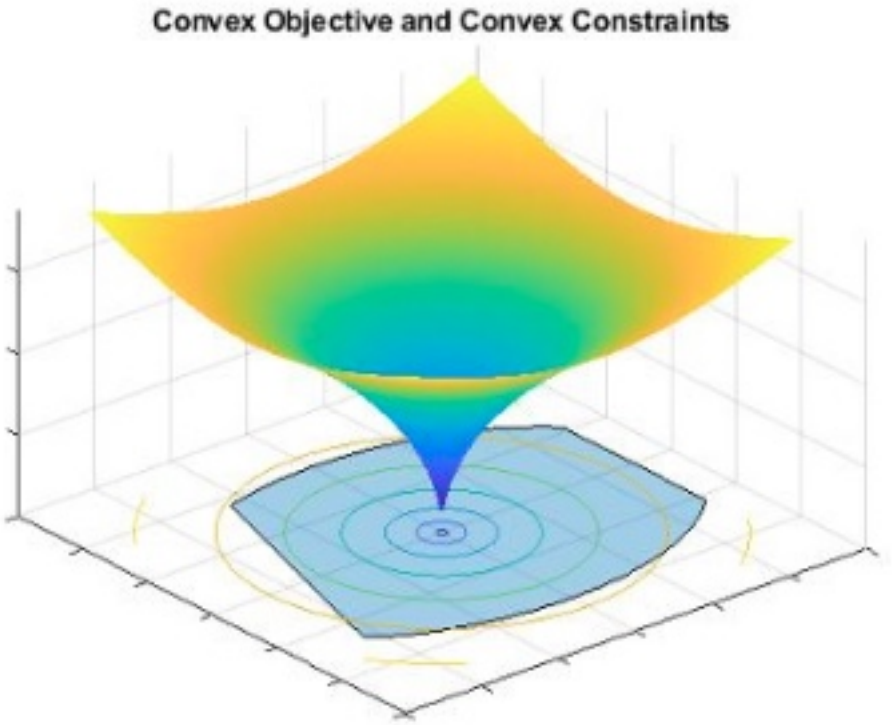


Figure 1: Convex Plot

3 Method

3.1 Simpler Example

To better understand the solution presented in the next subsection it is better to look at a simpler example. Fig 2 a) shows a noisy signal, let's refer to it as x, and our task it smooth out the signal, get an output signal, let's refer to it as r, similar to Fig 2 b).

We think of this problem as trying to minimize the following function:

$$f(r) = \frac{1}{2}\sum(r-x)^2 + Reg(r)$$

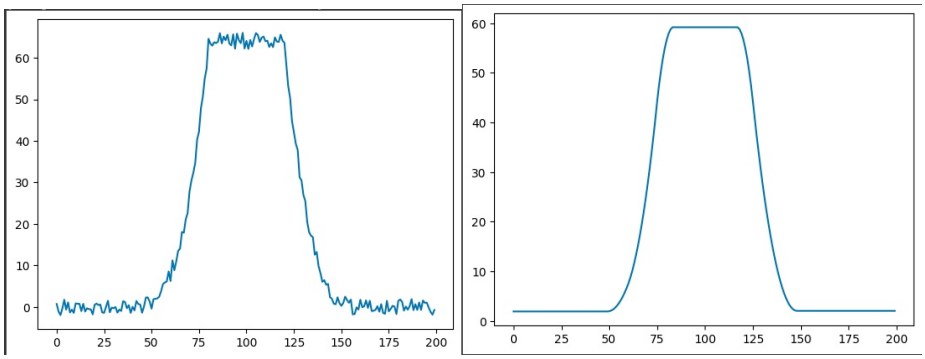
The above expression has two parts. The first is the sum of the squared differences of the 2 vectors r and x. The minimization of this part of the expression will ensure that the output signal r will try to match the the input signal as much as possible.

The second part of the expression is $Reg(r)$, which is a regularization term introduced to prevent the output signal from matching the noisy input signal exactly. More precisely this regularization term looks to penalize high differences between consecutive values in the output signal.

According to Dr Vineet Gandhi's Paper[1], a linear combination of the first order l1 norm and the third order l1 norm of the output signal r is a very effective regularizer for this task. The first order l1 term penalizes any cropping window motion if the actor included in shot specification is static while the third order l1 term will give jerk free transitions at the start and stop of the camera movement, with segments of constant acceleration and deceleration.

So the final function to minimize is the following:

$$f(r) = \tfrac{1}{2}\sum(r-x)^2 + 100\sum_{i=1}^{N-1}|r_{i+1}-r_i| + 100\sum_{i=1}^{N-3}|r_{i+3}-3r_{i+2}+3r_{i+1}-r_i|$$



(a) Noisy Signal (b) Smoothed out signal

Figure 2

3.2 Smoothing Out the Bounding Box Movement

Now in the data set, we are given the bounding box in terms of the x and y coordinates of the center of the bounding box and half the height of the bounding box for each frame. If we plot the X coordinate values, Y coordinate values and the height values separately for each of the frames, we get the graphs seen in Fig 3.

Now we can see the similarity to the example above. Each of the attributes of the bounding box over all the frames can be seen as a noisy signal that needs to be smoothed out. We can treat each attribute the same way we treated the

above signal and smooth it out the same way. The results can be seen in Fig 4.

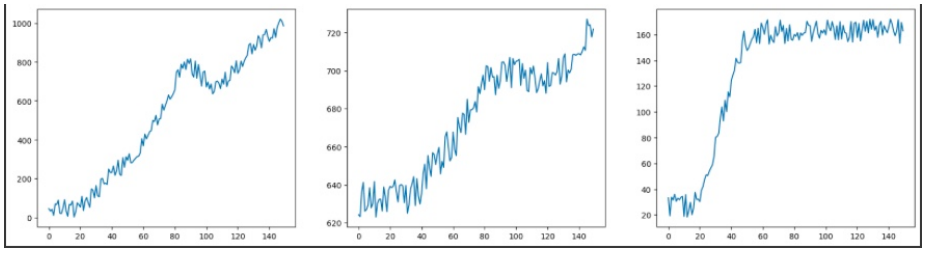


Figure 3: The left most images is the plot for X coordinate of the center across the frames, the middle image is the plot for the Y coordinate and the right most image is the plot for the height

4 Result

Figure 4 shows the result of the smoothing out of the values of the bounding boxes for the frames of the video. Now upon rejoining the frames the final result will be smooth movement of the bounding boxes appropriately following the movement of the dancers.

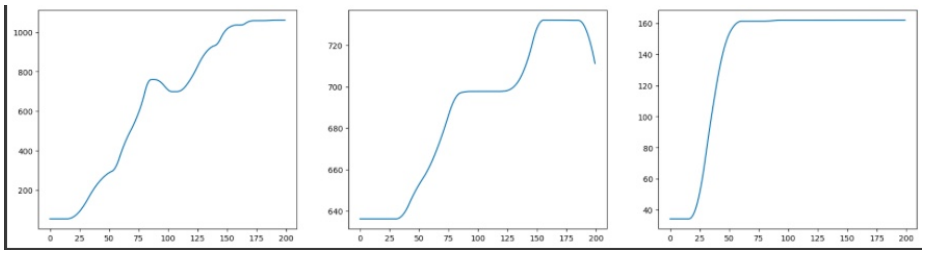


Figure 4: Figure shows the smoothed out values.The left most images is the plot for X coordinate of the center across the frames, the middle image is the plot for the Y coordinate and the right most image is the plot for the height

References

[1] Vineet Gandhi, Remi Ronfard, and Michael Gleicher. Multi-Clip Video Editing from a Single Viewpoint. In *European Conference on Visual Media Production (CVMP)*, 2014.