Theo Diamandis: 2/2
Your introduction explains the problem you wish to solve in a clear,
concise manner. The previous works section provided a good background,
which made clear your contribution to the literature. I was surprised
that few previous works looked at the convergence of projected GD. It
sounds like you all found a great problem to tackle. The following
section proving your convergence result was organized, clear, and
concise. Looking forward to seeing your future work on convergence
guarantees.

I list more specific comments and suggestions below.

First, three minor points. In Section 1, I noticed that you used the
ball B(R) notation before it is introduced in section 1.2. In Section
1.1, should the second sentence say "have been proven" rather than
"have proven"? In Section 1.2, you don't mention your Otilde(n)
notation. If you use this notation in Section 3, it might be useful to
put in Section 1.2 as a reference to the reader.

In Section 2, I found the figure difficult to read, especially when
printed out. I would make the lines thicker and darker, possibly using
markers rather than color to differentiate. In addition, since the
figure was not referenced in your paper, I was unsure if it served to
support a particular point or simply show GD steps. Also, since the
example is low dimensional, I would have liked to know the matrix A
and vector b used to create the instance of (1).

My theory background is lacking, so take what follows with a grain of
salt. I found two parts of Section 2 tricky to follow: the beginning
of the Lemma 2.5 proof, which invokes Lemma 2.4, was not obvious to
me; and I did not follow the ||x_tilde|| = R case in your final proof.
In the latter case, I was not sure where the gradient expression of
(1−c)*eta^−1*x came from.

--
Rahul Trivedi: 2/2
The authors mention that they exclude Lemma 2.3 and 2.4 due to space
constraints. It might be a good idea to provide a sketch of the proof
here since both of them seem crucial to the proof in Lemma 2.5 and
proposition 2.7.
The statement that 'The bound in (6) implies that the displayed series
is convergent as T → ∞ and thus φ(x(k)) → 0' is a bit unclear. It
might be a good idea to expand on this a bit more. In particular, on
reading the proof it seems that the fact that \sum ||x^(k) − x^(k+1)||
^2 goes is bounded by f(x^(0))−f(x∗) implies that x^(k) −> x∗ as k−>
infty. I am not quite sure why this is necessarily true, unless some
specific properties of the projected gradient descent are exploited

Other than that, the paper was very well written and the proofs were
very connected and understandable. I look forward to seeing the

convergence guarantees that you come up with.

--
Pulkit Tandon: 2/2
Problem is very clearly motivated and the idea seems neat. Paper is
well written.
Figure 1 is unclear to me. What is the experiment? It is not explained
in the text. Also, it will be nice to have a clearer figure: what
exactly are the 'orange' and 'green' lines and the dots on them
(caption of the figure is unclear).
In proposition 2.7, convergence of ϕ is unclear to me.

Hope this helps and I would love to see the complete proofs of Lemma
2.3, 2.4 and the convergence rate results in your final report.

--
Georgia Murray: 2/2
1) Can you provide some intuition into the assumptions in section 2.1?
The first seems like an actually necessary assumption (although I
haven't thought about why it is strictly necessary), but the second
just seems like it is for convenience. Is this the case, or is
initializing at the origin strictly necessary for some reason?

2) Just to clarify, you have already proven all the lemmas in section
2.2? You imply that you have, but I wasn't entirely sure.

3) Since (if I understand correctly) the point of your work is that
you can use gradient descent to solve the trust-region problem despite
its nonconvexity, I think it would make more sense to have Figure 1
show a nonconvex example.

Sorry, I wish I had more over-arching feedback, but great job!

--
Akshay Rajagopal: 2/2
For the problem instance used for Figure 1, is A an indefinite matrix?
If not, since projected gradient descent is able to solve the trust
region problem even with A indefinite, it might be nice to use an
indefinite A to illustrate that.
I know there's not much extra space, but if you can include it, maybe
an exact statement of the result you will prove going forward ("the
equivalent of Theorem 3.1, [CD16].") could help.
I was curious how the results in section 2.2 would fail if assumptions
2.1 and 2.2 did not hold. That's probably not necessary for this
midterm report, but a brief mention of it could be good for the final
report.

Otherwise, I think it looks good. Looking forward to seeing how it
progresses!