

1 To begin, we would like to extend our thanks to the reviewers for their detailed and insightful remarks. We reply to the  
2 reviewers’ major concerns below:

3 **Reviewer 1.** We thank the reviewer for their positive evaluation and encouraging remarks!

- 4 1. Regarding the lack of numerical experiments: we would kindly refer the reviewer to Section D of the  
5 supplement where we perform a series of numerical experiments on the resource sharing problem discussed  
6 in the paper’s body. In these experiments, we compare the convergence rate of Korpelevich’s extra-gradient  
7 algorithm to mirror-prox with a suitable Bregman function, and our adaptive mirror-prox variant. We wanted  
8 to include these experiments in the main body of the paper, but we ran out of space – we would be happy to  
9 use the extra page to do so if the paper is accepted.
- 10 2. On the complexity of calculating mirror steps: indeed, in full generality, calculating mirror steps can be  
11 computationally expensive (just as Euclidean projections can be bottlenecks in many problems). As is common  
12 in the literature (see e.g., Nemirovski, 2004, Nesterov, 2007, Juditsky et al., 2011, and the many works  
13 citing these papers), we were tacitly assuming a “prox-friendly” setting where mirror steps can be computed  
14 efficiently. The example we presented is indeed prox-friendly because it only involves the calculation of a  
15 single Lagrange multiplier (which can be accomplished efficiently via a simple line search). We will gladly  
16 explain this in more detail in the paper.
- 17 3. Beyond monotonicity: excellent suggestion! Albeit outside the scope of this work, local convergence in  
18 non-monotone VIs is one of our future research agendas and we would be happy to discuss it in more detail.

19 **Reviewer 2.** We thank the reviewer for their positive recommendation! We address their main concerns below:

- 20 1. On numerical experiments: please see Section D of our paper’s supplement and our reply to Reviewer 1 above.
- 21 2. On adaptation via backtracking: this is a very interesting suggestion and we thank the reviewer for making it.  
22 Two particularly promising works in that direction are the 2006 JCAM paper of Y. He and the 2018 SIOPT  
23 paper of Malitsky and Pock. Carrying out a linesearch analysis along these lines lies beyond the scope of the  
24 current work, but we would be happy to discuss this promising agenda in detail in a revision.
- 25 3. On (4.2b): yes, this was supposed to refer to the dual of the local norm. Apologies for any confusion!
- 26 4. On the continuity condition of #4247: we believe there may have been a misunderstanding here. To put  
27 things on an equal footing, consider the problem of minimizing a given convex function  $f$  (the common  
28 denominator of variational inequalities and online/stochastic optimization problems). The growth condition of  
29 #4247 extends the notion of **Lipschitz continuity** of  $f$  to singular problems; by contrast, Bregman continuity  
30 (this paper) extends the notion of **Lipschitz continuity \*of the gradient of  $f^*$**  to singular problems.  
31 The optimization literature makes a clear distinction between these two settings (often referred to as “non-  
32 smooth” and “smooth” respectively), because they lead to very different algorithms and oracle complexity  
33 bounds – see e.g., Nesterov’s (2004) classical textbook on the topic. In particular, relative to #4247: (a) the  
34 algorithms are different (mirror-prox vs. FTRL); (b) the rates obtained are different (mirror-prox achieves  
35 a  $\mathcal{O}(1/T)$  convergence rate in deterministic minimization problems whereas, without further assumptions,  
36 FTRL only achieves  $\mathcal{O}(1/\sqrt{T})$ ); and, of course, (c) as we stated above, the regularity conditions considered  
37 are clearly different, even in convex minimization problems.
- 38 5. On (5.2): yes, the  $\rho$  below (5.2) should be  $\theta$ , thanks!

39 **Reviewer 3.** We thank the reviewer for their positive evaluation and constructive feedback! Our replies follow below:

- 40 1. Regarding the “Lipschitz-like” condition of [6]: to clarify, this is an extension of Lipschitz gradient continuity  
41 for functions with singularities, and is designed to yield an appropriate descent lemma. Since there is no longer  
42 an objective function to minimize, descent lemmas are not applicable to variational inequality problems; as a  
43 result, we had to seek a different starting point altogether.
- 44 2. Regarding the statement in Lines 13-15: we see the reviewer’s point, thanks! We did not mean for the sentences  
45 to be parsed in that order, we will of course change the phrasing accordingly.
- 46 3. Regarding the definition of the local norm in lines 170-171: by a “continuous assignment of a norm” we had  
47 in mind the definition of an absolutely homogeneous Finsler metric as in the classic textbook of Bao, Chern  
48 and Shen, 2000. Specifically, a local norm is a continuous nonnegative function  $F: \mathcal{X} \rightarrow \mathcal{V} \rightarrow \mathbb{R}_+$  with the  
49 following properties for all  $x \in \mathcal{X}$ ,  $z_1, z_2 \in \mathcal{V}$ : (i)  $F(x, z_1 + z_2) \leq F(x, z_1) + F(x, z_2)$ ; (ii)  $F(x, \lambda z) =$   
50  $|\lambda|F(x, z)$ ; and (iii)  $F(x, z) > 0$  whenever  $z \in \mathcal{V} \setminus 0$ . The “local norm” at  $x$  would then be  $\|z\|_x = F(x, z)$   
51 in this language. We will be happy to provide this more detailed definition in the paper.
- 52 4. Regarding the numerical experiments in the supplement: AMP was run with  $\theta = 0.9$ . In practice, we saw very  
53 little difference for values of  $\theta$  between 0.5 and 0.99.