1 12/15/2015: Łojasiewicz-like inequality proved.

I have proved the following result:

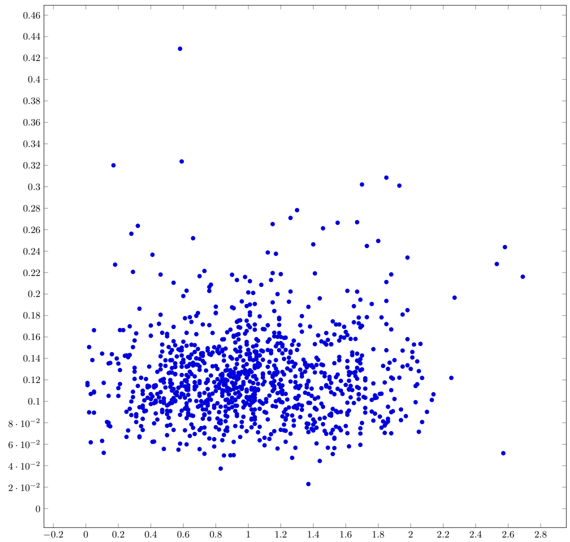
Theorem 1.1 *Suppose E is a bounded multivariate rational function over $\mathbb{R}^m$. Given any point $\mathbf{p} \in \mathbb{R}^m$ and almost any direction $\mathbf{d} \in \mathbb{R}^m$, there exist an open set $U_* \subset \mathbb{R}^m$ and constants $\theta \in \left(0, \frac{1}{2}\right]$ and $k > 0$ such that for all $\mathbf{x} \in \operatorname{domain}(E) \cap U_*$,*

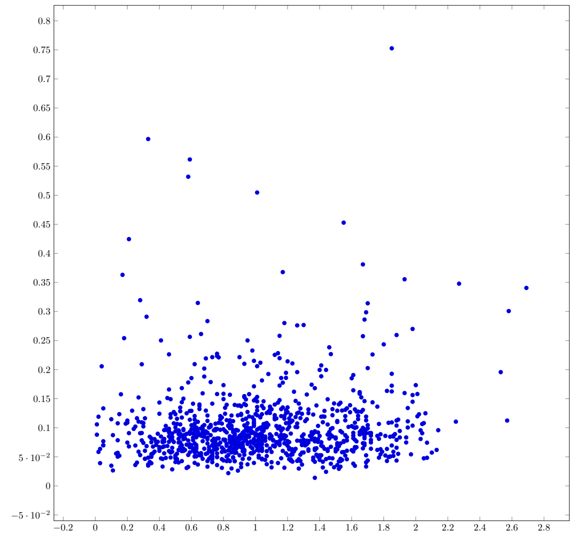
$$\left| E(\mathbf{x}) - \lim_{s \to 0} E(\mathbf{p} + s\mathbf{d}) \right|^{1-\theta} \leq k \left\| \nabla E(\mathbf{x}) \right\| .$$

Further, U_* may be chosen such that there exists some $s \in (0, 1]$ for which $\mathbf{p} + s\mathbf{d} \in U_*$, and such that for all $\mathbf{y} \in U_*$ and $t \in [-1, 0) \cup (0, 1]$, $t\mathbf{y} + (1 - t)\mathbf{p} \in U_*$.

This partially completes a project that I have been working on since late October.
It remains to be shown that the cones U_* may be used to contain the sequence produced by ALS. However, if this can be proved, the above result implies that the ALS algorithm creates convergent sequences.

2 01/19/2016: "Badness" estimates and greased ALS





Badness $= \frac{f}{\|\nabla f^2\|^2}$ Badness $= \frac{f^2}{\|\nabla f^2\|^2}$

The graphs above compare two formulations of Dr. Mohlenkamp's "badness" estimate. In particular, I attempted to determine experimentally whether the expected value of the badness estimates would be affected by the parameter used for greasing. The horizontal axis of each graph indicates a choice of greasing parameter, while the vertical axis is the "badness" estimate attained after several iterations of greased ALS.

The expected value of the badness estimate appears to be independent of the greasing parameter.

3 03/15/2016: Line-search extrapolation advice from Łojasiewicz-based convergence theorems

The bound on the rate of linear convergence given by the Łojasiewicz-based convergence theorems is

$$1 - 2\sigma\kappa\,,$$

where σ and κ are defined as in the convergence theorems. Under some simplifying assumptions, I have obtained a proof which indicates that line search should be extrapolated, and which suggests a value for the extrapolation constant.

Theorem 3.1 *Suppose that $f(\mathbf{x}) = \mathbf{x}^* A\mathbf{x} + \mathbf{b}^*\mathbf{x} + c$ and that $\mathbf{x}_{k+1}$ has the following property:*

$$f(\mathbf{x}_{k+1}) = \arg\min_{\alpha\in\mathbb{R}} f(\mathbf{x}_k + (\mathbf{x}_{k+1} - \mathbf{x}_k)\alpha)$$

Then $\sigma\kappa$ is maximized (over linear extrapolations) by replacing $\mathbf{x}_{k+1}$ by

$$\widehat{\mathbf{x}_{k+1}} = \mathbf{x}_k + \frac{4}{3}(\mathbf{x}_{k+1} - \mathbf{x}_k)$$

Some things remaining in the line of investigation are:

- Less well-behaved objective functions should be well-approximated by quadratics within some sufficiently small neighborhood. I need to formalize the following notions:
  - If f is only approximately quadratic, it doesn't move the minimizing point "too much".
  - The value of f at the extrapolated point is "sufficiently similar" to that of the quadratic approximation. This should follow from the previous notion.
- The result can't be directly applied to ALS, as ALS does not minimize along a line. More work is required.
- Greasing a full round of ALS is not quite the same as linearly extrapolating the result of the round. This needs to be addressed.

4 03/22/2016: Two steps forward, one step back

4.1 Dissertation proposal feedback

Dr. Just provided feedback on the draft of my dissertation proposal. Most of the feedback suggested small changes, but there was a suggestion which I interpret as follows:

"The proposal does not well indicate (to a non-specialist) the direction in which I want to take the research."

I think the following would fix this:

1. Add a mini-introduction and mini-conclusion to each research problem, and to the research section itself.
2. Make the introduction chapter more focused. Much of the information in the introduction is made redundant by the background, so I can remove all but the core concepts.
3. Make the proofs in the research chapter visually distinct from the rest of the writing. Though they provide evidence of capability, many of the proofs are at least a page long, and are densely packed with technical details, which would hamper "big-picture" reading.
4. Add a more detailed blurb about the reason for each theorem and lemma. Some of my current blurbs resemble "here's my result", without specifying why I wanted to find it.

4.2 Greased ALS is not as new as I thought.

On further investigation, I am forced to conclude that "greased" ALS, when considered on a second-order approximation of the objective function, is exactly a Block Successive Over-Relaxation algorithm. I may be able to extract useful comparisons from this, but the algorithm is not as new as I had hoped.

4.3 Łojasiewicz inequality

My proof of the Łojasiewicz-like inequality uses an auxiliary function

$$F(\mathbf{x}, t) \quad = \quad \lim_{s \to t} E(\mathbf{p} + s\mathbf{x})$$

which provides the limiting value of E when approaching along a first-order curve. The auxiliary function F is analytic on its domain, so I can apply the original Łojasiewicz inequality to it, and relate the gradients of F and E. This past week, I have been investigating how to deal with the case where F is undefined.

1. Decreasing the dimension of the domain of F would eventually reduce the problem to that of a bounded univariate rational function, which has a continuous extension. I have not been able to make this work, however. Though I can reformulate F such that its domain has dimension identical to that of E, this doesn't reduce the dimension of the problem.
2. A higher-order approach seems like my best bet.

$$\hat{F}(t, \mathbf{x}_0, \ldots, \mathbf{x}_n) \quad = \quad \lim_{s \to t} E\left(\sum_{i=0}^{n} s^i \mathbf{x}_i\right)$$

By requiring that $[i \neq j] \iff [\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0]$, or otherwise require linear independence, I would be able to exhaust the dimension of the domain of E. Most of the supporting theorems I created for F

apply also to $\hat{F}$, and most of the arguments from the original proof would also apply. Despite this, the open sets on which the Łojasiewicz-like inequality holds would not be cones if I repeated the previous argument. I cannot yet say anything deep about when to increase n, and I have not yet justified requiring linear independence, so I cannot guarantee that that the argument will terminate for any finite n.

5 03/29/2016: More thoughts on the Łojasiewicz inequality

5.1 Dissertation proposal feedback

I have made (almost all of) the small changes suggested by Dr. Just.

More than 50% of the original introduction was made redundant by the background chapter. I have pared the background part of the introduction down to only those concepts used heavily in my research. The introduction is now almost 50% shorter, and far more focused.

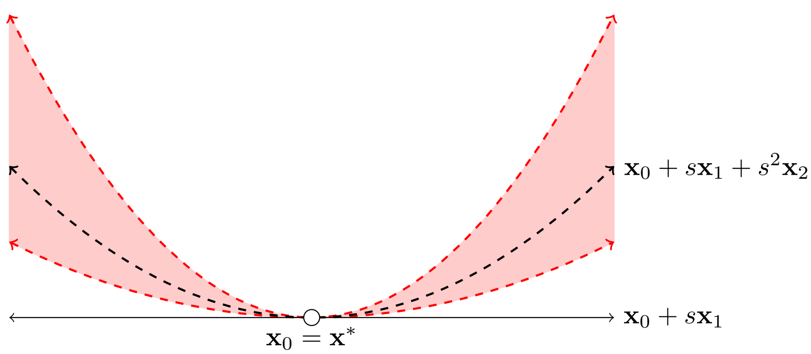
5.2 Łojasiewicz inequality

I considered this week the possibility that a previous argument of mine (the proof of the Łojasiewicz cones) would work for neighborhoods of the coefficients $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ from

$$\hat{F}(t, \mathbf{x}_0, \ldots, \mathbf{x}_n) \quad = \quad \lim_{s \to t} E\left(\sum_{i=0}^{n} s^i \mathbf{x}_i\right) .$$

This is not the case. I was able to find a counterexample and to identify the place where the previous argument would break down if applied to these neighborhoods.

I will return to my previous approach, which holds $\mathbf{x}_0, \ldots, \mathbf{x}_{n-1}$ constant, and examines neighborhoods of $\mathbf{x}_n$. In the example below, $\mathbf{x}_0$ and $\mathbf{x}_1$ are held constant, but $\mathbf{x}_2$ is allowed to vary.



I am still unable to justify requiring that the vectors $\mathbf{x}_i$ be orthogonal.

My argument for using cones surrounding $\mathbf{x}_0 + s\mathbf{x}_1$ relies on the cluster points of $\frac{\mathbf{x}_0 - \mathbf{v}_n}{\|\mathbf{x}_0 - \mathbf{v}_n\|}$, where $\mathbf{v}_n$ is a sequence of parameters produced by some tensor fitting algorithm. This relies on the tangential limit

$$\hat{F}(t, \mathbf{x}_0, \ldots, \mathbf{x}_1) \quad = \quad \lim_{s \to t} E\left(\sum_{i=0}^{1} s^i \mathbf{x}_i\right)$$

being insensitive to the norm of $\mathbf{x}_1$, so I may not be able to apply the same method for $\mathbf{x}_2$ or higher.

6 04/05/2016: A search for counterexamples

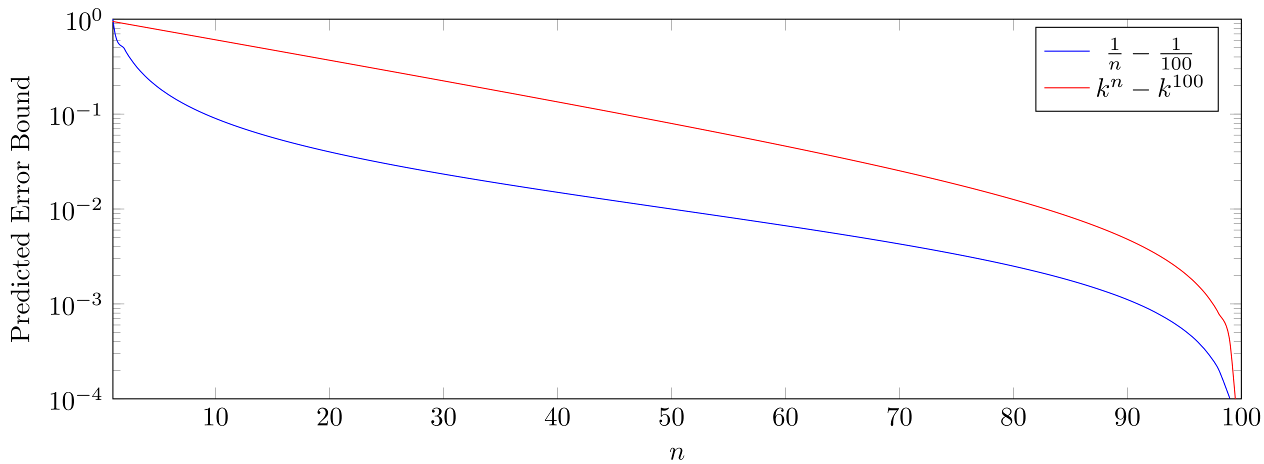
This week, I have been searching for counterexamples to my conjecture that sequences produced by ALS will converge in direction. My process is:

1. Select the size of the tensor (eg. $3 \times 3 \times 5$).
2. Select target and decomposition ranks.
3. Generate, randomly, a target and initial approximation with given ranks.
4. Record approximately 5000 iterations of ALS.
5. Compare each approximation to the final approximation. Compare norms of summands, and angles between corresponding summands.[1]

So far, this has yielded:

1. Many examples of sequences which appear to converge.
2. Several examples of rank-3 sequences for which all 2 of 3 summands appear to diverge (but appear to converge in direction)
3. One example of a rank-3 sequence for which all 3 summands appear to diverge (but appear to converge in direction). This tensor is in $\mathbb{R}^{3\times3\times5}$.
4. One example with unusually slow convergence, but for which 2^{17} iterations reveal the conjectured behavior.

If my conjecture holds, and if ALS satisfies descent conditions, then every sequence produced by ALS will eventually exhibit one of the two types of behavior, the graphs of which are below. Note that I compare each iteration to the most recent iteration, rather than to a limit point. To clarify how I look for counterexamples, I look for one of the following behaviors:



[1] Convergence in direction was also examined for each dimension individually, because that is the form of convergence I seek in my conjecture. Convergence of sequences was identical.

7 04/12/2016: Partial results in many areas, completed results in none.

This week,

1. I have partially completed slides for a proposal defense.
2. I have almost completed a proof of a supporting conjecture for my study of convergence in direction.
   - My originally intended approach for the conjecture was unnecessarily complicated. I have since found a different approach, which simplifies the proof, if not the notation, immensely.
   - Some details of the proof still elude me.
3. I have examined a portion of the literature related to Successive Over-Relaxation, and found two papers in particular which appear promising.
   - A paper by Vrahatis et. al. contains a result showing convergence of nonlinear SOR. It does not explicitly examine Block SOR.
   - A paper by Hadjidimos summarizes many convergence results for SOR and block SOR.

   The results of the two papers should hold also for greased ALS, and the conditions under which the results hold are effectively identical to the assumptions I was using in my attempts to show convergence.

   The papers examined did not contain a method for optimal selection of the extrapolation parameter, so I may still be able to introduce something new by exploiting the tensor structure. *This may already be studied in the literature. I just haven't yet found an example.*