Alternating Optimizations:
Geometric and Physical Interpretations

Rami Zamir

MIT student seminar
Wed. Sept. 25

Alternating projections between two convex sets in Euclidean space converge to the minimum distance between the sets. In the case of intersecting sets this fact follows easily from the Pythagoras Theorem (used in the form of inequality). The closest pair of points on the two sets is a unique, stable, "fixed point" of the projection mapping. If the sets are non-convex or the projections are not accurate then the alternating process may get stuck on a local minimum, converge to a biased solution, enter a limit cycle, or not converge at all.

Csiszar and Tusnady demonstrated that alternating minimizations of the divergence (or Kullback Leibler distance) between sets of probability distributions behave in a similar way; convexity of the distribution sets guarantees convergence to the closest distributions. The key element in the proof of convergence is the ``Pythagoras theorem for divergence'' (Csiszar's inequality). Their work suggests a unified view of few well known iterative optimization algorithms: EM for maximum-likelihood estimation, and Arimoto-Blahut for rate-distortion and capacity computation.

Richardson develops a geometric picture for the well known Turbo decoding algorithm. His aim is to investigate the convergence properties of the algorithm and its proximity to the desired maximum likelihood solution. Unlike the problems above, here the geometric picture is not as simple; that is, each iteration step is not directly related to the true likelihood and therefore cannot be regarded as an exact projection. Nevertheless he succeeds to prove the existence of fixed point(s) and draw some conclusions about stability, uniqueness and proximity to maximum likelihood.

Richardson's geometric picture does not quite parallel that of Csiszar and Tusnady, yet in his case the physical interpretation of the iterations is clear. In the seminar I will highlight the other side of the story: can we associate physical meaning to the intermediate quantities in the Arimoto-Blahut algorithm?