WEEKLY PARTICIPATION 4: PROXIMAL GRADIENT DESCENT

Proximal gradient descent (aka composite gradient descent, generalized gradient descent, or the prox-linear algorithm) is used to solve problems of the form

$$\text{(Comp)} \qquad \min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x})$$

where f is a differentiable convex function and g is a convex function. This method is particularly useful when g is not smooth. To move from iterate $\mathbf{x}_t$ to $\mathbf{x}_{t+1}$, the method takes a gradient descent step of f, then chooses $\mathbf{x}_{t+1}$ to balance between minimizing $g(\mathbf{x})$ and staying close to the intermediate iterate:

$$\text{(Prox)} \qquad \mathbf{x}_{t+1} = \text{argmin}_{\mathbf{x}} \frac{1}{2\alpha_t} \|\mathbf{x} - (\mathbf{x}_t - \alpha_t \nabla f(\mathbf{x}_t))\|^2 + g(\mathbf{x}).$$

Note that (Prox) is always strictly convex, so $\mathbf{x}_{t+1}$ is always a well-defined single point.

Assume g is smooth. Argue that if $\mathbf{x}^\star$ is a fixed point of (Prox) then it is a minimizer of (Comp).

Discussion. The fixed points of the proximal gradient descent process are minimizers of the original composite optimization problem even if g is non-smooth. In fact this is one case where the proximal gradient algorithm is frequently employed. A canonical example of this usecase is the ISTA algorithm for the LASSO problem – you will see this example on your next homework. Prox-gradient updates of the form (Prox) also play a key role in the ADMM method for distributed optimization that we will see later in the course.