1. Let \( f : \mathbb{R}^d \to \mathbb{R} \). Assume that for any \( x \) there exists \( g_x \in \mathbb{R}^d \) such that for all \( y \) one has \( f(x) - f(y) \leq g_x \cdot (x - y) \). Show that \( f \) is convex. Furthermore if \( f \) is differentiable, show that one can take \( g_x = \nabla f(x) \).

2. Prove that \( X \mapsto -\log \det(X) \) is convex on the set of positive definite matrices. Hint: to exhibit a subgradient at \( x = X \), consider some other point \( y = X + Y \) and write it as \( X^{1/2} (I + X^{-1/2}YX^{-1/2})X^{1/2} \) (recall also the AM-GM inequality).

3. For a convex set \( K \subset \mathbb{R}^d \) we define the normal cone of \( K \) at \( x \in K \) by \( N_K(x) = \{ \theta \in \mathbb{R}^d : \forall y \in K, \theta \cdot (y - x) \leq 0 \} \). What are the normal cones of the Euclidean ball? Draw a picture for a 2-dimensional polytope.

4. Show that for a convex function one has \( x = \arg\min_{x' \in K} f(x') \iff \nabla f(x) + \lambda = 0 \) for some \( \lambda \in N_K(x) \).

5. Let \( x \in \mathbb{R}^d \) and \( y \) be its Euclidean projection on \( K \). Show that for any \( z \in K \), one has \( \|y - z\| \leq \|x - z\| \).

6. Consider a polytope written as \( K = \{ x : a_j \cdot x \leq b_j \forall j \in [m] \} \). Show that \( N_K(x) = \{ \sum_j \alpha_j a_j : \alpha_j \geq 0 \} \).

7. Show that the algorithm \( x_{t+1} = \arg\min_{x \in \text{span}(\nabla f(x_0),...,\nabla f(x_t))} f(x) \) always terminate in at most \( d \) steps.

8. Consider a quadratic function \( f(x) = \frac{1}{2} x^\top Ax + b^\top x \). Derive a simple formula for the update of the above algorithm (Hint: you may want to work with the inner product \( \langle x, y \rangle_A = x^\top Ay \), and first note that the updates \( x_{t+1} - x_t \) form an orthogonal basis for that inner product).

9. Prove that for a smooth function \( f \), with a noisy oracle such that \( \mathbb{E}[\|g - \nabla f(x)\|^2] \leq \sigma^2 \) one has that the rate of gradient descent is of order \( \frac{\sigma^2}{\tau} + \frac{\sigma}{\sqrt{\tau}} \).

10. How can the above result be used in a machine learning setting? Think about doing mini-batches (to leverage parallel processing).