EE364b Homework 2

1. **Subgradient optimality conditions for nondifferentiable inequality constrained optimization.** Consider the problem

$$\text{minimize } f_0(x)$$
$$\text{subject to } f_i(x) \leq 0, \quad i = 1, \ldots, m,$$

with variable $x \in \mathbb{R}^n$. We do not assume that $f_0, \ldots, f_m$ are convex. Suppose that $\tilde{x}$ and $\tilde{\lambda} \succeq 0$ satisfy primal feasibility,

$$f_i(\tilde{x}) \leq 0, \quad i = 1, \ldots, m,$$

dual feasibility,

$$0 \in \partial f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \partial f_i(\tilde{x}),$$

and the complementarity condition

$$\tilde{\lambda}_i f_i(\tilde{x}) = 0, \quad i = 1, \ldots, m.$$

Show that $\tilde{x}$ is optimal, using only a simple argument, and definition of subgradient. Recall that we do not assume the functions $f_0, \ldots, f_m$ are convex.

2. **Optimality conditions and coordinate-wise descent for $\ell_1$-regularized minimization.** We consider the problem of minimizing

$$\phi(x) = f(x) + \lambda \|x\|_1,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable, and $\lambda \geq 0$. The number $\lambda$ is the regularization parameter, and is used to control the trade-off between small $f$ and small $\|x\|_1$. When $\ell_1$-regularization is used as a heuristic for finding a sparse $x$ for which $f(x)$ is small, $\lambda$ controls (roughly) the trade-off between $f(x)$ and the cardinality (number of nonzero elements) of $x$.

(a) Show that $x = 0$ is optimal for this problem (i.e., minimizes $\phi$) if and only if $\|\nabla f(0)\|_\infty \leq \lambda$. In particular, for $\lambda \geq \lambda_{\text{max}} = \|\nabla f(0)\|_\infty$, $\ell_1$ regularization yields the sparsest possible $x$, the zero vector.

**Remark.** The value $\lambda_{\text{max}}$ gives a good reference point for choosing a value of the penalty parameter $\lambda$ in $\ell_1$-regularized minimization. A common choice is to start with $\lambda = \lambda_{\text{max}}/2$, and then adjust $\lambda$ to achieve the desired sparsity/fit trade-off.
(b) **Coordinate-wise descent.** In the coordinate-wise descent method for minimizing a convex function $g$, we first minimize over $x_1$, keeping all other variables fixed; then we minimize over $x_2$, keeping all other variables fixed, and so on. After minimizing over $x_n$, we go back to $x_1$ and repeat the whole process, repeatedly cycling over all $n$ variables.

Show that coordinate-wise descent fails for the function $g(x) = |x_1 - x_2| + 0.1(x_1 + x_2)$.

(In particular, verify that the algorithm terminates after one step at the point $(x_2^{(0)}, x_2^{(0)})$, while $\inf_x g(x) = -\infty$.) Thus, coordinate-wise descent need not work, for general convex functions.

(c) Now consider coordinate-wise descent for minimizing the specific function $\phi$ defined above. Assuming $f$ is strongly convex (say) it can be shown that the iterates converge to a fixed point $\bar{x}$. Show that $\bar{x}$ is optimal, i.e., minimizes $\phi$.

Thus, coordinate-wise descent works for $\ell_1$-regularized minimization of a differentiable function.

(d) Work out an explicit form for coordinate-wise descent for $\ell_1$-regularized least-squares, i.e., for minimizing the function

$$\|Ax - b\|_2^2 + \lambda\|x\|_1.$$  

You might find the deadzone function

$$\psi(u) = \begin{cases} 
  u - 1 & u > 1 \\
  0 & |u| \leq 1 \\
  u + 1 & u < -1 
\end{cases}$$

useful. Generate some data and try out the coordinate-wise descent method. Check the result against the solution found using CVX, and produce a graph showing convergence of your coordinate-wise method.