1 Review Concepts

Decide whether the following claims are true or false. Assume functions are twice continuously differentiable unless otherwise stated.

**Unconstrained Optimization**

*Optimality Conditions*

(01) For an unconstrained problem, if $x^*$ satisfies $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succeq 0$, then $x^*$ is a local minimum.

(02) The converse of the previous claim.

(03) $\nabla f(x^*) = 0$ is a necessary and sufficient condition for $x^*$ to a strict local minimum if $f$ is convex.

(04) For the quadratic $\frac{1}{2}x^T Q x - b^T x$, $Q \succ 0$, $Q^{-1} b$ is the global minimum.

(05) If instead $Q \succeq 0$ in the previous claim, then it is possible to have zero solutions or infinite solutions.

(06) If the feasible set of an optimization problem is closed, then there exists a global minimum.

**Gradient Methods**

*dbbrown@mit.edu*
(07) Our convergence results refer only to convergence towards local minima, but not maxima.

(08) Our convergence results refer only to convergence towards local optima.

(09) In order to prove a gradient method converges, it is crucial to show that the sequence of direction vectors is gradient related.

(10) A gradient related method using the Armijo or minimization stepsize rules is guaranteed to converge to a stationary point.

(11) The same applies if we use a constant or diminishing stepsize, but the rate of convergence is typically much slower.

(12) The condition number of the Hessian near a stationary point is a good estimate of the asymptotic convergence rate of the steepest descent method.

*Newton’s Method*

(13) Newton’s method has a quadratic rate of convergence.

(14) Since we minimize the second-order Taylor series at each step in Newton’s method, and, close to a stationary point the second-order Taylor series approximates the function well, it is impossible for (pure) Newton’s method to converge to a local maximum.

(15) If a problem is ill-conditioned, we would be wise to scale it before using Newton’s method.

(16) If we use a line minimization stepsize rule in conjunction with Newton’s method, then there is no need to correct the Hessian when it is not positive definite (e.g., with modified Cholesky factorization).

*Least-Squares Problems*

(17) The gradient of $g : \mathbb{R}^n \mapsto \mathbb{R}^m$ is an $n \times m$ matrix.

(18) The gradient of $1/2\|g(x)\|^2$ (where $g$ is as in the previous claim) is an $n$-vector written as $\nabla g(x)g(x)$. 

\[ \]
The Gauss-Newton method is computationally more involved than Newton’s method, but it tends to converge in fewer iterations for least-squares problems.

If you do not understand extended Kalman filters, you are in trouble for this exam.

Conjugate Direction Methods

The conjugate gradient method minimizes a strictly convex quadratic function in no more than $k$ steps, where $k$ is the number of distinct eigenvalues of the Hessian.

We converge faster in the previous claim if we invoke the preconditioned conjugate gradient method (i.e., we use scaling).

Optimization Over a Convex Set

Optimality Conditions

$\nabla f(x^*)^T(x - x^*) \geq 0 \forall x \in X$ where $f$ is convex and $X$ is convex holds if and only if $x^*$ is a global minimum of the problem.

The problem $\min f(x)$ subject to $h_i(x) = 0$, $i = 1, \ldots, m$ and $g_j(x) \geq 0$, $j = 1, \ldots, r$ is convex if $f$ and $g_j$ are convex and $h_i$ is affine.

If $X$ is a nonempty, closed subset of $\mathbb{R}^n$, then $\forall z \in \mathbb{R}^n$, $[z]_+$ is unique.

Feasible Direction Methods

The convergence results for feasible direction methods are completely analogous to the gradient descent convergence results from unconstrained optimization.

The conditional gradient method with a minimization stepsize would solve a linear program (assuming it has an optimal solution) in a single step.

The gradient projection method is computationally more expensive per iteration than the conditional gradient method.

Two-metric projection methods are computationally more expensive per iteration than the gradient projection method.
Affine scaling is a technique for solving linear programs.

Lagrange Multiplier Theory

**Necessary and Sufficient Conditions for Equality Constraints**

(31) If $x^*$ is a local minimum of $f$ subject to $h(x) = 0$ and a regular point, then there exists a unique vector $\lambda^*$ such that

\[
\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0,
\]

\[
y^T \left( \nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*) \right) y \geq 0 \quad \forall y \in V(x^*),
\]

where $V(x^*) = \{ y | \nabla h_i(x^*)^T y = 0, \quad i = 1, \ldots, m \}$.

(32) The sufficiency condition is just the converse of the previous statement (with the second-order condition being strict).

(33) The sensitivity theorem requires regularity of the unperturbed optimal solution $x^*$.

**Inequality Constraints**

(34) If $x^*$ is a local minimum of $f$ subject to $h_i(x) = 0, \quad i = 1, \ldots, m$, $g_j(x) \leq 0, \quad j = 1, \ldots, r$, and a regular point, then there exist unique vectors $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}^r$ such that

\[
\nabla_x L(x^*, \lambda^*, \mu^*) = 0,
\]

\[
\mu_j = 0 \quad \forall j \notin A(x^*),
\]

\[
y^T \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) y \geq 0, \quad \forall y \in V(x^*),
\]

where $V(x^*) = \{ y | \nabla h_i(x^*)^T y = 0, \quad i = 1, \ldots, m \}$ and $A(x^*)$ is the set of indices of active constraints at $x^*$.

(35) The KKT sufficiency condition is the same as the KKT necessity condition except that we have strict inequalities and we do not require regularity.
A general sufficiency condition states that if $x^*$ is feasible, $\mu_j^* \geq 0$, $\mu_j = 0 \ \forall \ j \notin A(x^*)$, and $x^* = \arg\min_{x \in X} L(x, \mu^*)$, then $x^*$ is a global minimum.

A corollary to the above statement is that, if $f$ and $g$ are convex (and $X = \mathbb{R}^n$), we can replace the last condition with $\nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*) = 0$.

If the constraint inequalities are all affine, then we do not require regularity of $x^*$ in the necessity condition.

If we have an optimal solution to minimizing a convex function over a convex set, then the corresponding dual problem has an optimal solution, and the primal and dual optimal values are equal.

The dual function (for primal minimization problems) $q(\mu)$ is a convex function of $\mu$.

2 Examples

Example 2.1. Equality-constrained least-squares.
(Taken from [1], chapter 5). Compute the optimal solution to

$$\begin{align*}
\text{minimize} & \quad \|Ax - b\|_2^2 \\
\text{subject to} & \quad Gx = h,
\end{align*}$$

where $A \in \mathbb{R}^{m \times n}$ has rank $n$ and $G \in \mathbb{R}^{p \times n}$ has rank $p$.

Example 2.2. Robust LP.
(Modified from [1], chapter 5). Consider an LP of the form

$$\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax \leq b.
\end{align*}$$

Assume now that the data matrix $A \in \mathbb{R}^{m \times n}$ is not completely known. In particular, assume for each row $a_i$ of $A$ we have polyhedral uncertainty; that is, we know only that $a_i$ belongs to some known polyhedron:

$$a_i \in \{y \mid D_i y \leq d_i\} \quad \forall i = 1, \ldots, m.$$  

The robust counterpart of the LP above is the minimum value of $c^T x$ over all $x$ such that $x$ is feasible for all possible data in the uncertainty set. Show that the robust counterpart may be formulated as another LP.
References