Optimization and Numerical Linear Algebra Qualifying Exam Study Sheet

Trevor Loe

January 22, 2025

This document contains a brief overview of all the topics that I have found may appear on the optimization and numerical linear algebra qualifying exam. I have tried to organize the topics by similarity. The list of topics was started based off the UCLA qual website `https://ww3.math.ucla.edu/qualifying-exam-dates/` and appended as past qual problems were done.

Past ONLA qualifying exams can be found at `https://ww3.math.ucla.edu/past-qualifying-exams/`

Basic optimization theory material was taken from [1] and [2]

1 Optimization

1.1 Unconstrained Optimization

1.1.1 Differentiable f

Let $f : \mathbb{R}^n \to \mathbb{R}$ be C^1. Then the **first order necessary condition** is
If x^* is a local minimizer for f over $\Omega \subset \mathbb{R}^n$, then, for any feasible direction d, we have

$$d^T \nabla f(x^*) \geq 0$$

If $\Omega = \mathbb{R}^n$ then every direction is feasible and the condition becomes

$$\nabla f(x^*) = 0$$

For $f \in C^2$, the **Second order necessary conditions** state
If x^* is a local minimizer for f over $\Omega \subseteq \mathbb{R}^n$, then for any feasible direction d, whenever

$$d^T \nabla f(x^*) = 0$$

We have

$$d^T F(x^*) d \geq 0$$

where F is the Hessian of f. When $\Omega = \mathbb{R}^n$ this becomes that $F(x^*)$ is positive semi-definite.
For the case where $\Omega = \mathbb{R}^n$ we have the **Second order sufficient condition** which states
For $f \in C^2$ if

- $\nabla f(x^*) = 0$
- $F(x^*) > 0$, that is F is positive definite

then x^* is a strict local minimizer for f.

1.2 Constrained Optimization

For $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, and $g : \mathbb{R}^n \to \mathbb{R}^p$

$$\min f(x)$$

$$\text{subject to } h_i(x) = 0, \quad i = 1, ..., m$$

$$g_j(x) \leq 0, \quad j = 1, ..., p$$

1.2.1 Lagrange Conditions

If we have no inequality conditions, so we just have f and h, then **Lagrange's Theorem** states
Let x^* be a local minimizer of f subject to $h(x) = 0$ with $h : \mathbb{R}^n \to \mathbb{R}^m$. Assume that x^* is a regular point for h, that is $\nabla h_1(x^*), ..., \nabla h_m(x^*)$ are linearly independent. Then there exists $\lambda^* \in \mathbb{R}^m$ such that

$$Df(x^*) + (\lambda^*)^T Dh(x^*) = 0$$

1.2.2 Karush-Kuhn-Tucker (KKT) Conditions

Let $f, g, h \in C^1$. Let x^* be a regular point minimizing f subject to $h(x) = 0, g(x) \leq 0$. Then there exists $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that

- $\mu^* \geq 0$
- $Df(x^*) + (\lambda^*)^T Dh(x^*) + (\mu^*)^T Dg(x^*) = 0$
- $(\mu^*)^T g(x^*) = 0$

The second order KKT conditions can be formulated by defining

$$T(x^*) = \{y \in \mathbb{R}^n : Dh(x^*)y = 0, Dg_j(x^*)y = 0, j \in J(x^*)\}$$

where J is the set of indices representing active contraints at x^*. Also define

$$L(x, \lambda, \mu) = F(x) + \lambda_1 H_1(x) + ... + \lambda_m H_m(x) + \mu_1 G_1(x) + ... + \mu_p G_p(x)$$

where H_i is the Hessian of h_i, etc. Then we have
If $f, g, h \in C^2$ and x^* is a local minimizer that is also regular then there exists λ^* and μ^* satisfying the first order KKT conditions and also, for all $y \in T(x^*)$ we have

$$y^T L(x, \lambda^*, \mu^*)y \geq 0$$

The sufficient conditions can be obtained by defining

$$T(x^*, \mu^*) = \{y : Dh(x^*) = 0, Dg_i(x^*) = 0, i \in J(x^*, \mu^*)$$

where $J(x^*, \mu^*) = \{i : g_i(x^*) = 0, \mu_i^* > 0\}$. Then we have
If $f, g, h \in C^2$, and there is a feasible point x^* along with vectors λ^*, μ^* such that

- $\mu^* \geq 0$
- $Df(x^*) + (\lambda^*)^T Dh(x^*) + (\mu^*)^T Dg(x^*) = 0$
- $(\mu^*)^T g(x^*) = 0$
- For all $y \in T(x^*, \mu^*)$ with $y \neq 0$ we have $y^T L(x, \lambda^*, \mu^*)y \geq 0$

Then x^* is a strict local minimizer for the above problem.

1.2.3 Lagrange Duel Problem

Similarly, define the Lagrangian

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x)$$

Then the Lagrange duel function is given by

$$\mathcal{G}(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu)$$

Then if $\lambda \geq 0$ we have

$$\mathcal{G}(\lambda, \mu) \leq p^* = f(x^*)$$

The lagrange duel problem is

$$\max_{\lambda, \mu} \mathcal{G}(\lambda, \mu)$$

to find the greatest lower bound on the optimal value p^*. This will be a convex problem, even if the original optimization of f is not.

2 Iterative Methods

2.1 Iterative Methods for Nonlinear Optimization

2.1.1 The Gradient Descent Algorithm

In order to find a minima of the function $f \in C^1$ we compute the iterations in the following way

$$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f(x^{(k)})$$

The **steepest descent** algorithm uses the step size chosen so

$$\alpha_k = \text{argmin}_{\alpha>0} f(x^{(k)} - \alpha \nabla f(x^{(k)}))$$

The Baillon-Haddad Theorem: If f is a convex function, it is L-Lipschitz differentiable, if and only if

$$||\nabla f(x) - \nabla f(y)||^2 \leq L\langle \nabla f(x) - \nabla f(y), x - y\rangle$$

for all x, y
Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable, convex, and L-Lipschitz differentiable. Then gradient descent with fixed step size $t < 1/L$ satisfies

$$f(x^k) - f(x^*) \leq \frac{||x^0 - x^*||^2}{2tk}$$

Proof. First note that $\nabla f(x)$ is Lipshitz with constant L. So for any x, y, z we have

$$||\nabla^2 f(z)(x - y)|| \leq L||x - y||$$

and similarly

$$(x - y)^T \nabla^2 f(z)(x - y) \leq L||x - y||^2$$

Taylor expanding f gives us

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(\xi)(y - x)$$

$$\leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}||y - x||^2$$

If we let $y = x - t\nabla f(x)$ then we get

$$f(y) \leq f(x) + \nabla f(x)^T (-t\nabla f(x)) + \frac{L}{2}|| - t\nabla f(x)||^2$$

$$= f(x) + (\frac{Lt^2}{2} - t)||\nabla f(x)||^2$$

As $t < 1/L$ we have

$$\frac{Lt^2}{2} - t = \left(\frac{Lt}{2} - 1\right) t$$

and $Lt/2 < 1/2$ so $Lt/2 - 1 < 1/2$. So we get

$$f(y) \leq f(x) + \frac{t}{2}||\nabla f(x)||^2$$

We can similarly argue that $Lt/2 - 1 \leq -\frac{1}{2}$. As f is convex we have

$$f(x) \leq f(x^*) + \nabla f(x)^T (x - x^*)$$

Plugging this into the other equation yields

$$f(y) \leq f(x^*) + \nabla f(x)^T (x - x^*) - \frac{t}{2}||\nabla f(x)||^2$$

So

$$f(y) - f(x^*) \leq \frac{1}{2t}\left(2t\nabla f(x)^T(x - x^*) - t^2||\nabla f(x)||^2\right)$$
$$= \frac{1}{2t}\left(2t\nabla f(x)^T(x - x^*) - t^2||\nabla f(x)||^2 - ||x - x^*||^2 + ||x - x^*||^2\right)$$

Observing that

$$2t\nabla f(x)^T(x - x^*) - t^2||\nabla f(x)||^2 - ||x - x^*||^2 = -||x - 2t\nabla f(x) - x^*||^2 = -||y - x^*||^2$$

We get

$$f(y) - f(x^*) \leq \frac{1}{2t}\left(||x - x^*||^2 - ||y - x^*||^2\right)$$

As this will true for any iteration of the gradient descent method, we get

$$\sum_{i=1}^{k} f(x^{(k)}) - f(x^*) \leq \frac{1}{2t}\sum_{i=1}^{k}\left(||x^{(i-1)} - x^*||^2 - ||x^{(i)} - x^*||^2\right)$$

Note that each iteration, the value of f is decreasing. Thus

$$f(x^{(k)}) - f(x^*) \leq \frac{1}{k}\sum_{i=1}^{k} f(x^{(k)}) - f(x^*) \leq \frac{1}{2tk}\sum_{i=1}^{k}\left(||x^{(i-1)} - x^*||^2 - ||x^{(i)} - x^*||^2\right)$$

Note that the series on the right is telecoping, resulting in

$$||x^0 - x^*||^2 - ||x^{(k)} - x^*||^2 \leq ||x^0 - x^*||^2$$

So we have

$$f(x^{(k)}) - f(x^*) \leq \frac{1}{2tk}||x^0 - x^*||^2$$

□

2.1.2 Newton's Method

Newton's method for a function $f \in C^2$ is given by

$$x^{(n+1)} = x^{(n)} - F(x^{(n)})^{-1}\nabla f(x^{(n)})$$

Note that F must be positive semi-definite, otherwise Newton's method will be driven towards a maximizer rather than a minimizer.
If the function is well-behaved in the interval around the optimum and the initial guess is chosen sufficiently close, then Newton's method will converge quadratically.

The Levenberg-Marquardt modification to Newton's method is

$$x^{(k+1)} = x^{(k)} - (F(x^{(k)}) + \mu_k I)^{-1}\nabla f(x^{(k)}$$

3 Convex Functions

A function is convex if its domain is a convex set and for any $x, y \in \text{dom}(f)$ and $\lambda \in [0, 1]$ we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

The epigraph of a function

$$\text{epi}(f) = \{(x, t) \in \text{dom}(f) \times \mathbb{R} : f(x) \leq t\}$$

We have that a function is convex if and only if its epigraph is a convex set.

3.1 Subgradients

A vector g is a subgradient of a convex function f at x_0 if, for any $y \in \text{dom}(f)$ we have

$$f(y) \geq f(x_0) + \langle g, y - x_0 \rangle$$

The set of subgradients of f is called the subdifferential of f at x_0, denoted $\partial f(x_0)$.
The first-order optimality condition in terms of subgradients is x^* is a minimizer for a convex function f if and only if $0 \in \partial f(x_0)$.

References

[1] E.K.P. Chong and S.H. Zak. *An Introduction to Optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimi. Wiley, 2004. ISBN: 9780471654001. URL: `https://books.google.com/books?id=bKdibaXVuUEC`.

[2] Yurii Nesterov. *Lectures on Convex Optimization*. Springer Optimization and Its Applications. Springer Cham, 2018. ISBN: 978-3-319-91578-4. DOI: `https://doi.org/10.1007/978-3-319-91578-4`. URL: `https://link.springer.com/book/10.1007/978-3-319-91578-4`.