EE227BT: Convex Optimization

September 9, 2014

1 Lecture 1—8/28/2014

Def: A set C in $\mathbb{R}^n$ is convex if for any $X, y \in C$, the line segment $\theta x + (1-\theta)y(\theta \geq 0)$ also lies in C

Convex: $\theta_1 x + \theta_2 y \in C$, where $\theta_1, \theta_2 \geq 0$, and $\theta_1 + \theta_2 = 1$ Linear: restrictions on θ_1, θ_2 are dropped Conic: if restriction $\theta_1 + \theta_2 = 1$ is dropped

$S^n_+ = \{X positive semidefinite\}$

Is a convex cone, because $V^T Z V \geq 0 \forall V$

Theorem: C_1, C_2 are convex, $C_1 \cap C_2$ is also convex

Convex hull is the intersection of all convex sets that contain a point

2 Lecture 2—9/2/2014

epigraph: $epi f => \{(x,t), t \geq f(x)\}$

$epi f$ is convex if f is convex, and vv

composition with affine functions: if we have a convex function, and we compose it with a linear/affine function, the resulting function composition will be convex

adding convex functions will maintain convexity, subtracting will not necessarily how do you prove convexity with subtraction?

$$\hat{f}(y) -> f(Ay+b) \nabla \hat{f}(y) = A^T \nabla f(Ay+b) \nabla^2 \hat{f}(y) = A^T \nabla^2 f(Ay+b) A$$

2.1 Fenchel Conjugate

Fenchel Conjugate: take a function, make another function that is always convex

Let $||.||$ be a norm on $\mathbb{R}^n$. Its dual norm is $||u||_* := \sup\{u^T x | ||x|| \leq 1\}$

The Fenchel Conjugate of a function f is:

$f^*(z) = \sup_{x \in f} x^T z - f(x)$

indicator function is convex iff set is convex

If $f(x)$ is convex (and a technical condition is met), $f^{**} = f(x)$

2.2 Subdifferentials

First order global under estimator: $f(x) \geq f(y) + < g, x - y >$

Line supporting a graph: whole epigraph is on one side of the line, line passes through the point of interest

If function is differentiable, there is a single line (slope matches the gradient), if the function is undifferentiable, there are ¿¿ 1 functions that can support the graph

g is a valid sub gradient if

$$x \in \mathbb{R}^n g \in \mathbb{R}^n f(y) \geq f(x) + g^T(y-x)$$

$$x -> x^T S x G = x x^T (x^T S x) = (S x x^T)$$

3 Lecture 3—9/4/2014

3.1 Homework hints

- Sum of k largest eigenvalues of a matrix $\rightarrow$ not obvious to prove that it is convex
  - Know that the largest eigenvalue is convex
  - If you have a matrix S, you can add a positive semidefinite matrix, which gives another symmetric matrix; how do the eigenvalues change $\rightarrow$ they increase
  - Theorem $\rightarrow$ minimax representation of eigenvalues

3.2 Optimization problem

Let $f_i : \mathbb{R}^n \rightarrow \mathbb{R}(0 \leq u \leq m)$, generic nonlinear program is:

$$\min f_0(x) s.t. f_i(x) \leq 0, 1 \leq i \leq m x \in \{f_0 \cap \ldots \cap f_m\}$$

Drop condition on domains for brevity.

- If f_i are differentiable $\rightarrow$ smooth optimization

Convex optimization is specifically:

$$\begin{gathered}\min f_0(x)\\ s.t. f_i(x) \leq 0, 1 \leq i \leq m\\ Ax = b\end{gathered}$$

Observe:

- All f_i are convex
- Direction of inequality $f_i(x) \leq 0$ is critical
- Only allow affine equality $\rightarrow$ affine functions are both convex and concave

We denote the feasible set $\mathcal{X}$:

$$\mathcal{X} = x \in \mathbb{R}^n | f_i(x) \leq 0, 1 \leq i \leq m, Ax = b$$

The optimal value p^* is defined by:

$$p^* := \inf f_0(x) | x \in \mathcal{X}$$

IFF $\mathcal{X} =$, problem is infeasible, $p^* = +\infty$ If $p^* = -\infty$, we call problem *unbounded below*

3.3 Optimality

A point $x^* \in \mathcal{X}$ is *locally optimal* if $f(x^*) \geq f(x) \forall x$ in a neighborhood of x^*. x^* is *globally optimal* if $f(x^*) \geq f(x) \forall x \in \mathcal{X}$.

Only allow a single constraint; can turn multiple constraints, e.g.:

$$\begin{gathered}f_1(x) \leq 0\\ f_2(x) \leq 0\end{gathered}$$

Into a single constraint via:

$$\max_x f_1(x), f_2(x) \leq 0$$

We are at a stable point if $f(x) = 0$. If we are convex, x is a global minimum. If we are convex and at a non differentiable point, we are at a minimum if 0 is a valid subgradient (e.g., $0 \in \delta f(x)$).

Subgradient:

$g \in \mathbb{R}^n$ is a sub gradient of $f()x$ if $\forall y, f(y) \geq f(x) + g^T(y - x)$.

3.4 Monotonic Transformation

Say $\phi_0, \mathbb{R} \to \mathbb{R}$ is monotonically increasing, ϕ_i satisfies $\phi_i(u) \leq 0, \iff u \leq 0$, can transform to convex via composition.

E.g., $f(x_1, x_2, x_3) = x_1^{1.5} x_2^{-2} x_3^3$ is non-convex. However, $\log f(x_1, x_2, x_3)$ is convex ($f(x_1, x_2, x_3) \to 1.5y_1 - 2y_2 + 3y_3, x_i = logy_i$).

3.5 Slack variables

Turn inequalities into equalities $\to Ax \geq b \to Ax + s = b, s \geq 0$. Useful if you can use variables to reduce the dimension of the problem.

4 Lecture 4—9/9/2014

Quiz content: Quizzes will not be as hard as homework, but expect very strong linear algebra background. E.g., use SVD to solve a matrix optimization problem, solve QP unconstrained.

4.1 Linear Programming

$\min c^T x$
$s.t. b - Ax \in T\mathbb{R}_+^n$

Algorithm for solving LP in polynomial time ($O(n^3)$) found in 1970's.

Can generalize to:

$$b - Ax \in K$$

For K being a non-linear convex cone.

LP can be used to optimize non-linear objective functions, e.g.:

$\min ||Ax - b||_1, x \in \mathbb{R}^n$
$\min \sum_i |a_i^T x - b_i|, x \in \mathbb{R}^n$
$\min \sum_i t_i, |a_i^T x - b_i| \leq t_i$
$\min 1^T t, -t_i \leq a_i^T x - b_i \leq t_i$

Augmenting number of variables can help solve problems, as we can "refactor" problems into LP. LP can be used to minimize any polyhedral function; polyhedral being max of sum of affine.

4.2 Quadratic Programming

Like linear programming, but with a quadratic term. E.g.,

$$\min \frac{1}{2} x^T Ax + b^T x + c, s.t. Gx \leq h$$

Can apply to nonnegative least squares (NNLS), regularized least-squares (RLS).

Transform Lasso to QP:

$$\min \frac{1}{2} ||Ax - b||_2^2 + \lambda ||x||_1$$

Translate variables:

$$\min_z z^T Qx + q^T z, Gz \leq d$$

A is positive semidefinite.

If $\lambda = 0$,

$$Q = \frac{1}{2} A^T A$$

$$q = A^T b$$

If $\lambda \neq 0$,

$$\min_{z,t} z^T Q x + q^T z + \sum_i t_i, Gz \leq d$$

$$-x_i \leq t_i \leq x_i$$

QP is a slight generalization of LP.

4.3 Second Order Cone Program

A variation of LP, with much more expressive power:

$$\min f^T x$$

$$s.t., ||A_i x + b_i||_2 \leq c_i^T x + d_i, i \in 1, \ldots, m$$

4.4 Semidefinite Problem

$$\min_{x \in \mathbb{R}^n} c^T x$$

$$s.t. A(x) = A_0 + x_1 A_1 + \ldots + x_n A_n \succeq 0$$