CSCI7000-016: Optimization and Control of Networks

Review on Convex Optimization

Convex set

 $S \subset \mathcal{R}^n$ is convex if

$$x, y \in S, \ \lambda, \mu \ge 0, \ \lambda + \mu = 1 \quad \Rightarrow \quad \lambda x + \mu y \in S$$

geometrically: $x, y \in S \implies$ line segment through $x, y \in S$

examples (one convex, two nonconvex sets):



Hyperplanes and halfspaces

hyperplane: set of the form $\{x \mid a^T x = b, a \neq 0\}$

halfspace: set of the form $\{x \mid a^T x \leq b, a \neq 0\}$



- *a* is the normal vector
- hyperplanes and halfspaces are convex

Euclidean balls and ellipsoids

Euclidean ball with center x_c and radius r:

$$B(x_c, r) = \{x \mid ||x - x_c||_2 \le r\} = \{x_c + ru \mid ||u||_2 \le 1\}$$

ellipsoid: set of the form

$$\{x \mid (x - x_c)^T P^{-1} (x - x_c) \le 1\}$$

with $P \in \mathbf{S}_{++}^n$ (*i.e.*, P symmetric positive definite)



other representation: $\{x_c + Au \mid ||u||_2 \leq 1\}$ with A square and nonsingular

Polyhedra

solution set of finitely many linear inequalities and equalities

$$Ax \leq b, \qquad Cx = d$$

 $(A \in \mathbf{R}^{m \times n}, C \in \mathbf{R}^{p \times n}, \preceq \text{ is componentwise inequality})$



polyhedron is intersection of finite number of halfspaces and hyperplanes

Convex functions

 $f: \mathbf{R}^n \to \mathbf{R}$ is convex if $\operatorname{\mathbf{dom}} f$ is a convex set and

$$f(\theta x + (1 - \theta)y) \le \theta f(x) + (1 - \theta)f(y)$$

for all $x, y \in \operatorname{\mathbf{dom}} f$, $0 \le \theta \le 1$



- f is concave if -f is convex
- f is strictly convex if $\operatorname{dom} f$ is convex and

$$f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)$$

for $x, y \in \operatorname{\mathbf{dom}} f$, $x \neq y$, $0 < \theta < 1$

Examples on R

convex:

- affine: ax + b on **R**, for any $a, b \in \mathbf{R}$
- exponential: e^{ax} , for any $a \in \mathbf{R}$
- powers: x^{α} on $\mathbf{R}_{++},$ for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on ${\bf R},$ for $p\geq 1$
- negative entropy: $x \log x$ on \mathbf{R}_{++}

concave:

- affine: ax + b on **R**, for any $a, b \in \mathbf{R}$
- powers: x^{α} on \mathbf{R}_{++} , for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on \mathbf{R}_{++}

Examples on \mathbb{R}^n and \mathbb{R}^{m \times n}

affine functions are convex and concave; all norms are convex

examples on \mathbf{R}^n

• affine function $f(x) = a^T x + b$

examples on $\mathbf{R}^{m \times n}$ ($m \times n$ matrices)

• affine function

$$f(X) = \mathbf{tr}(A^T X) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} X_{ij} + b$$

Restriction of a convex function to a line

 $f: \mathbf{R}^n \to \mathbf{R}$ is convex if and only if the function $g: \mathbf{R} \to \mathbf{R}$,

 $g(t) = f(x + tv), \qquad \operatorname{dom} g = \{t \mid x + tv \in \operatorname{dom} f\}$

is convex (in t) for any $x \in \operatorname{\mathbf{dom}} f$, $v \in \mathbf{R}^n$.

can check convexity of f by checking convexity of functions of one variable

First-order condition

f is differentiable if $\operatorname{\mathbf{dom}} f$ is open and the gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n}\right)$$

exists at each $x \in \operatorname{\mathbf{dom}} f$

1st-order condition: differentiable f with convex domain is convex iff

$$f(y) \ge f(x) + \nabla f(x)^T (y - x)$$
 for all $x, y \in \operatorname{dom} f$



first-order approximation of f is global underestimator

Second-order conditions

f is twice differentiable if dom f is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \dots, n,$$

exists at each $x \in \operatorname{\mathbf{dom}} f$

2nd-order conditions for twice differentiable f with convex domain

• f is convex if and only if

$$\nabla^2 f(x) \succeq 0$$
 for all $x \in \operatorname{\mathbf{dom}} f$

• if $\nabla^2 f(x) \succ 0$ for all $x \in \operatorname{\mathbf{dom}} f$, then f is strictly convex

Examples

quadratic function: $f(x) = (1/2)x^T P x + q^T x + r$ (with $P \in \mathbf{S}^n$)

$$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$$

convex if $P \succeq 0$

least-squares objective: $f(x) = ||Ax - b||_2^2$

$$\nabla f(x) = 2A^T (Ax - b), \qquad \nabla^2 f(x) = 2A^T A$$

convex (for any A)

quadratic-over-linear: $f(x,y) = x^2/y$ $\nabla^2 f(x,y) = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0$

convex for y > 0



Epigraph and sublevel set

 α -sublevel set of $f : \mathbf{R}^n \to \mathbf{R}$:

$$C_{\alpha} = \{ x \in \operatorname{dom} f \mid f(x) \le \alpha \}$$

sublevel sets of convex functions are convex (converse is false) epigraph of $f : \mathbb{R}^n \to \mathbb{R}$:

$$\mathbf{epi}\,f = \{(x,t) \in \mathbf{R}^{n+1} \mid x \in \mathbf{dom}\,f, \ f(x) \le t\}$$



f is convex if and only if $\operatorname{\mathbf{epi}} f$ is a convex set

Jensen's inequality

basic inequality: if f is convex, then for $0 \le \theta \le 1$,

$$f(\theta x + (1 - \theta)y) \le \theta f(x) + (1 - \theta)f(y)$$

extension: if f is convex, then

$$f(\mathbf{E}\,z) \le \mathbf{E}\,f(z)$$

for any random variable z

basic inequality is special case with discrete distribution

$$\operatorname{prob}(z=x) = \theta, \quad \operatorname{prob}(z=y) = 1 - \theta$$

Optimization problem in standard form

minimize
$$f_0(x)$$

subject to $f_i(x) \le 0$, $i = 1, ..., m$
 $h_i(x) = 0$, $i = 1, ..., p$

- $x \in \mathbf{R}^n$ is the optimization variable
- $f_0: \mathbf{R}^n \to \mathbf{R}$ is the objective or cost function
- $f_i : \mathbf{R}^n \to \mathbf{R}, i = 1, \dots, m$, are the inequality constraint functions
- $h_i: \mathbf{R}^n \to \mathbf{R}$ are the equality constraint functions

optimal value:

$$p^{\star} = \inf\{f_0(x) \mid f_i(x) \le 0, \ i = 1, \dots, m, \ h_i(x) = 0, \ i = 1, \dots, p\}$$

- $p^{\star} = \infty$ if problem is infeasible (no x satisfies the constraints)
- $p^{\star} = -\infty$ if problem is unbounded below

Optimal and locally optimal points

x is feasible if $x \in \operatorname{dom} f_0$ and it satisfies the constraints

- a feasible x is optimal if $f_0(x) = p^*$; X_{opt} is the set of optimal points
- x is locally optimal if there is an R > 0 such that x is optimal for

minimize (over z)
$$f_0(z)$$

subject to $f_i(z) \le 0, \quad i = 1, \dots, m, \quad h_i(z) = 0, \quad i = 1, \dots, p$
 $\|z - x\|_2 \le R$

examples (with n = 1, m = p = 0)

- $f_0(x) = 1/x$, dom $f_0 = \mathbf{R}_{++}$: $p^* = 0$, no optimal point
- $f_0(x) = -\log x$, **dom** $f_0 = \mathbf{R}_{++}$: $p^* = -\infty$
- $f_0(x) = x \log x$, $\operatorname{dom} f_0 = \mathbf{R}_{++}$: $p^* = -1/e$, x = 1/e is optimal
- $f_0(x) = x^3 3x$, $p^* = -\infty$, local optimum at x = 1

Review on Convex Optimization

Implicit constraints

the standard form optimization problem has an implicit constraint

$$x \in \mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom} f_{i} \cap \bigcap_{i=1}^{p} \operatorname{dom} h_{i},$$

- \bullet we call ${\mathcal D}$ the domain of the problem
- the constraints $f_i(x) \leq 0$, $h_i(x) = 0$ are the explicit constraints
- a problem is unconstrained if it has no explicit constraints (m = p = 0)

example:

minimize
$$f_0(x) = -\sum_{i=1}^k \log(b_i - a_i^T x)$$

is an unconstrained problem with implicit constraints $a_i^T x < b_i$

Feasibility problem

find
$$x$$

subject to $f_i(x) \le 0, \quad i = 1, \dots, m$
 $h_i(x) = 0, \quad i = 1, \dots, p$

can be considered a special case of the general problem with $f_0(x) = 0$:

minimize 0
subject to
$$f_i(x) \le 0$$
, $i = 1, \dots, m$
 $h_i(x) = 0$, $i = 1, \dots, p$

- $p^{\star} = 0$ if constraints are feasible; any feasible x is optimal
- $p^{\star} = \infty$ if constraints are infeasible

Convex optimization problem

standard form convex optimization problem

minimize
$$f_0(x)$$

subject to $f_i(x) \le 0$, $i = 1, ..., m$
 $a_i^T x = b_i$, $i = 1, ..., p$

- f_0 , f_1 , . . . , f_m are convex; equality constraints are affine
- problem is quasiconvex if f_0 is quasiconvex (and f_1, \ldots, f_m convex)

often written as

minimize
$$f_0(x)$$

subject to $f_i(x) \le 0$, $i = 1, ..., m$
 $Ax = b$

important property: feasible set of a convex optimization problem is convex

example

$$\begin{array}{ll} \mbox{minimize} & f_0(x) = x_1^2 + x_2^2 \\ \mbox{subject to} & f_1(x) = x_1/(1+x_2^2) \leq 0 \\ & h_1(x) = (x_1+x_2)^2 = 0 \end{array}$$

- f_0 is convex; feasible set $\{(x_1, x_2) \mid x_1 = -x_2 \leq 0\}$ is convex
- not a convex problem (according to our definition): f_1 is not convex, h_1 is not affine
- equivalent (but not identical) to the convex problem

minimize
$$x_1^2 + x_2^2$$

subject to $x_1 \le 0$
 $x_1 + x_2 = 0$

Local and global optima

any locally optimal point of a convex problem is (globally) optimal proof: suppose x is locally optimal and y is optimal with $f_0(y) < f_0(x)$ x locally optimal means there is an R > 0 such that

$$z$$
 feasible, $||z - x||_2 \le R \implies f_0(z) \ge f_0(x)$

consider $z = \theta y + (1 - \theta)x$ with $\theta = R/(2\|y - x\|_2)$

•
$$||y - x||_2 > R$$
, so $0 < \theta < 1/2$

- z is a convex combination of two feasible points, hence also feasible
- $||z x||_2 = R/2$ and

$$f_0(z) \le \theta f_0(x) + (1 - \theta) f_0(y) < f_0(x)$$

which contradicts our assumption that x is locally optimal

Optimality criterion for differentiable f_0

x is optimal if and only if it is feasible and

 $\nabla f_0(x)^T(y-x) \ge 0$ for all feasible y



if nonzero, $\nabla f_0(x)$ defines a supporting hyperplane to feasible set X at x

• unconstrained problem: x is optimal if and only if

 $x \in \operatorname{dom} f_0, \qquad \nabla f_0(x) = 0$

• equality constrained problem

minimize $f_0(x)$ subject to Ax = b

x is optimal if and only if there exists a ν such that

 $x \in \operatorname{dom} f_0, \qquad Ax = b, \qquad \nabla f_0(x) + A^T \nu = 0$

minimization over nonnegative orthant

minimize $f_0(x)$ subject to $x \succeq 0$

x is optimal if and only if

$$x \in \operatorname{dom} f_0, \qquad x \succeq 0, \qquad \left\{ \begin{array}{ll} \nabla f_0(x)_i \ge 0 & x_i = 0\\ \nabla f_0(x)_i = 0 & x_i > 0 \end{array} \right.$$

Linear program (LP)

minimize
$$c^T x + d$$

subject to $Gx \leq h$
 $Ax = b$

- convex problem with affine objective and constraint functions
- feasible set is a polyhedron



Examples

diet problem: choose quantities x_1, \ldots, x_n of n foods

- one unit of food j costs c_j , contains amount a_{ij} of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

to find cheapest healthy diet,

$$\begin{array}{lll} \mbox{minimize} & c^T x \\ \mbox{subject to} & Ax \succeq b, \quad x \succeq 0 \end{array}$$

piecewise-linear minimization

minimize
$$\max_{i=1,\ldots,m}(a_i^T x + b_i)$$

equivalent to an LP

$$\begin{array}{ll} \mbox{minimize} & t \\ \mbox{subject to} & a_i^T x + b_i \leq t, \quad i = 1, \dots, m \end{array}$$

Chebyshev center of a polyhedron

Chebyshev center of

$$\mathcal{P} = \{ x \mid a_i^T x \le b_i, \ i = 1, \dots, m \}$$

is center of largest inscribed ball

$$\mathcal{B} = \{x_c + u \mid ||u||_2 \le r\}$$

•
$$a_i^T x \leq b_i$$
 for all $x \in \mathcal{B}$ if and only if

$$\sup\{a_i^T (x_c + u) \mid ||u||_2 \leq r\} = a_i^T x_c + r ||a_i||_2 \leq b_i$$

• hence, x_c , r can be determined by solving the LP

maximize
$$r$$

subject to $a_i^T x_c + r ||a_i||_2 \le b_i, \quad i = 1, \dots, m$



Quadratic program (QP)

minimize
$$(1/2)x^T P x + q^T x + r$$

subject to $Gx \leq h$
 $Ax = b$

- $P \in \mathbf{S}_{+}^{n}$, so objective is convex quadratic
- minimize a convex quadratic function over a polyhedron



Examples

least-squares

minimize $||Ax - b||_2^2$

- analytical solution $x^* = A^{\dagger}b$ (A^{\dagger} is pseudo-inverse)
- can add linear constraints, e.g., $l \preceq x \preceq u$

linear program with random cost

$$\begin{array}{ll} \mbox{minimize} & \bar{c}^T x + \gamma x^T \Sigma x = \mathbf{E} \, c^T x + \gamma \, \mathbf{var}(c^T x) \\ \mbox{subject to} & G x \preceq h, \quad A x = b \end{array}$$

- c is random vector with mean \bar{c} and covariance Σ
- hence, $c^T x$ is random variable with mean $\bar{c}^T x$ and variance $x^T \Sigma x$
- $\gamma > 0$ is risk aversion parameter; controls the trade-off between expected cost and variance (risk)

Quadratically constrained quadratic program (QCQP)

minimize
$$(1/2)x^T P_0 x + q_0^T x + r_0$$

subject to $(1/2)x^T P_i x + q_i^T x + r_i \le 0, \quad i = 1, \dots, m$
 $Ax = b$

- $P_i \in \mathbf{S}_+^n$; objective and constraints are convex quadratic
- if $P_1, \ldots, P_m \in \mathbf{S}_{++}^n$, feasible region is intersection of m ellipsoids and an affine set

Lagrangian

standard form problem (not necessarily convex)

minimize
$$f_0(x)$$

subject to $f_i(x) \le 0$, $i = 1, ..., m$
 $h_i(x) = 0$, $i = 1, ..., p$

variable $x \in \mathbf{R}^n$, domain \mathcal{D} , optimal value p^{\star}

Lagrangian: $L: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with $\operatorname{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$,

$$L(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

Lagrange dual function: $g: \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$g(\lambda,\nu) = \inf_{x \in \mathcal{D}} L(x,\lambda,\nu)$$
$$= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$$

g is concave, can be $-\infty$ for some $\lambda,\,\nu$

lower bound property: if $\lambda \succeq 0,$ then $g(\lambda,\nu) \leq p^{\star}$

proof: if \tilde{x} is feasible and $\lambda \succeq 0$, then

$$f_0(\tilde{x}) \ge L(\tilde{x}, \lambda, \nu) \ge \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible \tilde{x} gives $p^{\star} \geq g(\lambda,\nu)$

Least-norm solution of linear equations

 $\begin{array}{ll} \text{minimize} & x^T x\\ \text{subject to} & Ax = b \end{array}$

dual function

- Lagrangian is $L(x,\nu) = x^T x + \nu^T (Ax b)$
- to minimize L over x, set gradient equal to zero:

$$\nabla_x L(x,\nu) = 2x + A^T \nu = 0 \quad \Longrightarrow \quad x = -(1/2)A^T \nu$$

• plug in in L to obtain g:

$$g(\nu) = L((-1/2)A^T\nu, \nu) = -\frac{1}{4}\nu^T A A^T\nu - b^T\nu$$

a concave function of ν

lower bound property: $p^{\star} \geq -(1/4)\nu^T A A^T \nu - b^T \nu$ for all ν

Standard form LP

minimize
$$c^T x$$

subject to $Ax = b$, $x \succeq 0$

dual function

• Lagrangian is

$$L(x,\lambda,\nu) = c^T x + \nu^T (Ax - b) - \lambda^T x$$
$$= -b^T \nu + (c + A^T \nu - \lambda)^T x$$

• L is affine in x, hence

$$g(\lambda,\nu) = \inf_{x} L(x,\lambda,\nu) = \begin{cases} -b^{T}\nu & A^{T}\nu - \lambda + c = 0\\ -\infty & \text{otherwise} \end{cases}$$

g is linear on affine domain $\{(\lambda,\nu)\mid A^T\nu-\lambda+c=0\},$ hence concave

lower bound property: $p^{\star} \geq -b^T \nu$ if $A^T \nu + c \succeq 0$

The dual problem

Lagrange dual problem

 $\begin{array}{ll} \text{maximize} & g(\lambda,\nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$

- finds best lower bound on $p^{\star}\textsc{,}$ obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted d^{\star}
- λ , ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \operatorname{dom} g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \operatorname{dom} g$ explicit

example: standard form LP and its dual (page 33)

$$\begin{array}{ll} \mbox{minimize} & c^T x & \mbox{maximize} & -b^T \nu \\ \mbox{subject to} & Ax = b & \mbox{subject to} & A^T \nu + c \succeq 0 \\ & x \succeq 0 & \end{array}$$

Weak and strong duality

weak duality: $d^{\star} \leq p^{\star}$

• always holds (for convex and nonconvex problems)

strong duality: $d^{\star} = p^{\star}$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called constraint qualifications

Slater's constraint qualification

strong duality holds for a convex problem

minimize
$$f_0(x)$$

subject to $f_i(x) \le 0$, $i = 1, ..., m$
 $Ax = b$

if it is strictly feasible, *i.e.*,

$$\exists x \in \operatorname{int} \mathcal{D}: \quad f_i(x) < 0, \quad i = 1, \dots, m, \quad Ax = b$$

- also guarantees that the dual optimum is attained (if $p^{\star} > -\infty$)
- can be sharpened: e.g., can replace int D with relint D (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, . . .
- there exist many other types of constraint qualifications

Inequality form LP

primal problem

 $\begin{array}{ll} \text{minimize} & c^T x\\ \text{subject to} & Ax \preceq b \end{array}$

dual function

$$g(\lambda) = \inf_{x} \left((c + A^T \lambda)^T x - b^T \lambda \right) = \begin{cases} -b^T \lambda & A^T \lambda + c = 0\\ -\infty & \text{otherwise} \end{cases}$$

dual problem

$$\begin{array}{ll} \text{maximize} & -b^T\lambda\\ \text{subject to} & A^T\lambda+c=0, \quad \lambda\succeq 0 \end{array}$$

- from Slater's condition: $p^{\star} = d^{\star}$ if $A\tilde{x} \prec b$ for some \tilde{x}
- in fact, $p^* = d^*$ except when primal and dual are infeasible

Quadratic program

primal problem (assume $P \in \mathbf{S}_{++}^n$) minimize $x^T P x$ subject to $Ax \leq b$

dual function

$$g(\lambda) = \inf_{x} \left(x^T P x + \lambda^T (A x - b) \right) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

dual problem

maximize
$$-(1/4)\lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

subject to $\lambda \succeq 0$

- from Slater's condition: $p^{\star} = d^{\star}$ if $A\tilde{x} \prec b$ for some \tilde{x}
- in fact, $p^{\star} = d^{\star}$ always

Complementary slackness

assume strong duality holds, x^{\star} is primal optimal, $(\lambda^{\star},\nu^{\star})$ is dual optimal

$$f_0(x^*) = g(\lambda^*, \nu^*) = \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right)$$
$$\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*)$$
$$\leq f_0(x^*)$$

hence, the two inequalities hold with equality

- x^* minimizes $L(x, \lambda^*, \nu^*)$
- $\lambda_i^{\star} f_i(x^{\star}) = 0$ for i = 1, ..., m (known as complementary slackness):

$$\lambda_i^{\star} > 0 \Longrightarrow f_i(x^{\star}) = 0, \qquad f_i(x^{\star}) < 0 \Longrightarrow \lambda_i^{\star} = 0$$

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i , h_i):

- 1. primal feasibility: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$
- 2. dual feasibility: $\lambda \succeq 0$
- 3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \dots, m$
- 4. first order condition (gradient of Lagrangian with respect to x vanishes):

$$\nabla_x L(x,\lambda,\nu) = \nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + \sum_{i=1}^p \nu_i \nabla h_i(x) = 0$$

from page 39: if strong duality holds and x, λ , ν are optimal, then they must satisfy the KKT conditions

KKT conditions for convex problem

if \tilde{x} , $\tilde{\lambda}$, $\tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

if Slater's condition is satisfied:

x is optimal if and only if there exist $\lambda,\,\nu$ that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

example: water-filling (assume $\alpha_i > 0$)

minimize
$$-\sum_{i=1}^{n} \log(x_i + \alpha_i)$$

subject to $x \succeq 0, \quad \mathbf{1}^T x = 1$

x is optimal iff $x\succeq 0,~\mathbf{1}^Tx=1,$ and there exist $\lambda\in\mathbf{R}^n,~\nu\in\mathbf{R}$ such that

$$\lambda \succeq 0, \qquad \lambda_i x_i = 0, \qquad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

• if
$$\nu < 1/\alpha_i$$
: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$

• if
$$\nu \ge 1/\alpha_i$$
: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$

• determine ν from $\mathbf{1}^T x = \sum_{i=1}^n \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

- n patches; level of patch i is at height α_i
- flood area with unit amount of water
- resulting level is $1/\nu^{\star}$



Solving unconstrained minimization

minimize f(x)

- f convex, twice continuously differentiable (hence dom f open)
- we assume optimal value $p^* = \inf_x f(x)$ is attained (and finite)

unconstrained minimization methods

• produce sequence of points $x^{(k)} \in \operatorname{\mathbf{dom}} f$, $k=0,1,\ldots$ with

 $f(x^{(k)}) \to p^{\star}$

• can be interpreted as iterative methods for solving optimality condition

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

Initial point and sublevel set

algorithms in this chapter require a starting point $x^{(0)}$ such that

- $x^{(0)} \in \operatorname{dom} f$
- sublevel set $S = \{x \mid f(x) \le f(x^{(0)})\}$ is closed

2nd condition is hard to verify, except when *all* sublevel sets are closed:

- equivalent to condition that epi f is closed
- true if $\operatorname{\mathbf{dom}} f = \mathbf{R}^n$
- true if $f(x) \to \infty$ as $x \to \operatorname{\mathbf{bd}} \operatorname{\mathbf{dom}} f$

examples of differentiable functions with closed sublevel sets:

$$f(x) = \log(\sum_{i=1}^{m} \exp(a_i^T x + b_i)), \qquad f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$$

Strong convexity and implications

f is strongly convex on ${\cal S}$ if there exists an m>0 such that

 $\nabla^2 f(x) \succeq mI$ for all $x \in S$

implications

• for $x, y \in S$,

$$f(y) \ge f(x) + \nabla f(x)^T (y - x) + \frac{m}{2} ||x - y||_2^2$$

....

hence, S is bounded

• $p^{\star} > -\infty$, and for $x \in S$,

$$f(x) - p^* \le \frac{1}{2m} \|\nabla f(x)\|_2^2$$

useful as stopping criterion (if you know m)

Descent methods

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)} \quad \text{with } f(x^{(k+1)}) < f(x^{(k)})$$

- other notations: $x^+ = x + t\Delta x$, $x := x + t\Delta x$
- Δx is the step, or search direction; t is the step size, or step length
- from convexity, $f(x^+) < f(x)$ implies $\nabla f(x)^T \Delta x < 0$ (*i.e.*, Δx is a *descent direction*)

General descent method.

given a starting point $x \in \operatorname{dom} f$.

repeat

1. Determine a descent direction Δx .

2. *Line search.* Choose a step size t > 0.

3. Update. $x := x + t\Delta x$.

until stopping criterion is satisfied.

Line search types

exact line search: $t = \operatorname{argmin}_{t>0} f(x + t\Delta x)$

backtracking line search (with parameters $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$)

• starting at t = 1, repeat $t := \beta t$ until

$$f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$$

• graphical interpretation: backtrack until $t \leq t_0$



Gradient descent method

general descent method with $\Delta x = -\nabla f(x)$

given a starting point $x \in \text{dom } f$. **repeat** 1. $\Delta x := -\nabla f(x)$. 2. *Line search*. Choose step size t via exact or backtracking line search. 3. *Update*. $x := x + t\Delta x$. **until** stopping criterion is satisfied.

- stopping criterion usually of the form $\|\nabla f(x)\|_2 \leq \epsilon$
- convergence result: for strongly convex f,

$$f(x^{(k)}) - p^* \le c^k (f(x^{(0)}) - p^*)$$

 $c \in (0,1)$ depends on m, $x^{(0)}$, line search type

• very simple, but often very slow; rarely used in practice

quadratic problem in ${\boldsymbol{\mathsf{R}}}^2$

$$f(x) = (1/2)(x_1^2 + \gamma x_2^2) \qquad (\gamma > 0)$$

with exact line search, starting at $x^{(0)} = (\gamma, 1)$:

$$x_1^{(k)} = \gamma \left(\frac{\gamma - 1}{\gamma + 1}\right)^k, \qquad x_2^{(k)} = \left(-\frac{\gamma - 1}{\gamma + 1}\right)^k$$

• very slow if
$$\gamma \gg 1$$
 or $\gamma \ll 1$

• example for $\gamma = 10$:



nonquadratic example

$$f(x_1, x_2) = e^{x_1 + 3x_2 - 0.1} + e^{x_1 - 3x_2 - 0.1} + e^{-x_1 - 0.1}$$



 $x^{(0)}$ $x^{(1)}$

backtracking line search

exact line search

a problem in $\boldsymbol{\mathsf{R}}^{100}$



'linear' convergence, i.e., a straight line on a semilog plot