Optimization Tutorial + Fast and Accurate Approximate Query Processing with Visual Predicates

Stephen Macke

18 March 2016
Survey Results

What is your preferred format for this Friday's presentation? (14 responses)

- 57.1%: Only optimization tutorial (full hour)
- 21.4%: Mostly optimization tutorial, some of Stephen's research (45 min + 15 min respectively)
- 21.4%: Half and half (30 min + 30 min)
- Less than 10%: Some optimization tutorial, mostly Stephen's research (15 min + 45 min)
- Less than 10%: All about Stephen's research (full hour)
- Less than 10%: Other

For the optimization tutorial format, I want (14 responses)

- 50%: an interactive tutorial with examples and programming exercises (e.g. accessible from an ipython notebook)
- 50%: just a lecture with some slides
- Less than 10%: I don't want an optimization tutorial.
- Less than 10%: Other
Survey Results (cont’d)

I want the following material covered in the optimization tutorial: (15 responses)

- Basics (conv...): 7 (46.7%)
- Convex optim.: 9 (60%)
- Strong/weak...: 8 (53.3%)
- I don’t want...: 0 (0%)
Presentation Format

• Convex sets
• Convex functions
• Convex optimization problems
• Duality
• Application of ideas in my approximate query processing research
Convex Optimization

• Slides stolen shamelessly from Stephen Boyd
• More complete ones available at: ee364a.stanford.edu
• Convex Optimization textbook available from same website
Mathematical optimization

(mathematical) optimization problem

minimize $f_0(x)$
subject to $f_i(x) \leq b_i$, $i = 1, \ldots, m$

• $x = (x_1, \ldots, x_n)$: optimization variables

• $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$: objective function

• $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \ldots, m$: constraint functions

optimal solution $x^*$ has smallest value of $f_0$ among all vectors that satisfy the constraints
Examples

portfolio optimization

- variables: amounts invested in different assets
- constraints: budget, max./min. investment per asset, minimum return
- objective: overall risk or return variance

device sizing in electronic circuits

- variables: device widths and lengths
- constraints: manufacturing limits, timing requirements, maximum area
- objective: power consumption

data fitting

- variables: model parameters
- constraints: prior information, parameter limits
- objective: measure of misfit or prediction error
Convex optimization problem

minimize \( f_0(x) \)
subject to \( f_i(x) \leq b_i, \quad i = 1, \ldots, m \)

- objective and constraint functions are convex:

  \[
  f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)
  \]

  if \( \alpha + \beta = 1, \alpha \geq 0, \beta \geq 0 \)

- includes least-squares problems and linear programs as special cases
solving convex optimization problems

• no analytical solution
• reliable and efficient algorithms
• computation time (roughly) proportional to $\max\{n^3, n^2m, F\}$, where $F$ is cost of evaluating $f_i$’s and their first and second derivatives
• almost a technology

using convex optimization

• often difficult to recognize
• many tricks for transforming problems into convex form
• surprisingly many problems can be solved via convex optimization
2. Convex sets

- affine and convex sets
- some important examples
- operations that preserve convexity
- generalized inequalities
- separating and supporting hyperplanes
- dual cones and generalized inequalities
Convex set

**line segment** between $x_1$ and $x_2$: all points

$$x = \theta x_1 + (1 - \theta)x_2$$

with $0 \leq \theta \leq 1$

**convex set**: contains line segment between any two points in the set

$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \implies \quad \theta x_1 + (1 - \theta)x_2 \in C$$

**examples** (one convex, two nonconvex sets)
Convex cone

**conic (nonnegative) combination** of $x_1$ and $x_2$: any point of the form

$$x = \theta_1 x_1 + \theta_2 x_2$$

with $\theta_1 \geq 0$, $\theta_2 \geq 0$

**convex cone**: set that contains all conic combinations of points in the set
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 are affine and convex; halfspaces are convex
Euclidean balls and ellipsoids

(Euclidean) ball with center $x_c$ and radius $r$:

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

ellipsoid: set of the form

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

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
Positive semidefinite cone

notation:
- $S^n$ is set of symmetric $n \times n$ matrices
- $S_+^n = \{ X \in S^n \mid X \succeq 0 \}$: positive semidefinite $n \times n$ matrices
  \[ X \in S_+^n \iff z^T X z \geq 0 \text{ for all } z \]
- $S_+^n$ is a convex cone
- $S_{++}^n = \{ X \in S^n \mid X \succ 0 \}$: positive definite $n \times n$ matrices

example: $\begin{bmatrix} x & y \\ y & z \end{bmatrix} \in S_+^2$
3. Convex functions

- basic properties and examples
- operations that preserve convexity
- the conjugate function
- quasiconvex functions
- log-concave and log-convex functions
- convexity with respect to generalized inequalities
Definition

\( f : \mathbb{R}^n \to \mathbb{R} \) is convex if \( \text{dom} \ f \) is a convex set and

\[
f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)
\]

for all \( x, y \in \text{dom} \ f, \ 0 \leq \theta \leq 1 \)

- \( f \) is concave if \(-f\) is convex
- \( f \) is strictly convex if \( \text{dom} \ f \) is convex and

\[
f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)
\]

for \( x, y \in \text{dom} \ f, \ x \neq y, \ 0 < \theta < 1 \)
Examples on $\mathbb{R}$

convex:

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

concave:

- affine: $ax + b$ on $\mathbb{R}$, for any $a, b \in \mathbb{R}$
- powers: $x^\alpha$ on $\mathbb{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbb{R}_{++}$
Second-order conditions

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

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

exists at each $x \in \text{dom } f$

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

- $f$ is convex if and only if
  $$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \text{dom } f$$

- if $\nabla^2 f(x) \succ 0$ for all $x \in \text{dom } f$, then $f$ is strictly convex
Examples

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

\[
\nabla f(x) = P x + q, \quad \nabla^2 f(x) = P
\]

convex if \( P \succeq 0 \)

**least-squares objective:** \( f(x) = \| A x - b \|_2^2 \)

\[
\nabla f(x) = 2 A^T (A x - b), \quad \nabla^2 f(x) = 2 A^T A
\]

convex (for any \( A \))

**quadratic-over-linear:** \( f(x, y) = x^2 / y \)

\[
\nabla^2 f(x, y) = \frac{2}{y^3} \left[ \begin{array}{c} y \\ -x \end{array} \right] \left[ \begin{array}{c} y \\ -x \end{array} \right]^T \succeq 0
\]

convex for \( y > 0 \)
**log-sum-exp**: \( f(x) = \log \sum_{k=1}^{n} \exp x_k \) is convex

\[
\nabla^2 f(x) = \frac{1}{1^T z} \text{diag}(z) - \frac{1}{(1^T z)^2} zz^T \quad (z_k = \exp x_k)
\]

to show \( \nabla^2 f(x) \succeq 0 \), we must verify that \( v^T \nabla^2 f(x) v \geq 0 \) for all \( v \):

\[
v^T \nabla^2 f(x) v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k v_k z_k)^2}{(\sum_k z_k)^2} \geq 0
\]

since \( (\sum_k v_k z_k)^2 \leq (\sum_k z_k v_k^2)(\sum_k z_k) \) (from Cauchy-Schwarz inequality)

**geometric mean**: \( f(x) = (\prod_{k=1}^{n} x_k)^{1/n} \) on \( \mathbb{R}_{++}^n \) is concave

(similar proof as for log-sum-exp)
Jensen’s inequality

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

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

**extension:** if $f$ is convex, then

$$f(Ez) \leq E f(z)$$

for any random variable $z$

basic inequality is special case with discrete distribution

$$\text{prob}(z = x) = \theta, \quad \text{prob}(z = y) = 1 - \theta$$
Operations that preserve convexity

practical methods for establishing convexity of a function

1. verify definition (often simplified by restricting to a line)

2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$

3. show that $f$ is obtained from simple convex functions by operations that preserve convexity
   - nonnegative weighted sum
   - composition with affine function
   - pointwise maximum and supremum
   - composition
   - minimization
   - perspective
Pointwise maximum

if \( f_1, \ldots, f_m \) are convex, then \( f(x) = \max\{f_1(x), \ldots, f_m(x)\} \) is convex

examples

• piecewise-linear function: \( f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i) \) is convex

• sum of \( r \) largest components of \( x \in \mathbb{R}^n \):

\[
  f(x) = x[1] + x[2] + \cdots + x[r]
\]

is convex (\( x[i] \) is \( i \)th largest component of \( x \))

proof:

\[
  f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}
\]
Composition with scalar functions

composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$:

$$f(x) = h(g(x))$$

$f$ is convex if $g$ convex, $h$ convex, $\tilde{h}$ nondecreasing

$g$ concave, $h$ convex, $\tilde{h}$ nonincreasing

• proof (for $n = 1$, differentiable $g, h$)

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

• note: monotonicity must hold for extended-value extension $\tilde{h}$

examples

• $\exp g(x)$ is convex if $g$ is convex

• $1/g(x)$ is convex if $g$ is concave and positive
Vector composition

composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$:

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), \ldots, g_k(x))$$

$f$ is convex if $g_i$ convex, $h$ convex, $\tilde{h}$ nondecreasing in each argument

$g_i$ concave, $h$ convex, $\tilde{h}$ nonincreasing in each argument

proof (for $n = 1$, differentiable $g, h$)

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$$

examples

- $\sum_{i=1}^{m} \log g_i(x)$ is concave if $g_i$ are concave and positive
- $\log \sum_{i=1}^{m} \exp g_i(x)$ is convex if $g_i$ are convex
Minimization

If \( f(x, y) \) is convex in \((x, y)\) and \( C \) is a convex set, then

\[
g(x) = \inf_{y \in C} f(x, y)
\]

is convex

examples

- \( f(x, y) = x^T A x + 2x^T B y + y^T C y \) with

\[
\begin{bmatrix}
  A & B \\
  B^T & C
\end{bmatrix} \succeq 0, \quad C \succ 0
\]

minimizing over \( y \) gives \( g(x) = \inf_y f(x, y) = x^T (A - BC^{-1}B^T)x \)

\( g \) is convex, hence Schur complement \( A - BC^{-1}B^T \succeq 0 \)

- distance to a set: \( \text{dist}(x, S) = \inf_{y \in S} \|x - y\| \) is convex if \( S \) is convex
4. Convex optimization problems

- optimization problem in standard form
- convex optimization problems
- quasiconvex optimization
- linear optimization
- quadratic optimization
- geometric programming
- generalized inequality constraints
- semidefinite programming
- vector optimization
Optimization problem in standard form

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p 
\end{align*}
\]

- \( x \in \mathbb{R}^n \) is the optimization variable
- \( f_0 : \mathbb{R}^n \rightarrow \mathbb{R} \) is the objective or cost function
- \( f_i : \mathbb{R}^n \rightarrow \mathbb{R}, \ i = 1, \ldots, m, \) are the inequality constraint functions
- \( h_i : \mathbb{R}^n \rightarrow \mathbb{R} \) are the equality constraint functions

**optimal value:**

\[
p^* = \inf \{ f_0(x) \mid f_i(x) \leq 0, \ i = 1, \ldots, m, \ h_i(x) = 0, \ i = 1, \ldots, p \}
\]

- \( p^* = \infty \) if problem is infeasible (no \( x \) satisfies the constraints)
- \( p^* = -\infty \) if problem is unbounded below
Feasibility problem

\[ \begin{align*}
\text{find} & \quad x \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*} \]

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

\[ \begin{align*}
\text{minimize} & \quad 0 \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*} \]

- \( p^* = 0 \) if constraints are feasible; any feasible \( x \) is optimal
- \( p^* = \infty \) if constraints are infeasible
example

\[
\begin{align*}
\text{minimize} & \quad f_0(x) = x_1^2 + x_2^2 \\
\text{subject to} & \quad f_1(x) = x_1/(1 + x_2^2) \leq 0 \\
& \quad h_1(x) = (x_1 + x_2)^2 = 0 \\
\end{align*}
\]

• \( f_0 \) 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

\[
\begin{align*}
\text{minimize} & \quad x_1^2 + x_2^2 \\
\text{subject to} & \quad x_1 \leq 0 \\
& \quad x_1 + x_2 = 0 \\
\end{align*}
\]
Linear program (LP)

\[
\begin{align*}
\text{minimize} & \quad c^T x + d \\
\text{subject to} & \quad Gx \preceq h \\
& \quad Ax = b
\end{align*}
\]

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

Convex optimization problems
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,

$$\begin{align*}
\text{minimize} \quad & c^T x \\
\text{subject to} \quad & Ax \succeq b, \quad x \succeq 0
\end{align*}$$

piecewise-linear minimization

$$\begin{align*}
\text{minimize} \quad & \max_{i=1,\ldots,m} (a_i^T x + b_i) \\
\end{align*}$$

equivalent to an LP

$$\begin{align*}
\text{minimize} \quad & t \\
\text{subject to} \quad & a_i^T x + b_i \leq t, \quad i = 1, \ldots, m
\end{align*}$$
Quadratic program (QP)

minimize \( \frac{1}{2}x^T Px + q^T x + r \)
subject to \( Gx \preceq h \)
\( Ax = b \)

- \( P \in S^n_+ \), so objective is convex quadratic
- minimize a convex quadratic function over a polyhedron
Examples

least-squares

\[
\text{minimize} \quad \|Ax - b\|_2^2
\]

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

linear program with random cost

\[
\begin{align*}
\text{minimize} & \quad \bar{c}^T x + \gamma x^T \Sigma x = E c^T x + \gamma \text{var}(c^T x) \\
\text{subject to} & \quad Gx \preceq h, \quad Ax = b
\end{align*}
\]

• \( c \) is random vector with mean \( \bar{c} \) and covariance \( \Sigma \)
• 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 \[ \frac{1}{2} x^T P_0 x + q_0^T x + r_0 \]
subject to \[ \frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \]
\[ Ax = b \]

- \( P_i \in S^n_+ \); objective and constraints are convex quadratic
- if \( P_1, \ldots, P_m \in S^n_+ \), feasible region is intersection of \( m \) ellipsoids and an affine set
Semidefinite program (SDP)

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad x_1 F_1 + x_2 F_2 + \cdots + x_n F_n + G \preceq 0 \\
& \quad Ax = b
\end{align*}
\]

with \( F_i, G \in \mathbf{S}^k \)

- inequality constraint is called linear matrix inequality (LMI)
- includes problems with multiple LMI constraints: for example,

\[
x_1 \hat{F}_1 + \cdots + x_n \hat{F}_n + \hat{G} \preceq 0, \quad x_1 \tilde{F}_1 + \cdots + x_n \tilde{F}_n + \tilde{G} \preceq 0
\]

is equivalent to single LMI

\[
x_1 \begin{bmatrix} \hat{F}_1 & 0 \\ 0 & \tilde{F}_1 \end{bmatrix} + x_2 \begin{bmatrix} \hat{F}_2 & 0 \\ 0 & \tilde{F}_2 \end{bmatrix} + \cdots + x_n \begin{bmatrix} \hat{F}_n & 0 \\ 0 & \tilde{F}_n \end{bmatrix} + \begin{bmatrix} \hat{G} & 0 \\ 0 & \tilde{G} \end{bmatrix} \preceq 0
\]
Eigenvalue minimization

\[
\text{minimize } \lambda_{\text{max}}(A(x))
\]

where \( A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n \) (with given \( A_i \in S^k \))

equivalent SDP

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad A(x) \preceq tI
\end{align*}
\]

- variables \( x \in \mathbb{R}^n, t \in \mathbb{R} \)
- follows from

\[
\lambda_{\text{max}}(A) \leq t \iff A \preceq tI
\]
Matrix norm minimization

\[
\text{minimize } \| A(x) \|_2 = \left( \lambda_{\text{max}}(A(x)^T A(x)) \right)^{1/2}
\]

where \( A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n \) (with given \( A_i \in \mathbb{R}^{p \times q} \))
equivalent SDP

\[
\text{minimize } t
\]
\[
\text{subject to } \begin{bmatrix} tI & A(x) \\ A(x)^T & tI \end{bmatrix} \succeq 0
\]

- variables \( x \in \mathbb{R}^n, t \in \mathbb{R} \)
- constraint follows from

\[
\| A \|_2 \leq t \iff A^T A \leq t^2 I, \quad t \geq 0
\]
\[
\iff \begin{bmatrix} tI & A \\ A^T & tI \end{bmatrix} \succeq 0
\]
5. Duality

- Lagrange dual problem
- weak and strong duality
- geometric interpretation
- optimality conditions
- perturbation and sensitivity analysis
- examples
- generalized inequalities
**Lagrangian**

**standard form problem** (not necessarily convex)

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

variable \( x \in \mathbb{R}^n \), domain \( \mathcal{D} \), optimal value \( p^* \)

**Lagrangian**: \( L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \), with \( \text{dom} \, L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \),

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

- weighted sum of objective and constraint functions
- \( \lambda_i \) is Lagrange multiplier associated with \( f_i(x) \leq 0 \)
- \( \nu_i \) is Lagrange multiplier associated with \( h_i(x) = 0 \)
Lagrange dual function

Lagrange dual function: \( g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}, \)

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

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

**lower bound property:** if \( \lambda \succeq 0 \), then \( g(\lambda, \nu) \leq p^* \)

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

\[
f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)
\]

minimizing over all feasible \( \tilde{x} \) gives \( p^* \geq g(\lambda, \nu) \)
Two-way partitioning

minimize \( x^T W x \)
subject to \( x_i^2 = 1, \quad i = 1, \ldots, n \)

- a nonconvex problem; feasible set contains \( 2^n \) discrete points

- interpretation: partition \( \{1, \ldots, n\} \) in two sets; \( W_{ij} \) is cost of assigning \( i, j \) to the same set; \( -W_{ij} \) is cost of assigning to different sets

**dual function**

\[
g(\nu) = \inf_{x} (x^T W x + \sum_{i} \nu_i (x_i^2 - 1)) = \inf_{x} x^T (W + \text{diag}(\nu)) x - 1^T \nu
\]

\[
= \begin{cases} 
-1^T \nu & W + \text{diag}(\nu) \succeq 0 \\
-\infty & \text{otherwise}
\end{cases}
\]

**lower bound property:** \( p^* \geq -1^T \nu \) if \( W + \text{diag}(\nu) \succeq 0 \)

example: \( \nu = -\lambda_{\min}(W) 1 \) gives bound \( p^* \geq n \lambda_{\min}(W) \)
The dual problem

Lagrange dual problem

\[
\begin{align*}
\text{maximize} & \quad g(\lambda, \nu) \\
\text{subject to} & \quad \lambda \succeq 0
\end{align*}
\]

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

**example:** standard form LP and its dual (page 5–5)

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b, x \succeq 0 \\
\text{maximize} & \quad -b^T \nu \\
\text{subject to} & \quad A^T \nu + c \succeq 0
\end{align*}
\]
Weak and strong duality

**Weak duality**: \( d^* \leq p^* \)

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

  for example, solving the SDP

  \[
  \begin{align*}
  \text{maximize} & \quad -1^T \nu \\
  \text{subject to} & \quad W + \text{diag}(\nu) \succeq 0
  \end{align*}
  \]

  gives a lower bound for the two-way partitioning problem on page 5–7

**Strong duality**: \( d^* = p^* \)

- 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

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

if it is strictly feasible, i.e.,

\[\exists x \in \mathbf{int} \mathcal{D} : \quad f_i(x) < 0, \quad i = 1, \ldots, m, \quad Ax = b\]

• also guarantees that the dual optimum is attained (if \( p^* > -\infty \))

• can be sharpened: e.g., can replace \( \mathbf{int} \mathcal{D} \) with \( \mathbf{relint} \mathcal{D} \) (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, . . .

• there exist many other types of constraint qualifications
Quadratic program

**primal problem** (assume $P \in \mathbb{S}_{++}^n$)

minimize $x^T P x$

subject to $Ax \preceq b$

**dual function**

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

**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: $p^* = d^*$ if $A\tilde{x} < b$ for some $\tilde{x}$

• in fact, $p^* = d^*$ always
Complementary slackness

assume strong duality holds, \( x^* \) is primal optimal, \((\lambda^*, \nu^*)\) 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 \( L(x, \lambda^*, \nu^*) \)

• \( \lambda_i^* f_i(x^*) = 0 \) for \( i = 1, \ldots, m \) (known as complementary slackness):

\[
\lambda_i^* > 0 \implies f_i(x^*) = 0, \quad f_i(x^*) < 0 \implies \lambda_i^* = 0
\]
Karush-Kuhn-Tucker (KKT) conditions

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

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

\[
\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0
\]

from page 5–17: if strong duality holds and $x$, $\lambda$, $\nu$ 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 $x \succeq 0, \; 1^T x = 1$

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

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

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine $\nu$ from $1^T x = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

- $n$ patches; level of patch $i$ is at height $\alpha_i$
- flood area with unit amount of water
- resulting level is $1/\nu^*$
Perturbation and sensitivity analysis

( unperturbed) optimization problem and its dual

\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p \\
\text{maximize} & \quad g(\lambda, \nu) \\
\text{subject to} & \quad \lambda \succeq 0
\end{align*}

perturbed problem and its dual

\begin{align*}
\text{min.} & \quad f_0(x) \\
\text{s.t.} & \quad f_i(x) \leq u_i, \quad i = 1, \ldots, m \\
& \quad h_i(x) = v_i, \quad i = 1, \ldots, p \\
\text{max.} & \quad g(\lambda, \nu) - u^T \lambda - v^T \nu \\
\text{s.t.} & \quad \lambda \succeq 0
\end{align*}

- $x$ is primal variable; $u, v$ are parameters
- $p^*(u, v)$ is optimal value as a function of $u, v$
- we are interested in information about $p^*(u, v)$ that we can obtain from the solution of the unperturbed problem and its dual
global sensitivity result

assume strong duality holds for unperturbed problem, and that $\lambda^*$, $\nu^*$ are dual optimal for unperturbed problem

apply weak duality to perturbed problem:

$$p^*(u, v) \geq g(\lambda^*, \nu^*) - u^T \lambda^* - v^T \nu^*$$

$$= p^*(0, 0) - u^T \lambda^* - v^T \nu^*$$

sensitivity interpretation

- if $\lambda^*_i$ large: $p^*$ increases greatly if we tighten constraint $i$ ($u_i < 0$)
- if $\lambda^*_i$ small: $p^*$ does not decrease much if we loosen constraint $i$ ($u_i > 0$)
- if $\nu^*_i$ large and positive: $p^*$ increases greatly if we take $v_i < 0$;
  if $\nu^*_i$ large and negative: $p^*$ increases greatly if we take $v_i > 0$
- if $\nu^*_i$ small and positive: $p^*$ does not decrease much if we take $v_i > 0$;
  if $\nu^*_i$ small and negative: $p^*$ does not decrease much if we take $v_i < 0$
Part II: Fast and Accurate Approximate Query Processing with Visual Predicates

• Format:
  • Problem definition / motivation
  • Approach
  • System architecture
  • Algorithm
  • Results
The Problem

- Database queries can *generate* visualizations:
  - SELECT x, AVG(y) FROM table WHERE z=z* GROUP BY x

- x, y, z are all columns

- Suppose: x==day_of_week, y==delay, z==airport: this can be used to plot delay vs. day of week for some airport

- Given: query visualization Q, columns x, y, z.

- We want to *find* instances of z such that the above query generates visualizations “similar looking” to Q

- More precisely: we want top-K closest visualizations to Q, where we define distance later
Example

- Left == Query viz, right == result returned to analyst.
Why is it important?

• Quickly identify commonalities among data following a particular distribution.

• Suggest explanations for outlier visualizations.
Why is it hard?

• Can’t precompute indexes over all possible columns

• Computing K-closest visualizations exactly then requires a full disc scan.

• Solution: take samples!
High Level Approach

• Analyst wants top-K closest visualizations to Q

• Compute confidence intervals around candidate visualization distance from the query.

• Keep sampling and recomputing confidence intervals until:
  • Current top-K closest candidate confidence intervals stop intersecting bottom N-K, OR
  • Tired of waiting (might make mistakes!)
Distance of Visualizations from Query

- Define $d(r, Q)$ as $\| \frac{\vec{r}}{\vec{1}^T \vec{r}} - \frac{\vec{Q}}{\vec{1}^T \vec{Q}} \|_2$

- We denote normalized vectors with a hat: $\| \hat{\vec{r}} - \hat{\vec{Q}} \|_2$

- This matches our intuition that analysts are mostly interested in relative deviations from query.
Why is it *still* hard?

• How to get confidence intervals around candidate distance from query?
  • Recall our defn. of candidate distance: \( \| \hat{r} - \hat{Q} \|_2 \),
  • Lots of literature on how to get per-group CIs around individual components of \( r \).
    • Approx CIs: bootstrap, normal asymptotics (Student T intervals)
    • Worst-case CIs: martingale and concentration bounds (e.g. Hoeffding-Serfling inequality)
  • We would like to be able to leverage any of these.
Simple Observation

• If B defines a $1-\alpha$ confidence hyperrectangle for per-group aggregates, then:

$$[\inf_{\mathbf{r} \in B} ||\mathbf{r} - \hat{Q}||_2, \sup_{\mathbf{r} \in B} ||\mathbf{r} - \hat{Q}||_2]$$ must contain a $1-\alpha$ confidence interval surrounding candidate distance to the query.

• New problem: how to compute this?
Finding the Infimum

- To compute $\inf_{\hat{r} \in B} ||\hat{r} - \hat{Q}||_2$, formulate the following optimization problem:

**Problem (opt-lower-1).**

Minimize: $||\frac{\hat{r}}{C} - \hat{Q}||_2^2$

subject to: $C = \hat{r}^T \hat{1}$

$L_j \leq r_j \leq U_j, \quad j = 1, \ldots, |Vx|$

- Nonconvex!
Solution: reformulate to QP

• opt-lower-1 turns out to be equivalent to:

\[\textbf{Problem (opt-lower-2).}\]

Minimize: \[\|\mathbf{\hat{r}} - \mathbf{\tilde{Q}}\|_2^2\]

subject to: \[\mathbf{1}^T \mathbf{\hat{r}} = 1\]

\[\kappa L_j \leq \hat{r}_j \leq \kappa U_j, \quad j = 1, \ldots, |V_x|\]

\[\mathbf{\hat{r}} \succeq 0\]

• This is a convex QP and is readily solved!
What about the supremum?

• Can only get upper bound through:
  • Ad-hoc relaxation of constraints
  • Semidefinite relaxation
  • Dual relaxation

• We skip technical details in the interest of time.
System architecture
Experimental Results

- Based on ~2.5 million rows in a flight delays table.
- Querying for top 30 closest visualizations.

Precision vs. $1 - \alpha$

Tuples sampled vs. $1 - \alpha$
Thanks!