indent best frequently has constraints

- nutrition
- groceries $\vec{g} \geq 0$
- prices $\vec{p}$
- price $\min_{\vec{g}} \vec{g} \cdot \vec{p}$
- minimum requirements $\vec{m}$
- nutrition value $\vec{N}$

$\vec{N} \cdot \vec{g} \geq \vec{m}$

defines linear program, LP

- price may be a function of quantity, not linear quadratic objective, quadratic program, QP
- general case mathematical program
- portfolios, routing airplanes, running a factory program as plan, not computer program, can be same electrical networks [Dennis, 1958]
- flow control [Low et al., 2002]
- layering [Chiang et al., 2007]
- sorting

variables $\vec{x}$, objective minimize $f(\vec{x})$, constraints $\vec{c}(\vec{x})$

max = -min

slack variables to convert inequality to equality

$$c(\vec{x}) \geq 0$$ (17.1)

replace with

$$c(\vec{x}) - s = 0$$

$$s \geq 0$$ (17.2)

combinatorial $x$ equals 1 or -1 relaxed as algebraic constraint $(x^2 - 1)^2 = 0$

L1 norm

$$|\vec{x}| = \sum_i |x_i|$$ (17.3)
compressed sensing, sparsity
non-differentiable
[Schmidt et al., 2007]
\((x)_+ = \max(x, 0)\)
\((x)_- = \max(-x, 0)\)

\[ |x| = (x)_- + (x)_+ \] (17.4)

\[ |x| \approx |x|_\alpha \]
\[ = \frac{1}{\alpha} \left[ \log \left( 1 + e^{-\alpha x} \right) + \log \left( 1 + e^{\alpha x} \right) \right] \] (17.5)

\[ \frac{d |x|_\alpha}{dx} = \frac{1}{1 + e^{-\alpha x}} - \frac{1}{1 + e^{\alpha x}} \] (17.6)

\[ \frac{d^2 |x|_\alpha}{dx^2} = \frac{2\alpha e^{\alpha x}}{(1 + e^{\alpha x})^2} \] (17.7)

minimize for increasing \(\alpha\)

### 17.1 LAGRANGE MULTIPLIERS

single equality constraint \(c(\vec{x}) = 0\)
step in direction \(\vec{d}\) to minimize \(f\) while satisfying the constraint

\[ 0 = c(\vec{x} + \vec{\delta}) \]
\[ \approx c(\vec{x}) + \nabla c \cdot \vec{\delta} \]
\[ = \nabla c \cdot \vec{\delta} \] (17.8)

step also minimizes \(f\)

\[ 0 > f(\vec{x} + \vec{\delta}) - f(\vec{x}) \]
\[ \approx f(\vec{x}) + \nabla f \cdot \vec{\delta} - f(\vec{x}) \]
\[ = \nabla f \cdot \vec{\delta} \] (17.9)

if \(\nabla c(\vec{x})\) and \(\nabla f(\vec{x})\) aligned not possible to find a direction, hence \(\vec{x}\) is a local minimizer

define Lagrangian

\[ \mathcal{L} = f(\vec{x}) - \lambda c(\vec{x}) \] (17.10)

solve for
\[ 0 = \nabla L = \nabla f - \lambda \nabla c \]  
\[ \text{(17.11)} \]

multiple constraints
linear combination

\[ \nabla f(\vec{x}) = \sum_i \lambda_i \nabla c_i(\vec{x}) \]  
\[ \text{(17.12)} \]

\[ f(\vec{x}) = \sum_i \lambda_i c_i(\vec{x}) \]  
\[ \text{(17.13)} \]

gives \( \vec{x}(\vec{\lambda}) \), substitute into constraints to find \( \vec{\lambda} \)

inequality constraint

\[ 0 \leq c(\vec{x} + \delta) \approx c(\vec{x}) + \nabla c \cdot \delta \]  
\[ \text{(17.14)} \]

if constraint not active \( (c > 0) \), can just do gradient descent \( \delta = -\alpha \nabla f \)

for an active constraint \( \nabla f \cdot \delta < 0 \) and \( \nabla c \cdot \delta \geq 0 \)
define half-planes
no intersection if point in same direction \( \nabla f = \lambda \nabla c \)
same condition, but now \( \lambda \geq 0 \)

## 17.2 Optimality

first-order

equality constraints \( c_i(\vec{x}), i \in \mathcal{E} \)
inequality constraints \( c_i(\vec{x}), i \in \mathcal{I} \)
inactive constraint \( \lambda_i = 0 \)
complementarity: \( \lambda_i c_i = 0 \): Lagrange multiplier only non-zero when constraint is active, otherwise reduces to gradient descent

\[ \nabla_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda}) = 0 \]
\[ c_i(\vec{x}) = 0 \quad (i \in \mathcal{E}) \]
\[ c_i(\vec{x}) \geq 0 \quad (i \in \mathcal{I}) \]
\[ \lambda_i \geq 0 \quad (i \in \mathcal{I}) \]
\[ \lambda_i c_i(x) = 0 \]  
\[ \text{(17.15)} \]

Karush-Kuhn-Tucker (KKT) conditions
necessary
second order: positive definite Lagrangian Hessian

sensitivity
replace \( c(x) = 0 \) with \( c(x) = \epsilon \)

minimizer \( \bar{x} \) goes to \( \bar{x}_\epsilon \)

\[
f(\bar{x}_\epsilon) - f(\bar{x}) \approx \nabla f \cdot (\bar{x}_\epsilon - \bar{x})
= \lambda \nabla c \cdot (\bar{x}_\epsilon - \bar{x})
\approx \lambda (c(\bar{x}_\epsilon) - c(\bar{x}))
= \lambda \epsilon
\]

\[
\frac{df}{d\epsilon} = \lambda
\]  \hspace{1cm} (17.16)

shadow prices: change in utility per change in constraint

\( \bar{x} \) primal \( \lambda \) dual

multi-objective Pareto

not possible to improve one constraint without making others worse
defines Pareto frontier
can combine in multi-objective function with relative weights

\section*{17.3 Solvers}

analytically can solve Lagrangian, then find Lagrange multipliers from constraints

\subsection*{17.3.1 Penalty}

penalty combine

\[
\mathcal{F} = f(\bar{x}) + \frac{\mu}{2} \sum_i c_i^2(\bar{x})
\]  \hspace{1cm} (17.17)

\[
\frac{\partial \mathcal{F}}{\partial x_j} = \frac{\partial f}{\partial x_j} + \mu \sum_i c_i \frac{\partial c_i}{\partial x_j}
\]  \hspace{1cm} (17.18)

\[
\mathcal{L} = f(\bar{x}) - \sum_i \lambda_i c_i(\bar{x})
\]  \hspace{1cm} (17.19)

\[
\frac{\partial \mathcal{L}}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_i \lambda_i \frac{\partial c_i}{\partial x_j}
\]  \hspace{1cm} (17.20)

effectively taking \( c_i = -\lambda_i / \mu \)
solving a different problem driven to 0 as \( \mu \to \infty \)
large \( \mu \) ill-conditioned

nonsmooth penalty
Constrained Optimization

\[ \mathcal{F} = f(\bar{x}) + \mu \sum_{i \in E} |c_i(\bar{x})| + \mu \sum_{i \in I} |c_i(\bar{x})|_+ \]  

(17.21)

exact for large enough \( \mu \) [Nocedal & Wright, 2006]

non-differentiable

sub-gradient approximate (17.5)

Newton steps, increase

17.3.2 Augmented Lagrangian

augmented Lagrangian

\[ \mathcal{L} = f(\bar{x}) - \sum_{i} \lambda_i c_i(\bar{x}) + \frac{\mu}{2} \sum_{i} c_i^2(\bar{x}) \]  

(17.22)

\[ \frac{\partial \mathcal{L}}{\partial \bar{x}_j} = \frac{\partial f}{\partial x_j} - \sum_{i} \lambda_i \frac{\partial c_i}{\partial x_j} + \mu \sum_{i} c_i \frac{\partial c_i}{\partial x_j} \]  

(17.23)

\[ \lambda_i^* = \lambda_i - \mu c_i \]

\[ c_i = (\lambda_i - \lambda_i^*)/\mu \]

vanishes much faster, as Lagrange multiplier estimates converge

\[ \lambda_i^{(n+1)} = \lambda_i^{(n)} - \mu c_i \]

minimize \( \bar{x} \), update \( \lambda \), increase \( \mu \)

17.3.3 Interior Point

inequality constraints

interior point

directly solve KKT system of equations

avoid boundaries

primal-dual

\[
\begin{align*}
\min_{\bar{x}} f(\bar{x}) \\
\text{subject to } & \, \bar{c}_E(\bar{x}) = 0 \\
& \, \bar{c}_I(\bar{x}) - \bar{s} = 0 \\
& \, \bar{s} \geq 0
\end{align*}
\]  

(17.24)

solve KKT, perturb from boundary

\[
\begin{align*}
\nabla f - \lambda_E \cdot \nabla c_E - \lambda_I \cdot \nabla c_I &= 0 \\
\bar{c}_E(\bar{x}) &= 0 \\
\bar{c}_I(\bar{x}) - \bar{s} &= 0 \\
\lambda_i s_i &= \mu
\end{align*}
\]  

(17.25)
Newton step on system
decrease $\mu$
same as barrier
minimize

$$\min_{\bar{x}, \bar{s}} f(x) - \mu \sum_i \log s_i$$
subject to $\bar{c}_E(\bar{x}) = 0$
$\bar{c}_I(\bar{x}) - \bar{s} = 0$

KKT for $s_i$

$$\frac{1}{s_i} - \lambda_i = 0$$

$$\lambda_i s_i = \mu$$

17.4 SELECTED REFERENCES

Unusually clear coverage of a field full of unusually opaque books.

17.5 PROBLEMS

(16.1) Given a point $(x_0, y_0)$, find the closest point on the line $y = ax + b$ by minimizing the distance $d^2 = (x_0 - x)^2 + (y_0 - y)^2$ subject to the constraint $y - ax - b = 0$.

(16.2) Consider a set of $N$ nodes that has each measured a quantity $x_i$. The goal is to find the best estimate $\bar{x}$ by minimizing

$$\min_{\bar{x}} \sum_{i=1}^{N} (\bar{x} - x_i)^2$$

however each node $i$ can communicate only with nodes $j$ in its neighborhood $j \in \mathcal{N}(i)$. This can be handled by having each node obtain a local estimate $\bar{x}_i$, and introducing a consistency constraint $c_{ij} = \bar{x}_i - \bar{x}_j = 0 \forall j \in \mathcal{N}(i)$.

(a) What is the Lagrangian?

(b) Find an update rule for the estimates $\bar{x}_i$ by evaluating where the gradient of the Lagrangian vanishes.

(c) Find an update rule for the Lagrange multipliers by taking a Newton step on their constraints.

(16.3) What is the Newton step for the interior point KKT system?

(16.4) Solve a 1D spin glass (Problem 14.2) as a constrained optimization with relaxed spins.
(16.5) compressed sensing ...  
  ... choose random frequencies and amplitudes  
  ... generate time series  
  ... sample random subset of points  
  ... equality constraint $A \cdot \vec{x} - \vec{b} = 0$  
  ... calculate minimum L2 norm $\vec{x}$ from SVD  
  ... calculate minimum L1 norm $\vec{x}$  
  ... approximate L1 norm, minimize exact penalty, increase  
  ... compare time series  
  ... compare Nyquist requirement