Convex Optimization and ℓ_1-minimization

Sangwoon Yun

Computational Sciences
Korea Institute for Advanced Study

December 11, 2009
2009 NIMS Thematic Winter School

Outline

I. Convex Optimization

II. ℓ_1-Minimization: Sparse Model

I. Convex Optimization

Content

Mathematical Optimization

(Mathematical) optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le b_i, \quad i = 1, ..., m \end{array}$$

- $x \in \Re^n$: optimization variable
- $f_0 : \Re^n \to \Re$: objective function
- $f_i : \Re^n \to \Re, i = 1, ..., m$: constraint functions

Optimization solution x^* has smallest value of f_0 among all vectors that satisfy the constraints.

Optimization Problem in Standard Form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, ..., m \\ & h_i(x) = 0, \quad i = 1, ..., p \end{array}$$

- $x \in \Re^n$: optimization variable
- $f_0 : \Re^n \to \Re$: objective or cost function
- $f_i : \Re^n \to \Re, \ i = 1, ..., m,$: inequality constraint functions
- $h_i : \Re^n \to \Re, \ i = 1, ..., p,$: equality constraint functions

Optimal Value

$$p^* = \inf\{f_0(x) \mid f_i(x) \leq 0, \quad i = 1, ..., m, \quad h_i(x) = 0, \quad i = 1, ..., p\}$$

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

Examples

Portfolio Optimization

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

Data Fitting

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

Solving Optimization Problems

General Optimization Problem

- very difficult to solve
- methods involve some compromise, e.g., very long computation time, or not always finding the solution

Exceptions: certain problem classes can be solved efficiently and reliably

- least-squares problems
- linear programming problems
- convex optimization problems

Convex Optimization

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m \end{array}$$

- objective and constraint functions are convex:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)$$

if $\alpha + \beta = 1, \quad \alpha \geq 0, \quad \beta \geq 0$

Standard Form Convex Optimization Problem

$$\begin{array}{rl} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, ..., m \\ & a_i^T x = b_i, \quad i = 1, ..., p \end{array}$$

- $f_0, f_1, ..., f_m$: convex
- equality constraints are affine

often written as

$$\begin{array}{rl} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, ..., m \\ & Ax = b \end{array}$$

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

Optimality Criterion for Differentiable f_0

- unconstrained problem: x is optimal if and only if

$$x \in \mathrm{dom} f_0, \quad \nabla f_0(x) = 0$$

- equality constrained problem

$$\text{minimize} \;\; f_0(x) \quad \text{subject to} \;\; Ax = b$$

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

$$x \in \mathrm{dom} f_0, \;\; Ax = b, \;\; \nabla f_0(x) + A^T v = 0$$

- minimization over nonnegative orthant

$$\text{minimize} \;\; f_0(x) \quad \text{subject to} \;\; x \geq 0$$

  x is optimal if and only if

$$x \in \mathrm{dom} f_0, \;\; x \geq 0, \;\; \begin{cases} \nabla f_0(x)_i \geq 0 & x_i = 0 \\ \nabla f_0(x)_i = 0 & x_i > 0 \end{cases}$$

Solving convex optimization problems

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

surprisingly many problems can be solved via convex optimization

many tricks for transforming problems into convex form

Linear Programming and Least Squares Problem

Linear Programming

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, ..., m \end{array}$$

Example

Diet problem: choose quantities $x_1, ..., x_n$ of n foods

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

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & Ax \geq b, \quad x \geq 0 \end{array}$$

Solving linear programmming problem

- no analytical formula for solution
- reliable and efficient algorithms and software
- computation time proportional to n^2m if $m \geq n$; less with structure

a few standard tricks used to convert problems into linear programs (e.g., problems involving ℓ_1- or ℓ_∞-norms, piecewise-linear functions)

Least Squares Problem

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

Solving least-squares problem

- analytical solution: $x^* = (A^T A)^{-1} A^T b$
- reliable and efficient algorithms and software
- computation time proportional to $n^2 m$ ($A \in \Re^{m \times n}$); less if structured

least-squares problems are easy to recognize

a few standard techniques increase flexibility (e.g., including weights, adding regularization terms)

Quadratic Program (QP)

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T P x + q^T x + r \\ \text{subject to} & Gx \leq h \\ & Ax = b \end{array}$$

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

Quadratically Constrained Quadratic Program (QCQP)

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & \frac{1}{2}x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, ..., m \\ & Ax = b \end{array}$$

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

Brief History of Convex Optimization

Theory (convex analysis): 1900–1970
Algorithms

- 1947: simplex algorithm for linear programming (Dantzig)
- 1960s: early interior-point methods (Fiacco & McCormick, Dikin, . . .)
- 1970s: ellipsoid method and other subgradient methods
- 1980s: polynomial-time interior-point methods for linear programming (Karmarkar 1984)
- late 1980s–now: polynomial-time interior-point methods for nonlinear convex optimization (Nesterov & Nemirovski 1994)

Applications

- before 1990: mostly in operations research; few in engineering
- since 1990: many new applications in engineering (control, signal/image processing, communications, circuit design,...); new problem classes (semidefinite and second-order cone programming)

Second-Order Cone Programming

$$
\begin{array}{rl}
\text{minimize} & f^T x \\
\text{subject to} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m \\
& Fx = g
\end{array}
$$

$(A_i \in \Re^{n_i \times n}, F \in \Re^{p \times n})$

- inequalities are called second-order cone (SOC) constraints:

$$(A_i x + b_i, c_i^T x + d_i) \in \text{second} - \text{order cone in } \Re^{n_i+1}$$

- for $n_i = 0$, reduces to an LP; if $c_i = 0$, reduces to a QCQP
- more general than QCQP and LP

Semidefinite Programming (SDP)

$$\begin{array}{ll} \text{minimize} & b^T x \\ \text{subject to} & x_1 F_1 + x_2 F_2 + \cdots + x_n F_n \preceq C \end{array}$$

with $F_i, C \in \mathcal{S}^n$

- inequality constraint is called linear matrix inequality (LMI)

Primal form

$$\begin{array}{ll} \text{minimize} & \langle C, X \rangle \\ \text{subject to} & \langle F_i, X \rangle = b_i, \quad i = 1, ..., m \\ & X \succeq 0 \end{array}$$

refer to this problem as a primal semidefinite program.

Unconstrained Minimization

$$\text{minimize} \quad f(x)$$

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

Unconstrained Minimization Methods

- produce sequence of points $x^k \in \mathrm{dom} f, \; k = 0, 1, \ldots$ with

$$f(x^k) \to p^*$$

- can be interpreted as iterative methods for solving optimality condition

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

Descent Methods

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

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

General Descent Method

given a starting point $x \in \mathrm{dom} f$
repeat

1. Determine a descent direction Δx
2. Line search. Choose a step size $t > 0$
3. Update. $x^{\text{new}} = x + t\Delta x$

until stopping criterion is satisfied.

Line Search Types

- exact line search: $t = \arg\min_{t>0} f(x + t\Delta x)$
- backtracking line search (with parameters $\alpha \in (0, 1/2),\ \beta \in (0, 1)$)
  starting at $t = 1$, repeat $t = \beta t$ until

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

Gradient Descent Methods

General Descent Method with $\Delta x = -\nabla f(x)$
given a starting point $x \in \mathrm{dom} f$
repeat

1. $\Delta x = -\nabla f(x)$
2. Line search. Choose step size t via exact or backtracking line search
3. Update. $x^{\mathrm{new}} = x + t\Delta x$

until stopping criterion is satisfied.

- stopping criterion usually of the form $\|\nabla f(x)\|_2 \leq \epsilon$
- convergence result: for strongly convex f ($\nabla^2 f(x) \succeq \delta I$, $\delta > 0$),

  $$f(x^k) - p^* \leq c^k(f(x^0) - p^*)$$

  $c^k \in (0, 1)$ depends on δ, x^0, line search type
- simple, but often very slow; rarely used in practice

Newton Step

$$\Delta x_{\mathrm{nt}} = -\nabla^2 f(x)^{-1} \nabla f(x)$$

Interpretations

- $x + \Delta x_{\mathrm{nt}}$ minimizes second order approximation

$$\hat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$$

- $x + \Delta x_{\mathrm{nt}}$ solves linearized optimality condition

$$\nabla f(x+v) \approx \nabla \hat{f}(x+v) = \nabla f(x) + \nabla^2 f(x) v = 0$$

Newton Decrement

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

a measure of the proximity of x to x^*

Properties

- gives an estimate of $f(x) - p^*$, using quadratic approximation $\hat{f}$:

$$f(x) - \inf_y \hat{f}(y) = \frac{1}{2}\lambda(x)^2$$

- equal to the norm of the Newton step in the quadratic Hessian norm

$$\lambda(x) = \left(\Delta x_{\mathrm{nt}}^T \nabla^2 f(x) \Delta x_{\mathrm{nt}}\right)^{1/2}$$

- directional derivative in the Newton direction: $\nabla f(x)^T \Delta x_{\mathrm{nt}} = -\lambda(x)^2$

Newton's Method

given a starting point $x \in \mathrm{dom} f$, tolerance $\epsilon > 0$

repeat

1. Compute the Newton step and decrement.

$$\Delta x_{\mathrm{nt}} = -\nabla^2 f(x)^{-1} \nabla f(x); \quad \lambda^2 = \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$$

2. Stopping criterion. quit if $\lambda^2/2 \leq \epsilon$
3. Line search. Choose step size t by backtracking line search
4. Update. $x = x + t\Delta x_{\mathrm{nt}}$

Classical Convergence Analysis

assumptions

- f strongly convex with constant δ
- $\nabla^2 f$ is Lipschitz continuous with constant $L > 0$:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2$$

(L measures how well f can be approximated by a quadratic function)

outline: there exist constants $\eta \in (0, \delta^2/L),\ \gamma > 0$ such that
if $\|\nabla f(x)\|_2 \geq \eta$, then $f(x^{k+1}) - f(x^k) \leq -\gamma$
if $\|\nabla f(x)\|_2 < \eta$, then

$$\frac{L}{2\delta^2}\|\nabla f(x^{k+1})\|_2 \leq \left(\frac{L}{2\delta^2}\|\nabla f(x^k)\|_2\right)^2$$

damped Newton phase ($\|\nabla f(x)\|_2 \geq \eta$)

- most iterations require backtracking steps
- function value decreases by at least γ
- if $p^* > -\infty$, this phase ends after at most $(f(x^0) - p^*)/\gamma$ iterations

quadratically convergent phase ($\|\nabla f(x)\|_2 > \eta$)

- all iterations use step size $t = 1$
- $\|\nabla f(x)\|_2$ converges to zero quadratically: if $\|\nabla^f(x)\|_2 < \eta$, then

$$\frac{L}{2\delta^2}\|\nabla f(x^l)\|_2 \leq \left(\frac{L}{2\delta^2}\|\nabla f(x^k)\|_2\right)^{2^{l-k}} \leq \left(\frac{1}{2}\right)^{2^{l-k}}, \quad l \geq k$$

conclusion: number of iterations until $f(x) - p^* \leq \epsilon$ is bounded above by

$$\frac{f(x^0) - p^*}{\gamma} + \log_2 \log_2(\epsilon_0/\epsilon)$$

- γ, ϵ_0 are constants that depend on δ, L, x^0
- second term is small and almost constant for practical purposes
- in practice, constants δ, L (hence γ, ϵ_0) are usually unknown
- provides qualitative insight in convergence properties (i.e., explains two algorithm phases)

Equality Constrained Minimization

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

- f convex, twice continuously differentiable
- $A \in \Re^{m \times n}$ with $\operatorname{rank} A = m$
- we assume $p^*$is finite and attained

Optimiality Conditions: x^* is optimal iff there exists a v^* such that

$$\nabla f(x^*) + A^T v^* = 0, \quad Ax^* = b$$

Newton Step

Newton step Δx_{nt} of f at feasible x is given by solution v of

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x); 0 \end{bmatrix}$$

Interpretations

- Δx_{nt} solves second order approximation (with variable v)

$$\begin{array}{ll} \text{minimize} & \hat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b \end{array}$$

- Δx_{nt} equations follow from linearizing optimality conditions

$$\nabla f(x+v) + A^T w \approx \nabla f(x) + \nabla^2 f(x) v + A^T w = 0, \quad A(x+v) = b$$

Newton's Method with Equality Constraints

given a starting point $x \in \mathrm{dom} f$ with $Ax = b$, tolerance $\epsilon > 0$

repeat

1. Compute the Newton step and decrement $\Delta x_{\mathrm{nt}}, \lambda(x)$
2. Stopping criterion. quit if $\lambda^2/2 \leq \epsilon$
3. Line search. Choose step size t by backtracking line search
4. Update. $x = x + t\Delta x_{\mathrm{nt}}$

a feasible descent method: x^k feasible and $f(x^{k+1}) < f(x^k)$

Interior-Point Methods

Inequality Constrained Minimization

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, ..., m \\ & Ax = b \end{array}$$

- f_i convex, twice continuously differentiable
- $A \in \Re^{p \times n}$ with $\mathrm{rank} A = p$
- we assume $p^*$is finite and attained
- we assume problem is strictly feasible: there exists $\tilde{x}$ with

$$\tilde{x} \in \mathrm{dom} f_0, \quad f_i(\tilde{x}) < 0, \quad i = 1, ..., m, \quad A\tilde{(x)} = b$$

Examples: LP, QP, QCQP, SDP, SOCP

Logarithmic Barrier

reformulation by using indicator function

$$\begin{array}{rl} \text{minimize} & f_0 + \sum_{i=1}^m I(f_i(x)) \\ \text{subject to} & Ax = b, \end{array}$$

where $I(u) = 0$ if $u \leq 0$, $I(u) = \infty$ otherwise

approximation via logarithmic barrier

$$\begin{array}{rl} \text{minimize} & f_0 - (1/t)\sum_{i=1}^m \log(-f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

- an equality constrained problem
- for $t > 0$, $-(1/t)\log(-u)$ is a smooth approximation of $I(\cdot)$
- approximation improves as $t \to \infty$

Central path

- for $t > 0$, define $x^*(t)$ as the solution of

$$\begin{array}{ll} \text{minimize} & tf_0(x) - \sum_{i=1}^m \log(-f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

  (for now, assume $x^*(t)$ exists and is unique for each $t > 0$)

- central path is $\{x^*(t) \mid t > 0\}$

Barrier Method

given strictly feasible $x, t = t^0 > 0, \mu > 1$, tolerance $\epsilon > 0$

repeat

1. Centering step. Compute $x^*(t)$
by minimizing $tf_0 - \sum_{i=1}^{m} \log(-f_i(x))$, subject to $Ax = b$
2. Update. $x = x^*(t)$
3. Stopping criterion. quit if $m/t \leq \epsilon$
4. Increase t. $t = \mu t$

- terminates with $f_0(x) - p^* \leq \epsilon$ (stopping criterion follows from $f_0(x^*(t)) - p^* \leq m/t$)
- centering usually done using Newton's method, starting at current x
- choice of μ involves a trade-off: large μ means fewer outer iterations, more inner (Newton) iterations;

Convergence Analysis

number of outer (centering) iterations: exactly

$$\left\lceil \frac{\log(m/(\epsilon t^0))}{\log \mu} \right\rceil$$

plus the initial centering step (to compute $x^*(t^0)$)

centering problem

$$\text{minimize} \quad tf_0(x) - \sum_{i=1}^{m} \log(-f_i(x))$$

see convergence analysis of Newton's method

- classical analysis requires strong convexity, Lipschitz condition

Primal-dual interior-point methods

more efficient than barrier method when high accuracy is needed

- update primal and dual variables at each iteration; no distinction between inner and outer iterations
- often exhibit superlinear asymptotic convergence
- search directions can be interpreted as Newton directions for modified KKT conditions
- can start at infeasible points
- cost per iteration same as barrier method

Conclusions

mathematical optimization

- problems in engineering design, data analysis and statistics, economics, management,..., can often be expressed as mathematical optimization problems

tractability

- roughly speaking, tractability in optimization requires convexity
- algorithms for nonconvex optimization find local (suboptimal) solutions, or are very expensive
- surprisingly many applications can be formulated as convex problems

Theoretical consequences of convexity

- local optima are global
- extensive duality theory (systematic way of deriving lower bounds on optimal value, necessary and sufficient optimality conditions)
- solution methods with polynomial worst-case complexity theory

Practical consequences of convexity

(most) convex problems can be solved globally and efficiently

- interior-point methods require 20 – 80 steps in practice
- basic algorithms (e.g., Newton, barrier method,...) are easy to implement and work well for small and medium size problems (larger problems if structure is exploited)
- more and more high-quality implementations of advanced algorithms and modeling tools are becoming available

Reference

Convex Optimization by Boyd and Vandenberghe
(http://www.stanford.edu/ boyd/cvxbook/)

II. ℓ_1-Minimization: Sparse Model

Content

- Underdetermined System of Linear Equations
- Sparse Model
- ℓ_1-Minimization & ℓ_1-regularization

Underdetermined System of Linear Equations

Many problems arising in signal/image processing and data mining/classification often entail finding a solution of an underdetermined system of linear equations:

$$Ax = b,$$

where $A \in \Re^{m \times n}$ ($m < n$) and $b \in \Re^m$, or a “best" solution of a inconsistent system of linear equations:

$$Ax \approx b,$$

i.e., the objective is to find a “best" guess of unknown u from linear measurement model

$$b = Au + \eta,$$

where η is measurement error.

Least Squares approach

Minimizing the 2-norm of the residual $Ax - b$ yields the well-known linear least squares problem:

$$\min_x \|Ax - b\|_2^2$$

In many cases, components of x are parameters that must lie within certain bounds:

$$\begin{aligned} \min_x \quad & \|Ax - b\|_2^2 \\ \text{s.t.} \quad & l \leq x \leq u, \end{aligned}$$

where $l \leq u$ (possibly with $-\infty$ or ∞ components).
This is a bound-constrained convex quadratic program.

There is considerable evidence that these problems arising in signal/image processing often have sparse solutions. Advances in finding sparse solutions to underdetermined systems have energized research on such signal and image processing problems.

- Signal Analysis
- Signal/image Compression
- Signal/image Denoising
- Inverse Problems: Even more generally, suppose that we observe not b, but a noisy indirect measurement of it, $\tilde{b} = Hb + \eta$ ($\tilde{b} = HAu + \eta$). Here the linear operator H generates blurring, masking, or some other kind of degradation, and η is noise as before.
- Compressed Sensing: For signals which are sparsely generated, one can obtain good reconstructions from reduced numbers of measurements - thereby compressing the sensing process rather than the traditionally sensed data.

- Morphological Component Analysis (MCA): Suppose that the observed signal is a superposition of two different subsignals b_1, b_2 (i.e., $b = b_1 + b_2$). Can we separate the two sources? Such source separation problems are fundamental in the processing of acoustic signals, for example, in the separation of speech from impulsive noise by independent component analysis (ICA) algorithms. An appealing image processing application that relies on MCA is inpainting, where missing pixels in an image are filled in, based on a sparse representation of the existing pixels. MCA is necessary because the piecewise smooth (cartoon) and texture contents of the image must be separated as part of this recovery process.
- image decomposition to cartoon and texture

This work lies at the intersection of signal processing and applied mathematics, and arose initially from the wavelets and harmonic analysis research communities.
Sparsity of representation is key to widely used techniques of transform-based image compression. Transform sparsity is also a driving factor for other important signal and image processing problems, including image denoising and image deblurring.

Sparsity Optimization (ℓ_0-norm minimization):

$$\begin{aligned} \min_{x} \quad & \|x\|_0 \\ \text{s.t} \quad & Ax = b \end{aligned}$$

(when the solution is sparse and the columns of A sufficiently incoherent)

NP hard!

Compressed Sensing: Is it possible to reconstruct a signal accurately from a few observed samples (measurements)?

Impossible in general, but if the signal is known to be sparse in some basis, then accurate recovery is possible by ℓ_1-minimization (known as Basis Pursuit) (Candés et al. 06, Donoho 06):

$$\begin{array}{rl} min_x & \|x\|_1 \\ \text{s.t} & Ax = b, \end{array}$$

where $A \in \Re^{m \times n}$ ($m \ll n$) satisfies certain restricted isometry property.

Definition: A matrix A satisfies the restricted isometry property of order s with constant $\delta_s \in (0, 1)$ if

$$(1 - \delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s)\|x\|_2^2$$

Linear Programming reformulation:

$$\begin{aligned} \min_{x^+, x^-} \quad & e^T x^+ + e^T x^- \\ \text{s.t} \quad & A(x^+ - x^-) = b \\ & x^+, x^- \geq 0, \end{aligned}$$

where $e = (1, 1, ..., 1) \in \Re^n$.

From Exact to Approximate Solution:

If the observation b is contaminated with noise, i.e., $b = Au + \eta$, then $Ax = b$ might not be feasible and so an appropriate norm of the residual $Ax - b$ should be minimized or considered as a constraint.

1. ℓ_1-minimization with quadratic constraints:

$$\begin{array}{cl} \min\limits_x & \|x\|_1 \\ \text{s.t} & \|Ax - b\|_2 \leq \lambda, \end{array}$$

where $\lambda > 0$.

2. Basis pursuit denoising (ℓ_1-regularized linear least squares):

$$\min_x \ \frac{1}{2}\|Ax - b\|^2 + \mu\|x\|_1,$$

where $\mu > 0$.

3. LASSO (least absolute shrinkage and selection operator):

$$\begin{array}{cl} \min_{x} & \frac{1}{2}\|Ax - b\|_2^2 \\ \text{s.t} & \|x\|_1 \leq \tau, \end{array}$$

where $\tau > 0$.

4. Dantzig selector (ℓ_1-minimization with bounded residual correlation):

$$\begin{array}{cl} \min_{x} & \|x\|_1 \\ \text{s.t} & \|A^T(Ax - b)\|_\infty \leq \sigma, \end{array}$$

where $\sigma > 0$.