# Complexity Analysis beyond Convex Optimization

Yinyu Ye

K. T. Li Professor of Engineering
Department of Management Science and Engineering
Stanford University

http://www.stanford.edu/˜yyye

August 1, 2013

# Outline

- Application arisen from Non-Convex Regularization
- Theory of the $L_p$-norm Regularization
- Selected Complexity Results for Non-Convex Optimization
- High-Level Complexity Analyses for Few cases
- Open Questions

# Unconstrained $L_2+L_p$ Minimization

Consider the problem:

$$\text{Minimize}_{\mathbf{x}\in R^n} \quad f_{2p}(\mathbf{x}) := \|A\mathbf{x}-\mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_p^p \qquad (1)$$

where data $A \in R^{m\times n}, \mathbf{b} \in R^m$, parameter $0 \le p \le 1$, and

$$\|x\|_p^p = \sum_{j=1}^{n} |x_j|^p.$$

# Unconstrained $L_2+L_p$ Minimization

Consider the problem:

$$\text{Minimize}_{\mathbf{x}\in R^n} \quad f_{2p}(\mathbf{x}) := \|A\mathbf{x}-\mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_p^p \tag{1}$$

where data $A \in R^{m\times n}, \mathbf{b} \in R^m$, parameter $0 \le p \le 1$, and

$$\|x\|_p^p = \sum_{j=1}^{n} |x_j|^p.$$

$$\|\mathbf{x}\|_0 := |\{j :\ x_j \neq 0\}|$$

that is, the number of nonzero entries in $\mathbf{x}$.

# Unconstrained $L_2+L_p$ Minimization

Consider the problem:

$$\text{Minimize}_{\mathbf{x}\in R^n} \quad f_{2p}(\mathbf{x}) := \|A\mathbf{x}-\mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_p^p \qquad (1)$$

where data $A \in R^{m\times n}, \mathbf{b} \in R^m$, parameter $0 \le p \le 1$, and

$$\|x\|_p^p = \sum_{j=1}^n |x_j|^p.$$

$$\|\mathbf{x}\|_0 := |\{j : \ x_j \neq 0\}|$$

that is, the number of nonzero entries in $\mathbf{x}$.

A more general model: for $q \ge 1$

$$\text{Minimize}_{\mathbf{x}\in R^n} \quad f_{qp}(\mathbf{x}) := \|A\mathbf{x}-\mathbf{b}\|_q^q + \lambda\|\mathbf{x}\|_p^p$$

# Constrained $L_p$ Minimization

Consider another problem:

$$\begin{array}{ll} \text{Minimize} & \sum_{1 \le j \le n} x_j^p \\ \text{Subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \ge \mathbf{0}, \end{array} \tag{2}$$

# Constrained $L_p$ Minimization

Consider another problem:

$$\begin{array}{ll} \text{Minimize} & \sum_{1 \le j \le n} x_j^p \\ \text{Subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \ge \mathbf{0}, \end{array} \tag{2}$$

or

$$\begin{array}{ll} \text{Minimize} & \sum_{1 \le j \le n} |x_j|^p \\ \text{Subject to} & A\mathbf{x} = \mathbf{b}. \end{array} \tag{3}$$

# Application and Motivation

The original goal is to minimize $\|\mathbf{x}\|_0 = |\{j : x_j \neq 0\}|$, the size of the support set of $\mathbf{x}$, such that $A\mathbf{x} = \mathbf{b}$, for

- Sparse image reconstruction
- Sparse signal recovering
- Sensor network localization

which is known to be an NP-Hard problem.

# Approximation of $\|\mathbf{x}\|_0$

- $\|\mathbf{x}\|_1$ has been used to approximate $\|\mathbf{x}\|_0$, and the regularization can be exact under certain strong conditions (Donoho 2004, Candès and Tao 2005, etc). This regularization model is actually a linear program.

# Approximation of $\|\mathbf{x}\|_0$

- $\|\mathbf{x}\|_1$ has been used to approximate $\|\mathbf{x}\|_0$, and the regularization can be exact under certain strong conditions (Donoho 2004, Candès and Tao 2005, etc). This regularization model is actually a linear program.
- Theoretical and empirical computational results indicate that $\|\mathbf{x}\|_p$ regularization, say $p = .5$, have better performances under weaker conditions, and it is solvable equally efficiently in practice (Chartrand 2009, Xu et al. 2009, etc).

# The Hardness of $L_p$ (Ge et al. 2011, Chen et al. 2012)

Question: is $L_2 + L_p$ minimization easier than $L_2 + L_0$ minimization?

# The Hardness of $L_p$ (Ge et al. 2011, Chen et al. 2012)

Question: is $L_2 + L_p$ minimization easier than $L_2 + L_0$ minimization?

**Theorem**

*Decide the global minimum of optimization problem $L_q + L_p$ is strongly NP-hard for any given $0 \leq p < 1$, $q \geq 1$ and $\lambda > 0$.*

# The Hardness of $L_p$ (Ge et al. 2011, Chen et al. 2012)

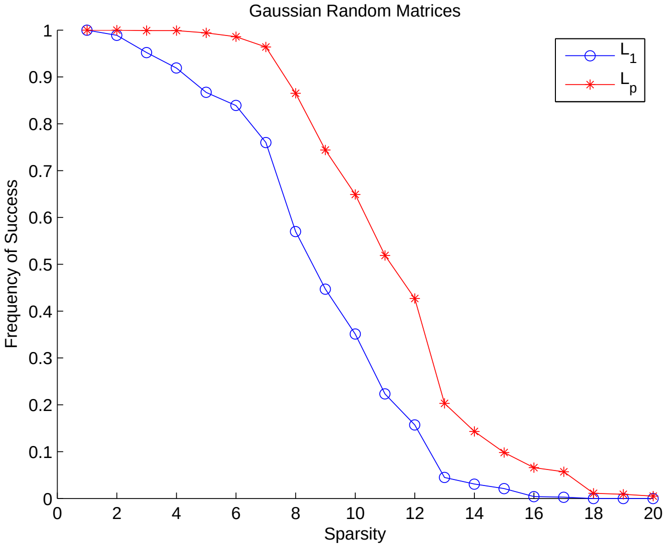
Question: is $L_2 + L_p$ minimization easier than $L_2 + L_0$ minimization?

**Theorem**

*Decide the global minimum of optimization problem $L_q + L_p$ is strongly NP-hard for any given $0 \leq p < 1$, $q \geq 1$ and $\lambda > 0$.*

Nevertheless, practitioners solve them using non-linear solvers to compute an KKT solution...

# Recover Result: $L_{0.5}$-Norm vs. $L_1$ Norm



# Sensor Network Localization

Given a graph $G = (V, E)$ and sets of non–negative weights, say $\{d_{ij} : (i, j) \in E\}$, the goal is to compute a realization of $G$ in the Euclidean space $\mathbf{R}^d$ for a given low dimension $d$, i.e.

- to place the vertexes of $G$ in $\mathbf{R}^d$ such that
- the Euclidean distance between a pair of adjacent vertexes $(i, j)$ equals to (or bounded by) the prescribed weight $d_{ij} \in E$.

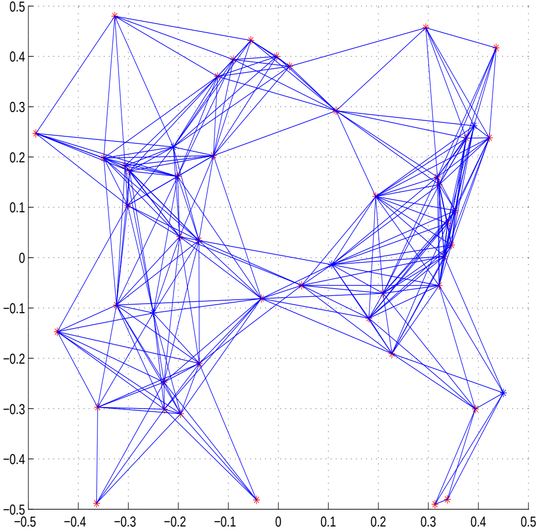



Figure: 50-node 2-D Sensor Localization

# Application to SNL

SNL-SDP:

$$\begin{array}{lll} \text{minimize} & \sum_{(i,j)\in E} \alpha_{ij}^2 & \\ \text{subject to} & (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \bullet Y & = d_{ij}^2 + \alpha_{ij},\ \forall\ (i,j) \in E, \\ & Y & \succeq \mathbf{0}. \end{array}$$

# Application to SNL

SNL-SDP:

$$\begin{array}{rcl} \text{minimize} & \sum_{(i,j)\in E} \alpha_{ij}^2 & \\ \text{subject to} & (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \bullet Y & = d_{ij}^2 + \alpha_{ij},\ \forall\,(i,j) \in E, \\ & Y & \succeq \mathbf{0}. \end{array}$$

Regularized SNL-SDP:

$$\begin{array}{rcl} \text{minimize} & \sum_{(i,j)\in E} \alpha_{ij}^2 + \lambda \|Y\|_p & \\ \text{subject to} & (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \bullet Y & = d_{ij}^2 + \alpha_{ij},\ \forall\,(i,j) \in E, \\ & Y & \succeq \mathbf{0}. \end{array}$$

# Schatten $p$-norm (Ji et al. 2013)

For any given symmetric matrix $Y \in S^n$,

$$\|Y\|_p = \left( \sum_j |\lambda(Y)_j|^p \right)^{1/p}, \; 0 < p \leq 1$$

is known as the Schatten $p$-quasi-norm of $Y$.

# Schatten $p$-norm (Ji et al. 2013)

For any given symmetric matrix $Y \in S^n$,

$$\|Y\|_p = \left( \sum_j |\lambda(Y)_j|^p \right)^{1/p}, \; 0 < p \leq 1$$

is known as the Schatten $p$-quasi-norm of $Y$.

When $p = 1$, it is called Nuclear norm.

# Schatten $p$-norm (Ji et al. 2013)

For any given symmetric matrix $Y \in S^n$,

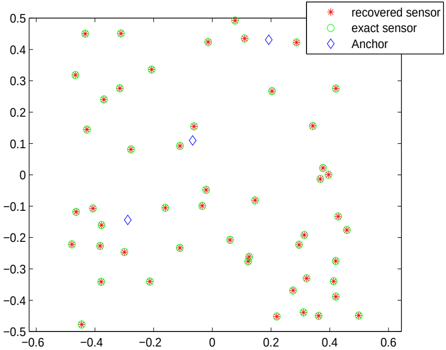
$$\|Y\|_p = \left( \sum_j |\lambda(Y)_j|^p \right)^{1/p}, \; 0 < p \leq 1$$

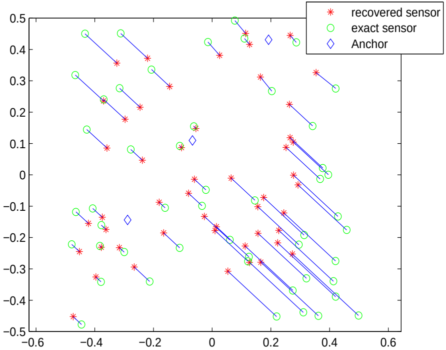
is known as the Schatten $p$-quasi-norm of $Y$.

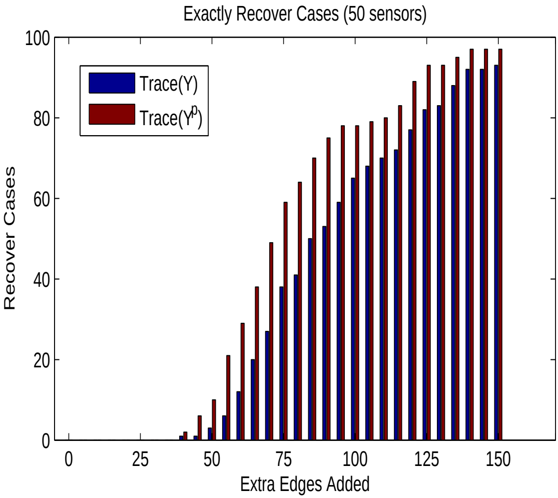
When $p = 1$, it is called Nuclear norm.

The Schatten $p$quasinorm has several nice analytical properties that make it a natural candidate for a regularizer.

# Recover Result: Schatten 0.5-Norm vs. Nuclear Norm




Exactly Recover Cases (50 sensors)
Trace(Y)
Trace(Y^p)
Recover Cases
Extra Edges Added
0
20
40
60
80
100
0
25
50
75
100
125
150

# Theory of $L_2$+$L_p$ Minimization I

**Theorem**

*(The first order bound) Let $\mathbf{x}^*$ be any first-order KKT point and let*

$$L_i = \left( \frac{\lambda p}{2\|\mathbf{a}_i\| \sqrt{f_{2p}(\mathbf{x}^*)}} \right)^{\frac{1}{1-p}}.$$

*Then*

$$\text{for any} \quad i \in \mathcal{N}, \qquad x_i^* \in (-L_i, L_i) \quad \Rightarrow \quad x_i^* = 0.$$

# Theory of $L_2$+$L_p$ Minimization I

Theorem

*(The first order bound) Let* $\mathbf{x}^*$ *be any first-order KKT point and let*

$$L_i = \left( \frac{\lambda p}{2\|\mathbf{a}_i\| \sqrt{f_{2p}(\mathbf{x}^*)}} \right)^{\frac{1}{1-p}}.$$

*Then*

$$\textit{for any} \quad i \in \mathcal{N}, \qquad x_i^* \in (-L_i, L_i) \quad \Rightarrow \quad x_i^* = 0.$$

"Lower Bound Theory of Nonzero Entries in Solutions of L2-Lp Minimization" (Chen, Xu and Y), SIAM J. Scientific Computing 32:5 (2010) 2832-2852.

Theorem

*(The second order bound) Let* $\mathbf{x}^*$ *be any second-order KKT point and let* $L_i = \left(\frac{\lambda p(1-p)}{2\|\mathbf{a}_i\|^2}\right)^{\frac{1}{2-p}}, i \in \mathcal{N}$. *Then*

(1)

$$\text{for any} \quad i \in \mathcal{N}, \qquad x_i^* \in (-L_i, L_i) \quad \Rightarrow \quad x_i^* = 0.$$

(2) *The support columns of* $\mathbf{x}^*$ *are linearly independent.*

# More Theoretical Developments ...

Markowitz Portfolio Model:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \mathbf{e}^T\mathbf{x} = 1,\ \mathbf{x} \geq \mathbf{0} \end{array}$$

# More Theoretical Developments ...

Markowitz Portfolio Model:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \mathbf{e}^T\mathbf{x} = 1, \ \mathbf{x} \geq \mathbf{0} \end{array}$$

Regularized Markowitz Portfolio Model:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} + \lambda\|\mathbf{x}\|_p^p \\ \text{subject to} & \mathbf{e}^T\mathbf{x} = 1, \ \mathbf{x} \geq \mathbf{0} \end{array}$$

# Linearly Constrained Optimization Problem

$$
\text{(LCOP)} \quad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}. \end{array}
$$

# Linearly Constrained Optimization Problem

$$\text{(LCOP)} \quad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}. \end{array}$$

The first-order KKT conditions:

$$\begin{array}{lcl} x_j(\nabla f(\mathbf{x}) - A^T\mathbf{y})_j & = & 0, \ \forall j \\ A\mathbf{x} & = & \mathbf{b} \\ \nabla f(\mathbf{x}) - A^T\mathbf{y} \geq \mathbf{0}, & & \mathbf{x} \geq \mathbf{0}. \end{array}$$

# Linearly Constrained Optimization Problem

$$\text{(LCOP)} \quad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \ge \mathbf{0}. \end{array}$$

The first-order KKT conditions:

$$\begin{array}{rcl} x_j(\nabla f(\mathbf{x}) - A^T\mathbf{y})_j & = & 0, \ \forall j \\ A\mathbf{x} & = & \mathbf{b} \\ \nabla f(\mathbf{x}) - A^T\mathbf{y} \ge \mathbf{0}, & & \mathbf{x} \ge \mathbf{0}. \end{array}$$

First-order $\epsilon$-KKT solution: $|x_j(\nabla f(\mathbf{x}) - A^T\mathbf{y})_j| \le \epsilon$ for all $j$.

# Linearly Constrained Optimization Problem

$$\text{(LCOP)} \quad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}. \end{array}$$

The first-order KKT conditions:

$$\begin{array}{lll} x_j(\nabla f(\mathbf{x}) - A^T\mathbf{y})_j & = & 0, \ \forall j \\ A\mathbf{x} & = & \mathbf{b} \\ \nabla f(\mathbf{x}) - A^T\mathbf{y} \geq \mathbf{0}, & & \mathbf{x} \geq \mathbf{0}. \end{array}$$

First-order $\epsilon$-KKT solution: $|x_j(\nabla f(\mathbf{x}) - A^T\mathbf{y})_j| \leq \epsilon$ for all $j$.
Second-order $\epsilon$-KKT solution if additionally the Hessian in the null space of active constraints is $\epsilon$-positive-semidefinite.

# Iteration Bound for an $\epsilon$-KKT Solution

| | Smooth | Lipschitz | Non-Lipschitz |
|---|---|---|---|
| $\log\log(\epsilon^{-1})$ | [Y 1992] Ball-IQP | | |
| $\epsilon^{-1}\log(\epsilon^{-1})$ | [Y 1998] IQP | | [Ge et al 2011] Constrained $L_p$ |
| $\epsilon^{-3/2}$ | [Nesterov et al 2006]; [Cartis et al 2011] | | [Bian et al 2012] |
| $\epsilon^{-2}$ | [Vavasis 1991], [Nesterov 2004]; [Gratton et al 2008] | [Vavasis 1991] [Cartis et al 2011] | [Bian et al 2012]; [Bian et al 2012] |
| $\epsilon^{-3}\log(\epsilon^{-1})$ | | [Garmanjani et al 2012] | |

Table: Selected worst-case complexity results for nonconvex optimization

# Ball or Sphere-Constrained Indefinite QP

$$\text{(BQP)} \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \|\mathbf{x}\|_2 = (\leq) 1. \end{array}$$

# Ball or Sphere-Constrained Indefinite QP

$$\text{(BQP)} \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \|\mathbf{x}\|_2 = (\leq) 1. \end{array}$$

The solution $\mathbf{x}$ of problem $(BQP)$ satisfies the following necessary and sufficient conditions (S-Lemma):

$$\begin{array}{lrcl} & (Q + \mu I)\mathbf{x} & = & -\mathbf{c}, \\ & (Q + \mu I) & \succeq & \mathbf{0}, \\ \text{and} & \|\mathbf{x}\| & = & 1. \end{array}$$

# Ball or Sphere-Constrained Indefinite QP

$$\text{(BQP)}\quad \begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \|\mathbf{x}\|_2 = (\leq) 1. \end{array}$$

The solution $\mathbf{x}$ of problem $(BQP)$ satisfies the following necessary and sufficient conditions (S-Lemma):

$$\begin{array}{rrcl} & (Q+\mu I)\mathbf{x} & = & -\mathbf{c}, \\ & (Q+\mu I) & \succeq & \mathbf{0}, \\ \text{and} & \|\mathbf{x}\| & = & 1. \end{array}$$

This is an SDP problem, and the simplest trust-region sub-problem (Moré, Sorensen, Dennis and Schnabel, etc. 1980).

# The Bisection Method

For any $\mu > -\underline{\lambda}(Q)$, where $\underline{\lambda}(Q)$ is the minimal eigenvalue of $Q$, denote by $\mathbf{x}(\mu)$ the solution to

$$(Q + \mu I)\mathbf{x} = -\mathbf{c}.$$

# The Bisection Method

For any $\mu > -\underline{\lambda}(Q)$, where $\underline{\lambda}(Q)$ is the minimal eigenvalue of $Q$, denote by $\mathbf{x}(\mu)$ the solution to

$$(Q + \mu I)\mathbf{x} = -\mathbf{c}.$$

Assume $\|\mathbf{x}(-\underline{\lambda}(Q))\| > 1$ (the othe case can be handled easily) and note

$$\mu \leq -\underline{\lambda}(Q) + \|\mathbf{c}\|.$$

# The Bisection Method

For any $\mu > -\underline{\lambda}(Q)$, where $\underline{\lambda}(Q)$ is the minimal eigenvalue of $Q$, denote by $\mathbf{x}(\mu)$ the solution to

$$(Q + \mu I)\mathbf{x} = -\mathbf{c}.$$

Assume $\|\mathbf{x}(-\underline{\lambda}(Q))\| > 1$ (the othe case can be handled easily) and note

$$\mu \leq -\underline{\lambda}(Q) + \|\mathbf{c}\|.$$

Thus, one can apply the bisection method to find the right $\mu$ and solve the problem in polynomial-time $\log(\epsilon^{-1})$ steps.

# The Bisection Method

For any $\mu > -\underline{\lambda}(Q)$, where $\underline{\lambda}(Q)$ is the minimal eigenvalue of $Q$, denote by $\mathbf{x}(\mu)$ the solution to

$$(Q + \mu I)\mathbf{x} = -\mathbf{c}.$$

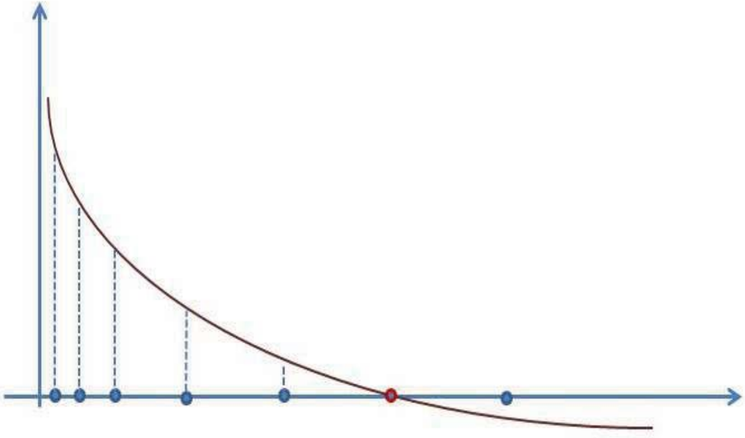
Assume $\|\mathbf{x}(-\underline{\lambda}(Q))\| > 1$ (the othe case can be handled easily) and note

$$\mu \le -\underline{\lambda}(Q) + \|\mathbf{c}\|.$$

Thus, one can apply the bisection method to find the right $\mu$ and solve the problem in polynomial-time $\log(\epsilon^{-1})$ steps.

We can do it in log-polynomial time using the Steve Smale 1986 work on Newton's method ...

# Combined Bisection and Newton's Method



# Potential Reduction Algorithm for LCOP

Consider the (concave+convex) Karmarkar potential function

$$\phi(\mathbf{x}) \quad = \rho \log(f(\mathbf{x})) - \sum_{j=1}^{n} \log x_j,$$

where we assume that $f(\mathbf{x})$ is nonnegative in the feasible region.

# Potential Reduction Algorithm for LCOP

Consider the (concave+convex) Karmarkar potential function

$$\phi(\mathbf{x}) \quad = \rho \log(f(\mathbf{x})) - \sum_{j=1}^{n} \log x_j,$$

where we assume that $f(\mathbf{x})$ is nonnegative in the feasible region. We start from the analytic center $\mathbf{x}^0$ of the feasible region–$\mathcal{F}_p$, so that if

$$\phi(\mathbf{x}^k) - \phi(\mathbf{x}^0) \leq \rho \log \epsilon, \tag{4}$$

$$\frac{f(\mathbf{x}^k)}{f(\mathbf{x}^0)} \leq \epsilon;$$

which implies that $\mathbf{x}^k$ is an $\epsilon$-global minimizer.

# Quadratic Over-Estimate of Potential Function I

Consider

$$f(\mathbf{x}) = q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x}.$$

# Quadratic Over-Estimate of Potential Function I

Consider

$$f(\mathbf{x}) = q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x}.$$

Given $\mathbf{0} < \mathbf{x} \in \mathcal{F}_p$, let $\Delta = q(\mathbf{x})$ and let $\mathbf{d}_x$, $A\mathbf{d}_x = \mathbf{0}$, be a vector such that $\mathbf{x}^+ := \mathbf{x} + \mathbf{d}_x > \mathbf{0}$. Then the non-convex part

$$\begin{aligned}
&\rho\log(q(\mathbf{x}^+)) - \rho\log(q(\mathbf{x})) \\
&= \rho\log(\Delta + \frac{1}{2}\mathbf{d}_x^T Q\mathbf{d}_x + (Q\mathbf{x}+\mathbf{c})^T\mathbf{d}_x) - \rho\log\Delta \\
&= \rho\log(1 + (\frac{1}{2}\mathbf{d}_x^T Q\mathbf{d}_x + (Q\mathbf{x}+\mathbf{c})^T\mathbf{d}_x)/\Delta) \\
&\leq \frac{\rho}{\Delta}(\frac{1}{2}\mathbf{d}_x^T Q\mathbf{d}_x + (Q\mathbf{x}+\mathbf{c})^T\mathbf{d}_x).
\end{aligned}$$

# Quadratic Over-Estimate of Potential Function II

On the other hand, if $\|X^{-1}\mathbf{d}_x\| \leq \beta < 1$, the convex part

$$-\sum_{j=1}^{n} \log(x_j^+) + \sum_{j=1}^{n} \log(x_j) \quad \leq \quad -\mathbf{e}^T X^{-1}\mathbf{d}_x + \frac{\beta^2}{2(1-\beta)}.$$

# Quadratic Over-Estimate of Potential Function II

On the other hand, if $\|X^{-1}\mathbf{d}_x\| \le \beta < 1$, the convex part

$$-\sum_{j=1}^{n} \log(x_j^+) + \sum_{j=1}^{n} \log(x_j) \quad \le \quad -\mathbf{e}^T X^{-1}\mathbf{d}_x + \frac{\beta^2}{2(1-\beta)}.$$

Thus, if $\|X^{-1}\mathbf{d}_x\| \le \beta < 1$, $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}_x > \mathbf{0}$ and

$$\phi(\mathbf{x}^+) - \phi(\mathbf{x}) \le \frac{\rho}{\Delta}\left(\frac{1}{2}\mathbf{d}_x^T Q\mathbf{d}_x + (Q\mathbf{x} + \mathbf{c} - \frac{\Delta}{\rho}X^{-1}\mathbf{e})^T\mathbf{d}_x\right) + \frac{\beta^2}{2(1-\beta)}. \tag{5}$$

# A Ball-Constrained Quadratic Subproblem I

We solve the following problem at the $k$th iteration:

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2}\mathbf{d}_x^T Q\mathbf{d}_x + (Q\mathbf{x}^k + \mathbf{c} - \tfrac{\Delta^k}{\rho}(X^k)^{-1}\mathbf{e})^T\mathbf{d}_x \\ \text{subject to} \quad & A\mathbf{d}_x = \mathbf{0}, \\ & \|(X^k)^{-1}\mathbf{d}_x\|^2 \le \beta^2. \end{aligned}$$

# A Ball-Constrained Quadratic Subproblem I

We solve the following problem at the $k$th iteration:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{d}_x^T Q\mathbf{d}_x + (Q\mathbf{x}^k + \mathbf{c} - \frac{\Delta^k}{\rho}(X^k)^{-1}\mathbf{e})^T\mathbf{d}_x \\ \text{subject to} & A\mathbf{d}_x = \mathbf{0}, \\ & \|(X^k)^{-1}\mathbf{d}_x\|^2 \le \beta^2. \end{array}$$

Using the affine-scaling, this problem can be reduced to the ball-constrained quadratic program, where the radius of the ball is $\beta$.

# Complexity Analysis

Each iteration can either make a constant reduction of the potential, or not.

In the latter case, the new iterate $\mathbf{x}^+$ becomes a second-order $\epsilon$-KKT solution with a suitable choice of $\rho$.

# Complexity Analysis

Each iteration can either make a constant reduction of the potential, or not.

In the latter case, the new iterate $\mathbf{x}^+$ becomes a second-order $\epsilon$-KKT solution with a suitable choice of $\rho$.

### Theorem

*Let $\beta = \frac{1}{3}$ and $\rho = \frac{3q(\mathbf{x}^0)}{\epsilon}$. Then the potential reduction algorithm returns a second-order $\epsilon$-KKT solution or global minimizer in no more than $O(\frac{q(\mathbf{x}^0)}{\epsilon} \log \frac{q(\mathbf{x}^0)}{\epsilon})$ iterations.*

# Complexity Analysis

Each iteration can either make a constant reduction of the potential, or not.

In the latter case, the new iterate $\mathbf{x}^+$ becomes a second-order $\epsilon$-KKT solution with a suitable choice of $\rho$.

**Theorem**

*Let $\beta = \frac{1}{3}$ and $\rho = \frac{3q(\mathbf{x}^0)}{\epsilon}$. Then the potential reduction algorithm returns a second-order $\epsilon$-KKT solution or global minimizer in no more than $O(\frac{q(\mathbf{x}^0)}{\epsilon} \log \frac{q(\mathbf{x}^0)}{\epsilon})$ iterations.*

This type of algorithm is called fully polynomial time approximation scheme.

# The PRA for Concave Minimization

The case when $f(\mathbf{x})$ being a concave function is even easier.

# The PRA for Concave Minimization

The case when $f(\mathbf{x})$ being a concave function is even easier.

Let $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}_x$. Then, we have a linear over-estimate of the potential function:

$$\phi(\mathbf{x}^+) - \phi(\mathbf{x}) \leq \left( \frac{\rho}{f(\mathbf{x})} \nabla f(\mathbf{x})^T - \mathbf{e}^T X^{-1} \right) \mathbf{d}_x + \frac{\beta^2}{2(1-\beta)},$$

as long as $\|X^{-1}\mathbf{d}_x\| \leq \beta$.

# The PRA for Concave Minimization

The case when $f(\mathbf{x})$ being a concave function is even easier.

Let $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}_x$. Then, we have a linear over-estimate of the potential function:

$$\phi(\mathbf{x}^+) - \phi(\mathbf{x}) \le \left( \frac{\rho}{f(\mathbf{x})} \nabla f(\mathbf{x})^T - \mathbf{e}^T X^{-1} \right) \mathbf{d}_x + \frac{\beta^2}{2(1-\beta)},$$

as long as $\|X^{-1}\mathbf{d}_x\| \le \beta$.

Let affine-scaling $\mathbf{d}' = X^{-1}\mathbf{d}_x$. Then, one can solve:

$$\begin{array}{rcc} z(\mathbf{d}') := & \text{Minimize} & \left( \frac{\rho}{f(\mathbf{x})} \nabla f(\mathbf{x})^T X - \mathbf{e}^T \right) \mathbf{d}' \\ & \text{Subject to} & AX\mathbf{d}' = \mathbf{0} \\ & & \|\mathbf{d}'\|^2 \le \beta^2. \end{array}$$

# Affine Scaling Direction

The optimal direction of the affine scaling sub-problem is given by

$$\mathbf{d}' = \frac{\beta}{\|\mathbf{p}(\mathbf{x})\|} \cdot \mathbf{p}(\mathbf{x}),$$

where

$$\begin{aligned} \mathbf{p}(\mathbf{x}) &= -\left(I - XA^T(AX^2A^T)^{-1}AX\right)\left(\tfrac{\rho}{f(\mathbf{x})}X\nabla f(\mathbf{x}) - \mathbf{e}\right) \\ &= \mathbf{e} - \tfrac{\rho}{f(\mathbf{x})}X\left(\nabla f(\mathbf{x}) - A^T\mathbf{y}\right). \end{aligned}$$

# Affine Scaling Direction

The optimal direction of the affine scaling sub-problem is given by

$$\mathbf{d}' = \frac{\beta}{\|\mathbf{p}(\mathbf{x})\|} \cdot \mathbf{p}(\mathbf{x}),$$

where

$$\begin{aligned} \mathbf{p}(\mathbf{x}) &= -\left(I - XA^T(AX^2A^T)^{-1}AX\right)\left(\tfrac{\rho}{f(\mathbf{x})}X\nabla f(\mathbf{x}) - \mathbf{e}\right) \\ &= \mathbf{e} - \tfrac{\rho}{f(\mathbf{x})}X\left(\nabla f(\mathbf{x}) - A^T\mathbf{y}\right). \end{aligned}$$

And the minimal value of the sub-problem

$$z(\mathbf{d}') = -\beta \cdot \|\mathbf{p}(\mathbf{x})\|.$$

# Complexity Analysis I

If $\|\mathbf{p}(\mathbf{x})\| \geq 1$, then the minimal objective value of the affine scaling sub-problem is less than $\beta$ so that

$$\phi(\mathbf{x}^+) - \phi(\mathbf{x}) < -\beta + \frac{\beta^2}{2(1-\beta).}$$

Thus, the potential value is reduced by a constant for choosing a suitable $\beta$.

# Complexity Analysis I

If $\|\mathbf{p}(\mathbf{x})\| \geq 1$, then the minimal objective value of the affine scaling sub-problem is less than $\beta$ so that

$$\phi(\mathbf{x}^+) - \phi(\mathbf{x}) < -\beta + \frac{\beta^2}{2(1-\beta).}$$

Thus, the potential value is reduced by a constant for choosing a suitable $\beta$.

If this case would hold for $O(\rho \log \frac{f(\mathbf{x}^0)}{\epsilon})$ iterations, we would have produced an $\epsilon$-global minimizer of LCOP.

# Complexity Analysis II

On the other hand, if $\|\mathbf{p}(\mathbf{x})\| < 1$, from

$$\mathbf{p}(\mathbf{x}) = \mathbf{e} - \frac{\rho}{f(\mathbf{x})} X \left( \nabla f(\mathbf{x}) - A^T \mathbf{y} \right),$$

we must have

$$\frac{\rho}{f(\mathbf{x})} X \left( \nabla f(\mathbf{x}) - A^T \mathbf{y} \right) \geq \mathbf{0}$$

and

$$\frac{\rho}{f(\mathbf{x})} X \left( \nabla f(\mathbf{x}) - A^T \mathbf{y} \right) \leq 2\mathbf{e}.$$

# Complexity Analysis III

In other words
$$\left(\nabla f(\mathbf{x}) - A^T\mathbf{y}\right) \geq \mathbf{0}$$
and
$$x_j \left(\nabla f(\mathbf{x}) - A^T\mathbf{y}\right)_j < \frac{2f(\mathbf{x})}{\rho}, \ \forall j.$$

# Complexity Analysis III

In other words

$$\left(\nabla f(\mathbf{x}) - A^T\mathbf{y}\right) \geq \mathbf{0}$$

and

$$x_j\left(\nabla f(\mathbf{x}) - A^T\mathbf{y}\right)_j < \frac{2f(\mathbf{x})}{\rho},\ \forall j.$$

The first condition indicate that the Lagrange multiplier $\mathbf{y}$ is valid, and the second inequality implies that the complementarity condition is approximately satisfied when $\rho$ chosen sufficiently large.

# Complexity Analyses IV

In particular, if we choose $\rho \geq \frac{2f(\mathbf{x}^0)}{\epsilon}$, then

$$\|X(\nabla f(\mathbf{x}) - A^T\mathbf{y})\|_\infty \leq \epsilon,$$

which implies that $\mathbf{x}$ is a first-order $\epsilon$-KKT solution.

# Complexity Analyses IV

In particular, if we choose $\rho \geq \frac{2f(\mathbf{x}^0)}{\epsilon}$, then

$$\|X(\nabla f(\mathbf{x}) - A^T \mathbf{y})\|_\infty \leq \epsilon,$$

which implies that $\mathbf{x}$ is a first-order $\epsilon$-KKT solution.

Theorem

*The algorithm then will provably return a first-order $\epsilon$-KKT solution of LCOP in no more than $O(\frac{f(\mathbf{x}^0)}{\epsilon} \log \frac{f(\mathbf{x}^0)}{\epsilon})$ iterations for any given $\epsilon < 1$, if the objective function is concave.*

# More Applications and Questions

- Could the time bound be further improved for QP or concave minimization?

# More Applications and Questions

- Could the time bound be further improved for QP or concave minimization?
- Would the $O(\epsilon^{-1} \log \epsilon^{-1})$ bound applicable to more general non-convex optimization problems?

# More Applications and Questions

- Could the time bound be further improved for QP or concave minimization?
- Would the $O(\epsilon^{-1} \log \epsilon^{-1})$ bound applicable to more general non-convex optimization problems?
- We are not even able to prove the $O(\epsilon^{-1} \log \epsilon^{-1})$ bound when

$$f(\mathbf{x}) = q(\mathbf{x}) + \|\mathbf{x}\|_p^p$$

.

# More Applications and Questions

- Could the time bound be further improved for QP or concave minimization?
- Would the $O(\epsilon^{-1}\log\epsilon^{-1})$ bound applicable to more general non-convex optimization problems?
- We are not even able to prove the $O(\epsilon^{-1}\log\epsilon^{-1})$ bound when

$$f(\mathbf{x}) = q(\mathbf{x}) + \|\mathbf{x}\|_p^p$$

.

- More structural properties on the final KKT solution.

# More Applications and Questions

- Could the time bound be further improved for QP or concave minimization?
- Would the $O(\epsilon^{-1} \log \epsilon^{-1})$ bound applicable to more general non-convex optimization problems?
- We are not even able to prove the $O(\epsilon^{-1} \log \epsilon^{-1})$ bound when

$$f(\mathbf{x}) = q(\mathbf{x}) + \|\mathbf{x}\|_p^p$$

.

- More structural properties on the final KKT solution.
- Applications to general sparse solution optimization, such as the cardinality constrained portfolio selection, sparse pricing reduction for revenue management, etc..