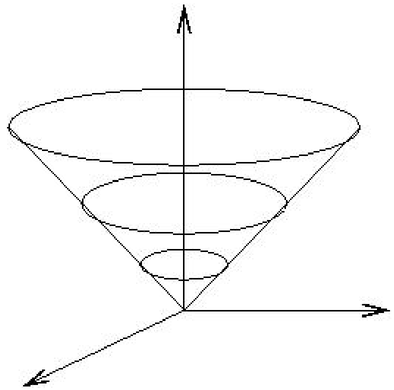
Coordinatewise Distributed Methods for Large Scale Convex Optimization

Paul Tseng
Mathematics, University of Washington
Seattle

ICCOPT, McMaster University
August 16, 2007

Talk Outline

- Sensor network localization and SDP, SOCP, ESDP relaxations
- Distributed methods for SOCP and ESDP relaxations
- Distributed method for TV-based image restoration
- Extensions



Sensor Network Localization

Basic Problem:

- n pts in $\Re^d$ ($d = 1, 2, 3$).

- Know last $n - m$ pts ('anchors') $x_{m+1}, ..., x_n$ and Eucl. dist. estimate for pairs of 'neighboring' pts

$$d_{ij} \geq 0 \quad \forall (i, j) \in \mathcal{A}$$

  with $\mathcal{A} \subseteq \{(i, j) : 1 \leq i < j \leq n\}$.

- Estimate first m pts ('sensors').

History? Graph realization, position estimation in wireless sensor network, determining protein structures, ...

Optimization Problem Formulation

$$\upsilon_{\text{opt}} := \min_{x_1,\ldots,x_m} \sum_{(i,j)\in\mathcal{A}} \left| \|x_i - x_j\|^2 - d_{ij}^2 \right|^2$$

- Objective function is smooth but nonconvex. m can be large ($m > 1000$).
- Problem is NP-hard (reduction from PARTITION).
- Use a convex (SDP, SOCP) relaxation. High soln accuracy unnecessary.
- Seek “simple” distributed methods (important for practical implementation).

SDP Relaxation

Let $X := [x_1 \ \cdots \ x_m], \quad A := [x_{m+1} \ \cdots \ x_n]$. Then

$$\begin{aligned} \upsilon_{\text{opt}} = \quad & \min_{X,Y} \sum_{(i,j)\in\mathcal{A}} \left|\operatorname{tr}\left(b_{ij}b_{ij}^T Z\right) - d_{ij}^2\right|^2 \\ & \text{s.t.} \quad Z = \begin{bmatrix} Y & X^T \\ X & I_d \end{bmatrix} \succeq 0, \quad \operatorname{rank} Z = d \end{aligned}$$

with $b_{ij} := \begin{bmatrix} I_m & 0 \\ 0 & A \end{bmatrix} (e_i - e_j)$.

SDP relaxation (Biswas,Ye '03):

$$\begin{aligned} \upsilon_{\text{sdp}} := \quad & \min_{X,Y} \sum_{(i,j)\in\mathcal{A}} \left|\operatorname{tr}\left(b_{ij}b_{ij}^T Z\right) - d_{ij}^2\right|^2 \\ & \text{s.t.} \quad Z = \begin{bmatrix} Y & X^T \\ X & I_d \end{bmatrix} \succeq 0 \end{aligned}$$

However, SDP relaxation is expensive to solve for m large..

SOCP Relaxation

$$\begin{aligned} \upsilon_{\text{opt}} = \quad & \min_{x_1,\ldots,x_m,y_{ij}} \sum_{(i,j)\in\mathcal{A}} \left| y_{ij} - d_{ij}^2 \right|^2 \\ & \text{s.t.} \quad y_{ij} = \|x_i - x_j\|^2 \quad \forall (i,j) \in \mathcal{A} \end{aligned}$$

Relax "$=$" to "$\geq$" constraint:

$$\begin{aligned} \upsilon_{\text{socp}} := \quad & \min_{x_1,\ldots,x_m,y_{ij}} \sum_{(i,j)\in\mathcal{A}} \left| y_{ij} - d_{ij}^2 \right|^2 \\ & \text{s.t.} \qquad y_{ij} \geq \|x_i - x_j\|^2 \quad \forall (i,j) \in \mathcal{A} \\ = \quad & \min_{x_1,\ldots,x_m} f(x_1,\ldots,x_m) := \sum_{(i,j)\in\mathcal{A}} \max\{0, \|x_i - x_j\|^2 - d_{ij}^2\}^2 \end{aligned}$$

This is an unconstrained problem, with f smooth, convex, partially separable.

Solve using a coordinate gradient descent (CGD) method (T, Yun '06):

- If $\|\nabla_{x_i} f\| \geq \mathrm{tol}$, then update x_i by moving it along $-H_i^{-1}\nabla_{x_i} f$, with $H_i \succ 0$ and stepsize by Armijo rule to decrease f, and re-iterate.

Computation is cheap and distributes. Only $\{x_j\}_{(i,j)\in\mathcal{A}}$ are needed to update x_i. Provable global convergence. Fast convergence in practice.

However, SOCP can be significantly weaker than SDP relaxation..

ESDP Relaxation

Idea: Further relax the constraint $Z \succeq 0$ in SDP relaxation.

ESDP relaxation (Wang, Zheng, Boyd, Ye '06):

$$
\begin{aligned}
\upsilon_{\text{esdp}} := \quad & \min_{X,Y} \quad \sum_{(i,j)\in\mathcal{A}} \left| \text{tr}\left(b_{ij}b_{ij}^T Z\right) - d_{ij}^2 \right|^2 \\
& \text{s.t.} \quad Z = \begin{bmatrix} Y & X^T \\ X & I_d \end{bmatrix} \\
& \begin{bmatrix} Y_{ii} & Y_{ij} & x_i^T \\ Y_{ij} & Y_{jj} & x_j^T \\ x_i & x_j & I_d \end{bmatrix} \succeq 0 \quad \forall (i,j)\in\mathcal{A} \text{ with } j \le m \\
& \begin{bmatrix} Y_{ii} & x_i^T \\ x_i & I_d \end{bmatrix} \succeq 0 \qquad \forall (i,j)\in\mathcal{A} \text{ with } j > m
\end{aligned}
$$

ESDP is stronger than SOCP, weaker than SDP relaxation. In simulation, ESDP is nearly as strong as SDP relaxation, and solvable much faster by SeDuMi. Distributed method?

Distributed Method for Partially Separable SDP

ESDP has the partially separable form

$$\min_z \quad h(z) := \sum_{k=1}^{K} h_k(z) \quad \text{s.t.} \quad A_k z + B_k \succeq 0,\ k = 1, ..., K$$

with A_k very sparse, B_k low-dim., and h_k convex, C^2, with $\nabla^2 h_k$ of the same sparsity pattern as A_k.

KKT Optimality conditions:

$$\nabla h(z) - \sum_k A_k^* \Lambda_k = 0,$$
$$0 \preceq \Lambda_k \perp A_k z + B_k \succeq 0,\ k = 1, ..., K$$

Unconstrained reformulation:

$$\min_{z,\Lambda} \quad f(z,\Lambda) := \sum_k \psi_{\rm FB}(A_k z + B_k, \Lambda_k) + \|\nabla h(z) - \sum_k A_k^* \Lambda_k\|^2$$

with

$$\psi_{\rm FB}(X,Y) = \|(X^2+Y^2)^{1/2} - X - Y\|_F^2.$$

Facts: (T '98, Sim, Sun, Ralph '06)

- f is smooth, partially separable, nonneg.
- If KKT soln exists, then (z,Λ) is KKT soln $\Longleftrightarrow \nabla f(z,\Lambda) = 0$.

Solvable by many methods, but most update all variables at once.
CGD-based distributed method:

- Choose a "small" subset of variables w of (z,Λ). If $\|\nabla_w f\| \geq \mathrm{tol}$, then move w along $-H^{-1}\nabla_w f$, with $H \succ 0$ and stepsize by Armijo rule to decrease f, and re-iterate.

TV-Based Image Restoration

Total variation-based problem for restoring a noisy image b on $\Omega \subset \Re^2$: (Rudin, Osher, Fatemi '92)

$$\min_u \int_\Omega \|\nabla u\| dx + \lambda \int_\Omega |b - u|^2 dx$$

Dual has form:

$$\min_w f(w) := \int_\Omega |\nabla \cdot w - \lambda b|^2 dx \quad \text{s.t.} \quad \|w\| \le 1 \text{ a.e. on } \Omega.$$

When discretized on a grid, reduces to minimizing a convex, partially separable quad. func. of $w_{ij} \in \Re^2$ subject to $\|w_{ij}\| \le 1$.

CGD-based distributed method:

- If $\|d_{ij}\| \geq \mathrm{tol}$, where

$$d_{ij} := \underset{\|w_{ij}+d\|\leq 1}{\arg\min} \, (\nabla_{w_{ij}} f)^T d + \frac{1}{2} d^T H_{ij} d$$

  with $H_{ij} \succ 0$, then move w_{ij} along d_{ij} with stepsize by Armijo rule to decrease f, and re-iterate.

If H_{ij} is a multiple of I_2, then d_{ij} has closed form solution.

Extensions

- Partially asynchronous computation, with constant stepsize?
- Simulation and numerical testing?
- Modifications to find a relative interior soln of ESDP?