Semidefinite and Second Order Cone Programming Seminar Final Project Part I

Wang Yao and Gyorgy Matyasfalvi
Instructor: Farid Alizadeh
RUTCOR, Rutgers University

December 3, 2012

Abstract

This project is dedicated to present the main ideas of [1]. We consider a Newton-CG augmented Lagrangian method for solving semidefinite programming (SDP) problems from the perspective of approximate semismooth Newton methods. For the inner problems, we show that the positive definiteness of the generalized Hessian of the objective function in these inner problems is equivalent to the constraint nondegeneracy of the corresponding dual problems. Numerical experiments on a variety of large-scale SDP problems with the matrix dimension n up to 4110 and the number of equality constraints m up to $2,156,544$ show that the proposed method is very efficient.

1 Introduction

Let $\mathcal{S}^n$ be the linear space of all $n \times n$ symmetric matrices and $\mathcal{S}^n_+$ be the cone of all $n \times n$ symmetric positive semidefinite matrix. The notation $X \succcurlyeq \mathbf{0}$ means that X is a symmetric positive semidefinite matrix. We study an augmented Lagrangian method for solving the following SDP problem:

$$\min \left\{ b^{\mathrm{T}} y \mid \mathcal{A}^* y - C \succcurlyeq \mathbf{0} \right\} \qquad \text{(D)}$$

where $C \in \mathcal{S}^n$, $b \in \mathbb{R}^m$, $\mathcal{A}$ is linear operator from $\mathcal{S}^n$ to $\mathbb{R}^m$ and $\mathcal{A}^* : \mathbb{R}^m \to \mathcal{S}^n$. Its dual is

$$\max \{\langle C, X\rangle \mid \mathcal{A}(X) = b, X \succcurlyeq \mathbf{0}\} \quad \text{(P)}$$

Given a penalty parameter $\sigma > 0$, the *augmented Lagrangian* function for (D) is

$$L_\sigma(y, X) = b^{\mathrm{T}} y + \frac{1}{2\sigma}\left(\left\|\Pi_{\mathcal{S}^n_+}\left(X - \sigma(\mathcal{A}^* y - C)\right)\right\|^2 - \|X\|^2\right)$$

where $(y, X) \in \mathbb{R}^m \times \mathcal{S}^n$. Since $\|\Pi(\cdot)\|^2$ is continuously differentiable [2], for any given $X \in \mathcal{S}^n$, we have

$$\nabla_y L_\sigma(y, X) = b - \mathcal{A}\Pi_{\mathcal{S}^n_+}\left(X - \sigma(\mathcal{A}^* y - C)\right)$$

Throughout the presentation, the following condition for (P) is assumed to hold.

ASSUMPTION 1. *Problem (P) satisfies the condition*

$$\begin{cases} \mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m \textit{is onto}, \\ \exists X_0 \in \mathcal{S}^n_+ \quad \textit{such that} \quad \mathcal{A}(X_0) = b, X_0 \succ \mathbf{0} \end{cases}$$

The remaining parts of this article are as follows. In Section 2, we give some preliminaries including a brief introduction about concepts related to the method of multipliers. In Section **??**, we introduce a semismooth Newton-CG method for solving the inner optimization problems. In Section 4, we shall give the complete algorithm. We list some numerical results in Section **??**. Finally we conclude this paper in Section **??**.

2 Preliminaries

It is well known that conjugate gradient method can solve the linear system

$$Ax = b$$

where $b \in \mathbb{R}^m$ and A is assumed to be a symmetric positive definite matrix. We will use a practical conjugate gradient algorithm in implementation.

Assume $f(x)$ is convex and twice continuously differentiable so its Hessian $\nabla^2 f(x)$ is positive definite. A typical iteration of Newton's method as follows:

$$x^{k+1} = x^k - \alpha_k [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$$

where α_k is step size of kth iteration. In practical, we do not directly compute $[\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$, instead, we solve

$$\nabla^2 f(x^k) d^k = -\nabla f(x^k)$$

with a practical CG method.

For given $X^0 \in \mathcal{S}^n, \sigma_0 > 0$ and $\rho \geq 1$ the augmented Lagrangian method for solving (D) and (P)generates sequences $\{y^k\} \subset \mathrm{R}^m$ and $\{X^k\} \subset \mathcal{S}^n$ as follows:

$$\begin{cases} y^{k+1} \approx \arg\min\limits_{y \in \mathbb{R}^m} L_{\sigma_k}(y, X), \leftarrow \text{inner problem} \\ X^{k+1} = \Pi_{\mathcal{S}^n_+} \left(X^k - \sigma_k (\mathcal{A}^* y^{k+1} - C) \right), \quad k = 0, 1, 2, \ldots, \\ \sigma_{k+1} = \rho \sigma_k \end{cases}$$

3 Solving the inner problem

We focus on inner problem, i.e.

$$\min \{ \varphi(y) = L_\sigma(y, X) \mid y \in \mathbb{R}^m \}$$

We know that $\varphi(y)$ is not twice continuously differentiable, but we can develop locally a Newton-CG method [3] to solve

$$\nabla \varphi(y) = b - \mathcal{A} \Pi_{\mathcal{S}^n_+} \left(X - \sigma (\mathcal{A}^* y - C) \right) = 0$$

Since $\Pi_{\mathcal{S}^n_+}$ is nonexpansive, it is Lipschitz continuous with modulus 1. Thus the mapping $\nabla \varphi(y)$ is Lipschitz continuous on $\mathbb{R}^m$. By Rademacher's theorem, $\nabla \varphi(y)$ is almost everywhere differentiable in $\mathbb{R}^m$.

Define the generalized Hessian of φ at y as

$$\hat{\partial}^2 \varphi(y) := \sigma \mathcal{A} \partial \Pi_{\mathcal{S}^n_+} \left(X - \sigma (\mathcal{A}^* y - C) \right) \mathcal{A}^*$$

where ∂ denotes the Clarke's generalized Jacobian [4]. To carry out the algorithm, we need to actually compute one element of $\hat{\partial}^2 \varphi(y)$. First since $X - \sigma(\mathcal{A}^* y - C) \in \mathcal{S}^n$, there exists an orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ such that

$$X - \sigma(\mathcal{A}^* y - C) = Q \Gamma_y Q^{\mathrm{T}}$$

where $\Gamma_y = \mathrm{diag}(\lambda_{[1]}, \lambda_{[2]}, \ldots, \lambda_{[n]})$ The recent progress shows that this spectral decomposition can be computed with a new iterative method [6]. the part II of this project cover the issue.

Define index set $\alpha := \{i \mid \lambda_i > 0\}$ and $\bar{\alpha} = \{1, 2, \ldots, n\} \backslash \alpha$ and

$$\Omega = \begin{bmatrix} E_{\alpha\alpha} & \nu_{\alpha\bar{\alpha}} \\ \nu_{\alpha\bar{\alpha}}^{\mathrm{T}} & \mathbf{0} \end{bmatrix}, \quad \nu_{\alpha\bar{\alpha}} := \frac{\lambda_i}{\lambda_i - \lambda_i}, \quad i \in \alpha, j \in \bar{\alpha}$$

where $E_{\alpha\alpha} \in \mathcal{S}^{|\alpha|}$ is the matrix of 1's Define the operator $W_y^0 : \mathcal{S}^n \to \mathcal{S}^n$ by

$$W_y^0(H) := Q\left(\Omega \circ \left(Q^{\mathrm{T}}(H)Q\right)\right)Q^{\mathrm{T}}, \quad H \in \mathcal{S}^n$$

where $\circ$ denotes the entry-wise product of two matrices Since, by [7], Lemma 11,

$$W_y^0 \in \partial \Pi_{\mathcal{S}_+^n}\left(X - \sigma(\mathcal{A}^* y - C)\right)$$

If we define $V_y^0 : \mathbb{R}^m \to \mathcal{S}^n$ by

$$V_y^0 d := \sigma \mathcal{A} \underbrace{\left[Q\left(\Omega \circ \left(Q^{\mathrm{T}}(\mathcal{A}^* d)Q\right)\right)Q^{\mathrm{T}}\right]}_{W_y^0(\mathcal{A}^* d)}, \quad d \in \mathbb{R}^m$$

thus we know that

$$V_y^0 = \sigma \mathcal{A} W_y^0 \mathcal{A}^* \in \hat{\partial}^2 \varphi(y)$$

It can be shown that under the primal constraint nondegenerate condition [5], every element in $\hat{\partial}^2 \varphi(y)$ is positive definite. Now we are ready to describe the algorithm for solving inner problem:

Data: $\mathcal{A}, C, b$
Result: return approximate minimizer of $L_{\sigma_k}(y, X^k)$
initialization;
while *the termination criteria are not satisfied* **do**
Step 1. Apply the practical CG algorithm to find solution d^j to

$$(V_j + \epsilon I)d = -\nabla \varphi(y)$$

where $V_j \in \hat{\partial}^2 \varphi(y^j)$;
Step 2. Choose step size α_j with Armijo's Rule;
Step 3. Update all the parameters and set

$$y^{j+1} = y^j + \alpha_j d^j$$

end
Algorithm 1: Solving inner problem at each outer iteration

We shall emphasize the termination criterion here. The old, e.g. *absolute summable* error criteria:

$$\varphi(y^j) - \inf \varphi \leq \epsilon_j, \quad \epsilon_k \geq 0, \quad \sum_0^\infty \epsilon < \infty$$

It is not practical because no direct rules provided for choosing ϵ_k. The new *relative error criterion* developed by [8] as follow:

$$\frac{2}{\sigma_k} \left| \left\langle w^{j-1} - y^j, \nabla_y \varphi(y^j) \right\rangle \right| + \|\nabla_y \varphi(y^j)\|^2 \leq$$
$$\tau \left(\min \left\{ \frac{1}{\sigma_k} X^{j-1}, (\mathcal{A}^* y^j - C) \right\} \right) \quad \tau \in (0,1)$$
$$w^j = w^{j-1} - \sigma_k \nabla_y \varphi(y^j) \quad \leftarrow \quad \text{update parameter}$$

Note that σ_k is fixed within inner problem. The new relative error criterion is practical because τ is easy to choose and all other information is known.

4 A Newton-CG augmented Lagrangian method

Now we integrated algorithm for inner problem into augmented Lagrangian method:

Data: $\mathcal{A}, C, b$
Result: return approximate optimal solutions of (D) and (P)
initialization;
while *the KKT conditions are not satisfied* **do**
  Step 1. Apply Algorithm 1 to get y^{k+1}, approximate solution for inner problem Step 2. $X^{k+1} = \Pi_{\mathcal{S}_+^n} \left(X^k - \sigma_k(\mathcal{A}^* y^{k+1} - C) \right)$ Step 3. $\sigma_{k+1} = \rho \sigma_k \qquad \rho \geq 1$
end

Algorithm 2: A Newton-CG augmented Lagrangian method for solving (D) and (P)

5 Highlights of numerical results

- SDP from frequency assignment problems (faq): Algorithm 2 takes only 41 seconds to solve faq09 and inexact IPM takes more than 2.5 hours.

- Able to solve faq36 in the 7th DIMACS Implementation Challenge faster and much more accurately than previous attempts
- SDP from maximum stable set problems: about 5-10 times faster than primal-dual IPM and modified barrier method.
- Binary integer quadratic programming problems: compared with dedicated augmented Lagrangian method with lift-and-project procedure (coded in C). Algorithm 2 is superior in terms of CPU time and the accuracy of the approximate optimal solution computed.

6 Conclusions

We introduce a Newton-CG augmented Lagrangian algorithm for solving (D) and (P). Convergence analysis is based on classical results of prximal point methods along with recent developments in perturbation analysis. Extensive numerical experiments demonstrated that the algorithm is very efficient on large-scale SDP problems. Application within a branch-and-bound algorithm for solving hard combinatorial problems.

References

[1] Xinyuan Zhao, Defeng Sun and Kim-Chuan Toh, *A Newton-CG Augmented Lagrangian Method for Semidefinite Programming.* SIAM J. OPTIM., Vol. 20, No. 4, pp. 1737-1765.

[2] E. H. Zarantonello, On this and that. *Projections on convex sets in Hilbert space and spectral theory* I *and* II, in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, New York, 1971, pp. 237-341 and 343-424.

[3] D.F. Sun, and J. Sun, *Semismooth matrix valued functions*, Math. Oper. Res., 27(2002), pp. 150-169.

[4] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983 pp. 8.

[5] F. Alizadeh, J. P. A. Haeberly, and O. L. Overton, *Complementarity and nondegeneracy in semidefinite programming*, Math. Programming, 77(1997), pp. 111-128.

[6] Siming Wei, Zhouchen Lin, *Accelerating Iterations Involving Eigenvalue or Singular Value Decomposition by Block Lanczos with Warm Start*, Microsoft Technical Report No. MSR-TR-2010-162.

[7] J. S. Pang, D. F. Sun, and J. Sun, *Semismooth homeomorphisms and strong stability of semidefinite and lorentz complementarity problems*, Math. Oper. Res., 28(2003), pp.272-284.

[8] J. Eckstein, and P. J.S. Silva, *A Practical Relative Error Criterion for Augmented Lagrangians*, Math. Programming, DOI 10.1007/s10107-012-0528-9 (online March 2012).