Convex Optimization : Conic Versus Functional Form

Erling D. Andersen
MOSEK ApS,
Fruebjergvej 3, Box 16, DK 2100 Copenhagen,
Blog: http://erlingdandersen.blogspot.com
Linkedin: http://dk.linkedin.com/in/edandersen
Email: e.d.andersen@mosek.com
WWW: http://www.mosek.com

Introduction

Topic

Given a convex optimization problem then is the **classical** form

$$\begin{array}{lrcl} \min & c(x) & & \\ \text{st} & a(x) & \geq & 0 \end{array}$$

or the **conic** form

$$\begin{array}{lrcl} \min & c^T x & & \\ \text{st} & Ax & = & b, \\ & x \in K. & & \end{array}$$

preferable from an user perspective?

Assumptions

- $c(x)$ is convex and $a(x)$ is concave.
- K is a convex cone.
- We can only solve convex problems efficiently.

Considerations about the forms

- Robustness.
  - Accuracy of the solution.
  - How sensitive is the optimizer to the problem data.
- Solution time.
- Does the form help the user to build a good model.
  - Does it prevent nonconvex problems because checking convexity is hard.
- Ease of use.

mosek http://www.mosek.com

Conic optimization. A recap

Primal problem:

$$\begin{array}{llll} (CO_P) & \min & c^T x & \\ & \text{st} & Ax & = & b, \\ & & x \in \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_k, & \end{array}$$

where each $\mathcal{K}_k$ is a cone (closed, pointed and solid).
Dual problem:

$$\begin{array}{llll} (CO_P) & \max & b^T y & \\ & \text{st} & A^T y + s & = & c, \\ & & s \in \mathcal{K}_1^* \times \mathcal{K}_2^* \times \cdots \times \mathcal{K}_k^*, & \end{array}$$

where each $\mathcal{K}_k^*$ is the dual convex cone i.e.

$$\mathcal{K}_k^* \equiv \{s : \; x^T s \geq 0, \forall x \in \mathcal{K}\}.$$

Basic cone types

- Linear:

$$\mathcal{K}_l := \{x \in R : \; x \geq 0\}$$

- Quadratic:

$$\mathcal{K}_q := \left\{ x \in R^n : \; x_1 \geq \sqrt{\sum_{j=2}^{n} x_j^2} \right\}$$

Comments

- We will only deal with the above 2 **self-dual** cones.
- A quote from Stephen Boyd of Stanford: "Almost all convex optimization problem can be formulated using a combination of linear, quadratic, semi-definite and the exponential cones".
- See also the book of Ben-Tal and Nemirovski for conic quadratic representable sets.

An illustrative example

Convex quadratic optimization

Functional formulation:

$$\begin{array}{lll} \min & \frac{1}{2}x^TQx + c^Tx & \\ \text{st} & Ax & = & b, \\ & x \geq 0. & \end{array}$$

Given convexity then there exists a H such that

$$Q = HH^T.$$

Conic reformulation

$$\begin{array}{llll} \min & \frac{1}{2}t + c^Tx & & \\ \text{st} & Ax & = & b, \\ & H^Tx - y & = & 0, \\ & s & = & 1, \\ & ||y||^2 \leq 2st, & & \\ & x \geq 0. & & \end{array}$$

example continued

- Conic reformulation requires more constraints and variables.
- The addition is highly structured and sparse.
- Conic form is convex by construction.
  - Q is frequently negative semi-definite due to rounding errors or mistakes. How to deal with:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1-\epsilon \end{bmatrix}?$$

- In practice Q is frequently low rank and dense. Implying the conic reformulation requires much less space to store.

Practical issues

- Many users of the nonlinear capabilities in MOSEK:
  - Do believe their nonlinear problems are convex.
  - Do believe convexity is not important.
  - Almost always make mistakes. (Errors in function or gradient computations).
- Checking and verifying convexity is an issue even for quadratic problems.
- Checking convexity and smoothness for arbitrary black box functions are hard.
- Data for the conic form is very simple. Just vectors, matrices and list of indexes.

Algorithms

A comparison

- The available algorithms determines which form is preferable.
- Only primal-dual interior-point algorithms are considered.
- They are efficient in practice.
- Polynomial complexity holds:
  - For some problems on functional form.
  - All conic quadratic problems.
- Claim: The algorithm for problems on conic form is more efficient and robust than problems on functional form.
- Justification of claim: Follows.

A primal-dual algorithm for problems on functional form

Problem:

$$\begin{array}{lrcl} \min & c(x) & & \\ \text{st} & a_i(x) & \geq & 0, \quad \forall i, \\ & x \geq 0. & & \end{array}$$

Assumptions:

1. $x \in \mathbf{R}^n$
2. c is convex and twice differentiable.
3. a_i is concave and twice differentiable.

Langrange function

Lagrange function

$$L(x,y) \equiv c(x) - y^T a(x),$$

where $y \geq 0$.
The dual problem:

$$\begin{array}{lllll} \max & L(x,y) - x^T s & & \\ \text{st} & \nabla_{(x)} L(x,y)^T & = & s, \\ & x, y, s \geq 0, & & \end{array}$$

$s \in \mathbf{R}^n$.
Alternatively:

$$\begin{array}{llll} \max & L(x,y) - \nabla_x L(x,y) x & & \\ \text{st} & \nabla_x L(x,y)^T & = & s, \\ & x, y, s \geq 0. & & \end{array}$$

mosek http://www.mosek.com

KKT optimality conditions

$$
\begin{array}{llllll}
(MCP) & \min & (x;y)^T(s;z) & & \\
 & \text{st} & a(x) & = & z, \\
 & \text{st} & \nabla_x L(x,y)^T & = & s, \\
 & & x,z,y,s \geq 0. & &
\end{array}
$$

where $z \in \mathbf{R}^m$ is slack variables.
Note:

- The KKT conditions is a monotone complementarity problem.
- Primal-dual interior-point type algorithm.
- Difficulties with detecting infeasibility!
- Feasibility constraints are potentially highly nonlinear!

The homogeneous model

The homogeneous model denoted $(HMCP)$:

$$
\begin{array}{lllll}
\min & (x;\tau)^T(s;\kappa) & & & \\
\text{st} & \tau a(x/\tau) & = & z, & \text{primal feasibility} \\
 & \tau\nabla_x L(x/\tau, y/\tau)^T & = & s, & \text{dual feasibility,} \\
 & -x^T\nabla_x L(x/\tau, y/\tau)^T - y^T a(x/\tau) & = & \kappa, & \text{zero gap,} \\
 & x, z, \tau, s, y, \kappa \geq 0. & & &
\end{array}
$$

- Has two additional variables τ, κ and one additional constraint.
- $(HMCP)$ is a homogeneous monotone complementarity problem.
- Always has solution (0).

A theorem

Theorem: (Andersen and Ye 95)
Let $(x^*, z^*, y^*, \tau^*, s^*, \kappa^*)$ be a maximal complementarity solution to $(HMCP)$. Then (MCP) has a solution if and only if $\tau^* > 0$. In this case, $(x^*, z^*, y^*, s^*)/\tau^*$ is a complementarity solution for (MCP).
Conclusion: Compute a maximal solution to $(HMCP)$.

The homogeneous algorithm

Algorithm: Apply a primal-dual interior-point method to $(HMCP)$ i.e. apply Newtons method to the perturbed optimality conditions:

$$
\begin{array}{rcl}
\tau \nabla_x L(x/\tau, y/\tau)^T & = & s, \\
\tau a(x/\tau) & = & z, \\
-x^T \nabla_x L(x/\tau, y/\tau)^T - y^T a(x/\tau) & = & \kappa, \\
Xs & = & \mu e, \\
Zy & = & \mu e, \\
\tau\kappa & = & \mu, \\
x, z, \tau, s, y, \kappa & > & 0.
\end{array}
$$

$\mu > 0$: Is a parameter.

- Primal feasibility cond. is nonlinear.
- Dual feasibility cond. is nonlinear.
- Perturbed complementarity cond. is nonlinear.

http://www.mosek.com

The algorithm

Step 0: Choose $(x; z; \tau; y; s; \kappa) > 0$.
Step 1: Choose parameters.
Step 2: Compute the residuals:

$$r \equiv \begin{bmatrix} r_D \\ r_G \\ r_P \end{bmatrix} \equiv - \begin{bmatrix} s - \tau g \\ \kappa + x^T g - y^T a(x/\tau) \\ z - \tau a(x/\tau) \end{bmatrix},$$

where

$$\begin{aligned} w &\equiv (x; z; \tau; y; s; \kappa) \\ J &\equiv \nabla a(x/\tau), \\ g &\equiv \nabla_x L(x/\tau, y/\tau)^T \\ H &\equiv \nabla_x^2 L(x/\tau, y/\tau). \end{aligned}$$

alg. cont

Step 3: Solve the Newton equation system:

$$\begin{bmatrix} H & v^2 & -J^T \\ (v^1)^T & H_G & -(v^3)^T \\ J & v^3 & 0 \end{bmatrix} \begin{bmatrix} d_x \\ d_\tau \\ d_y \end{bmatrix} - \begin{bmatrix} d_s \\ d_\kappa \\ d_z \end{bmatrix} = \eta r$$

and

$$\begin{array}{rcl} S d_x + X d_s & = & -Xs + \gamma\mu e, \\ Y d_z + Z d_y & = & -Zy + \gamma\mu e, \\ \kappa d_\tau + \tau d_\kappa & = & -\tau\kappa + \gamma\mu, \end{array}$$

where

$$\begin{array}{rcl} v^1 & \equiv & \nabla c(x/\tau)^T - Hx/\tau, \\ v^2 & \equiv & -\nabla c(x/\tau)^T - Hx/\tau, \\ v^3 & \equiv & a(x/\tau) - Jx/\tau, \\ H_g & \equiv & x/\tau^T H(x/\tau), \\ \mu & \equiv & (x; z; \tau)^T (s; y; \kappa)/(n + m + 1). \end{array}$$

mosek
http://www.mosek.com

alg. cont.

- $\eta = (\gamma - 1)$ where $\gamma \in [0, 1]$ is an algorithmic parameters.

Step 4: Update.

$$(x^+; \tau^+; y^+;) = (x; \tau; y) + \alpha(d_x; d_\tau; d_y)$$

Update of slacks:

$$(z^+; s^+; \kappa^+) = (z + \alpha d_z; s + \alpha d_s; \kappa + \alpha d_\kappa)$$

or

$$\begin{array}{rcl} z^+ & = & ((1 + \alpha\eta) r_P + \tau^+ a(x^+/\tau^+), \\ s^+ & = & ((1 + \alpha\eta) r_D + g^+), \\ \kappa^+ & = & ((1 + \alpha\eta) r_G + g_\kappa^+, \end{array}$$

where

$$\begin{array}{rcl} g^+ & \equiv & \tau^+ \nabla_x L(x^+/\tau^+, y^+/\tau^+)^T \\ g_\kappa^+ & \equiv & -(x^+)^T \nabla_x L(x^+/\tau^+, y^+/\tau^+)^T - (y^+)^T a(x^+/\tau^+) \end{array}$$

alg. cont.

Step 5: Repeat step 2 to 4.
Issues:

- Gap and residuals are reduced by the factor $(1+\alpha\eta)$ in every iteration (given the nonlinear update).
- Polynomial complexity holds under certain assumptions.
- **OBSERVE:** The nonlinear update

$$
\begin{array}{rcl}
z^{+} & = & ((1+\alpha\eta)r_P + \tau^{+}a(x^{+}/\tau^{+}), \\
s^{+} & = & ((1+\alpha\eta)r_D + g^{+}). \\
\kappa^{+} & = & ((1+\alpha\eta)r_G + g_{\kappa}^{+}).
\end{array}
$$

Does not always work well!!!

The nonlinear update

- Alternative. A linear update

$$
\begin{aligned}
z^+ &= x + \alpha d_z, \\
s^+ &= s + \alpha d_s \\
\kappa^+ &= \kappa + \alpha d_\kappa
\end{aligned}
$$

  but then a merit function is required e.g. two norm of the residuals.
- Works most of the time.
- Can lead to very slow convergence and robustness problems.
- A major weakness.
- Nonlinear primal and dual feasibility constraints are the cause of the problems.

A primal-dual algorithm for conic quadratic problems

The homogeneous model:

$$
\begin{array}{llcl}
(H) & Ax - b\tau & = & 0, \text{(primal feasibility)} \\
 & A^T y + s - c\tau & = & 0, \text{(dual feasibility)} \\
 & -c^T x + b^T y - \kappa & = & 0, \text{(zero gap)} \\
 & (x;\tau) \in \mathcal{K}, (s;\kappa) \in \mathcal{K} & &
\end{array}
$$

- In linear case suggested suggested Goldman and Tucker. Later rediscovered by Mizuno, Todd and Ye.
- All solutions satisfies: $x^T s + \tau\kappa = 0$.
- If and only if $\tau > 0$, then an optimal solution exists.
- Find a solution such that $\tau + \kappa > 0$.
- Linear primal and dual feasibility conditions!!!
- Algorithmic framework: The Nesterov-Todd direction.

One iteration

NT search direction:

$$
\begin{array}{rcl}
Ad_x - bd_\tau & = & \eta(Ax - b\tau), \\
A^T d_y + d_s - cd_\tau & = & \eta(A^T y + s - c\tau), \\
-c^T d_x + b^T d_y - d_\kappa & = & \eta(-c^T x + b^T y - \kappa), \\
\bar{X} T (\Theta W)^{-1} d_s + \bar{S} T \Theta W d_x & = & -\bar{X}\bar{S}e + \gamma\mu e, \\
\tau d_\kappa + \kappa d_\tau & = & -\tau\kappa + \gamma\mu.
\end{array}
$$

where $\eta := \gamma - 1$ and $\gamma \in [0, 1)$.

New iterate:

$$
\begin{bmatrix} x^+ \\ \tau^+ \\ y^+ \\ s^+ \\ \kappa^+ \end{bmatrix} = \begin{bmatrix} x \\ \tau \\ y \\ s \\ \kappa \end{bmatrix} + \alpha \begin{bmatrix} d_x \\ d_\tau \\ d_y \\ d_s \\ d_\kappa \end{bmatrix}.
$$

Properties of the search direction

Fact:

$$
\begin{aligned}
Ax^+ - b\tau^+ &= (1+\alpha\eta)(Ax - b\tau), \\
A^T y^+ + s^+ - c\tau^+ &= (1+\alpha\eta)(A^T y + s - c\tau), \\
-c^T x^+ + b^T y^+ - \kappa^+ &= (1+\alpha\eta)(-c^T x + b^T y - \kappa), \\
d_x^T d_s^T + d_\tau d_\kappa &= 0, \\
(x^+)^T s^+ + \tau^+\kappa^+ &= (1+\alpha\eta)((x)^T s + \tau\kappa).
\end{aligned}
$$

Observations:

- NT search direction can be seen as Newtons method.
- The method is symmetric!
- The complementarity gap is reduced by a factor of $(1+\alpha\eta) \in [0, 1)$.
- The infeasibility is reduced by the same factor.
- **No merit function is required.**

Summary about algorithms

- Both algorithms are called primal-dual interior-point.
- Only in the conic form are the feasibility constraints linear.
- The functional one is more like primal algorithm for the KKT system
- The conic algorithm is truly symmetric i.e. you can flip the primal and dual problem.
- Due to issues with the merit function then the functional form is not as robust as conic one.

http://www.mosek.com

An illustrative example

Conic form:

$$\begin{array}{ll} \min & y \\ \text{st} & x = y \\ & \|x\| \leq y, \\ & x \in \mathbf{R} \end{array}$$

Functional form:

$$\begin{array}{ll} \min & y \\ \text{st} & x = y \\ & \frac{x^2}{y} \leq y, \\ & y \geq 0, \\ & x, y \in \mathbf{R} \end{array}$$

MOSEK v6 on the conic formulation

ITE	PFEAS	DFEAS	KAP/TAU	POBJ	DOBJ	MU
0	1.0e+000	0.0e+000	1.0e+000	1.000e+000	0.0000e+000	1.0e+000
1	5.0e-002	0.0e+000	5.0e-002	5.000e-002	0.0000e+000	5.0e-002
2	2.5e-003	0.0e+000	2.5e-003	2.500e-003	0.0000e+000	2.5e-003
3	1.3e-004	0.0e+000	1.3e-004	1.250e-004	0.0000e+000	1.3e-004
4	6.3e-006	2.2e-016	6.3e-006	6.250e-006	0.0000e+000	6.3e-006
5	3.1e-007	0.0e+000	3.1e-007	3.128e-007	0.0000e+000	3.1e-007
6	1.6e-008	0.0e+000	1.6e-008	1.564e-008	0.0000e+000	1.6e-008
7	7.8e-010	1.1e-016	7.8e-010	7.821e-010	0.0000e+000	7.8e-010

MOSEK v6 on the functional formulation


```
ITE PFEAS    DFEAS    KAP/TAU  POBJ        DOBJ        MU
0   1.0e+000 1.4e+000 1.0e+000 1.000e+000  0.000e+000  1.0e+000
1   1.2e-001 6.4e-001 3.9e-003 1.648e-001  0.000e+000  1.1e-001
2   2.2e-003 6.8e-001 1.8e-003 1.972e-003  0.000e+000  2.0e-003
...
50  3.2e-009 3.5e-003 1.5e-009 2.040e-009  -4.135e-025 2.4e-009
51  2.9e-009 3.1e-003 1.3e-009 1.836e-009  0.000e+000  2.1e-009
52  2.6e-009 2.8e-003 1.2e-009 1.653e-009  0.000e+000  1.9e-009
...
122 1.6e-012 1.4e-006 7.4e-013 1.036e-012  0.0000e+000 1.2e-012
123 1.5e-012 1.2e-006 6.7e-013 9.327e-013  0.0000e+000 1.1e-012
```

- Performance on the functional formulation is bad.
- Merit function forces the alg. to take small steps.
- Behavior seen occasionally on large instances.
- Other interior-point optimizers may do better of course.

mosek http://www.mosek.com

Summary

The (restricted) conic formulation of nonlinear problems is advantageous because

- The problem data is simpler to deal in a software sense and no blackbox functions are required.
- It is impossible to formulate nonconvex problems.
- The available algorithm is much better.
- It easy to dualize the conic formulation.

Caveats:

- The conic formulation is usually bigger in terms of number of variables and constraints but also highly structured.
- Usually the storage and time spend per iteration will be same or less for the conic formulation.

Conclusion

- The conic form is better than the functional form!