LINMA2471: Optimization models and methods: course 11

BROGNIET Adrien, DEFOUR Loïc, NINANE Charles

December 1, 2015

1 Conic modeling

So far we have seen two types of cones:

- $\mathbb{R}_n^+$ first order cone to represent linear constraints.
- $\mathbb{L}^n$ second order cone to represent quadratics constraints like ratio, second and third power, ...

Theses cover a large class of problems but we will see a third type of cone.

Positive semi-definite cone

A positive semi-definite cone is defined as: $\{ \text{ M} \in \mathbb{S}^n | M \succeq 0 \}$ where $\mathbb{S}^n$ is the set of symmetric matrices of size n. The following definitions of a positive semi-definite matrix are equivalent:

- $\mathbb{M} \succeq 0$
- All eigenvalues of M are greater than 0.
- $\exists B \in \mathbb{R}^{n \times r} : M = BB^T$ (kind of Cholesky factorisation)
- $\forall x : x^T M x \geq 0$

Exemple conic optimization with semi-definite cone

Find a symmetric matrix M such that:

- $M = \begin{pmatrix} 1 & 2 & 5 \\ 2 & x & -1 \\ 5 & -1 & y \end{pmatrix}$
- λ_{max} is minimum.

Idea:
Let $M \in \mathbb{S}^n$ with eigenvalues $\lambda_1, \cdots, \lambda_n$. The eigenvalues are real because M is symmetric. It's easy to see that $M + \Delta I \in \mathbb{S}^n$ has eigenvalues $\lambda_1 + \Delta, \cdots, \lambda_n + \Delta$. By applying a shift on the diagonal of M we also shift the eigenvalues. By applying these observations to matrix M we can write the following:
All eigenvalues of M $\succeq 0 \Leftrightarrow M \geq 0 \Leftrightarrow \lambda_{min}(M) \geq 0$
All eigenvalues of M $\geq 1 \Leftrightarrow M - I \succeq 0 \Leftrightarrow \lambda_{min}(M) \geq 1$
All eigenvalues of M $\succeq c \Leftrightarrow M - cI \geq 0$ (linear even if c is a variable)$\Leftrightarrow \lambda_{min}(M) \geq c$
All eigenvalues of M $\leq c \Leftrightarrow M - cI \preceq 0 \Leftrightarrow \lambda_{max}(M) \leq c \Leftrightarrow -(M - cI) \succeq 0 = cI - M \succeq 0$ which is also a linear constraint.

We can now reformulate the optimization problem as:

$$\min_{x,y,t} t \text{ s.t } - \begin{pmatrix} 1 & 2 & 5 \\ 2 & x & -1 \\ 5 & -1 & y \end{pmatrix} + tI \succeq 0$$

2 Second order methods

2.1 Pros and cons (of the Newton's method)

Pros	Cons
- Faster convergence than first order methods	- More expensive computation
- Potential quadratic convergence under specific conditions	- Not always well defined
	- Not globally convergent

2.2 Computing the Newton's method

We consider the unconstrained optimization program :

$$\min_x f(x)$$

$$x \in \mathbb{R}^n$$

We use the second order Taylor expansion to write :

$$f(x) = f(x_k + h) \approx f(x_k) + \nabla f(x_k)^T h + \frac{1}{2} h^T \nabla^2 f(x_k) h$$

with x_k a fixed point and h a step such that $x_k, h \in \mathbb{R}^n$. We know that the stationary points of the function f are reached when $\frac{\partial f(x_k+h)}{\partial h} = 0$. Regarding the previous approximation, we have :

$$\frac{\partial f(x_k + h)}{\partial h} = \nabla f(x_k) + \nabla^2 f(x_k) h$$

Therefore we can choose the step h such that it satisfies the linear system :

$$\nabla^2 f(x_k) h = -\nabla f(x_k)$$

Assuming $\nabla^2 f(x_k)$ is invertible, we have :

$$h = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

The Newton's method is then :

Given x_0, f, ∇f, $\nabla^2 f$ invertible, $k = 0$

Repeat

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$$
$$k \leftarrow k + 1$$

Example 1 Using the Newton's method, we can find the global minimum of the function $f(x) = x^2$ in one iteration $\Longrightarrow x^+ = x - (2)^{-1}(2x) = 0$.

Example 2 Let us apply the Newton's method on $f(x) = -\cos(x)$.
We get that $x^+ = x - (\cos(x))^{-1}(\sin(x)) = x - \tan(x) = x - (x + \mathcal{O}(x^3))$ (near to 0, quadratic convergence)

Example 3 Let us apply the Newton's method on $f(x) = x^4$.
We get that $x^+ = x - (12x^2)^{-1}(4x^3) = x - \frac{x}{3} = \frac{2}{3}x$.
We have a linear convergence.
Why is that ? $f(x)$ has a "flat" minimum $\Longrightarrow$ in $x^* = 0, \nabla^2 f(x^*) = 12x^* = 0 \not\succeq \mu$.

2.3 Conditions for quadratic convergence

Theorem 1 (Quadratic convergence) *Let $f \in \mathcal{C}^2$. If $\nabla^2 f$ is M-Lipschitz and x^* is a minimum of f such that $\nabla^2 f(x^*) \succeq \mu I$ (with $\mu \in \mathbb{R}$ and I the identity matrix), then for any x such that $||x - x^*|| \leq \frac{\mu}{2M}$ we have :*

$$||x^+ - x^*|| \leq \frac{M}{\mu} ||x - x^*||^2$$

with $x^+ = x - \nabla^2 f(x)^{-1} \nabla f(x)$ well defined.

Example 4 $\frac{\mu}{2M} \longmapsto \frac{M}{\mu} \cdot \left(\frac{\mu}{2M}\right)^2 = \frac{\mu}{4M} \longmapsto \frac{M}{\mu} \cdot \left(\frac{\mu}{4M}\right)^2 = \frac{\mu}{16M}$ (quadratic convergence)
After k steps, we get $\frac{\mu}{M} \cdot \frac{1}{2^{2^{k-1}}}$.

2.4 Change of variables

We have seen that an iteration of the Newton's method is defined by $x \longmapsto x - \nabla^2 f(x)^{-1} \nabla f(x)$.
Let us define x such that $x \triangleq Ay$ and $g(y) = f(Ay)$.

We claim that $x^+ = Ay^+$ (hence : $f(x^+) = g(y^+)$).
Let us proof that claim. We compute that

$$\nabla g(y) = A^T \nabla f(Ay), \qquad \nabla^2 g(y) = A^T \nabla^2 f(Ay) A$$

So we have
$y^+ = y - \nabla^2 g(y)^{-1} \nabla g(y) \Longrightarrow y^+ = y - (A^T \nabla^2 f(Ay) A)^{-1} A^T \nabla f(Ay) \Longrightarrow y^+ = y - A^{-1} \nabla^2 f(Ay) A^{-T} A^T \nabla f(Ay)$

If we multiply the last expression by A on both sides, we get
$Ay^+ = Ay - AA^{-1} \nabla^2 f(Ay) \nabla f(Ay) \Longrightarrow Ay^+ = Ay - \nabla^2 f(Ay) \nabla f(Ay) \Longrightarrow x^+ = x - \nabla^2 f(x)^{-1} \nabla f(x)$

We have shown that Newton's method is affine invariant.

2.5 Self-concordent function

Definition 1 A function f is called self-concordent if

- $f \in \mathcal{C}^2(X)$ (open domain X)
- f convex
- $\nabla^3 f(x)[h, h, h] \leq 2(h^T \nabla^2 f(x)\ h)^{3/2} \quad \forall x \in X, \forall h$