Nonlinear Optimization: Lecture II

Ya-xiang Yuan

Inst of Computational Mathematics and Scientific/Engineering Computing
Chinese Academy of Sciences, P. O. Box 2719, Beijing 100080, China
Email: yyx@lsec.cc.ac.cn

Optimization and Applications, Beijing, Sept 18-Oct 15, 2006

Outline

Penalty Idea

Consider the nonlinear optimization problem:

$$\begin{array}{lll} \min & f(x) & \\ \text{subject to} & c_i(x) = 0, & i = 1, ..., m_e; \\ & c_i(x) \geq 0, & i = m_e + 1, ..., m. \end{array}$$

Feasible Set X

$$X = \{x| \quad c_i(x) = 0, i = 1, ..., m_e; c_i(x) \geq 0, i = m_e + 1, ..., m.\}$$

Objective: Find the best point in X.

Questions How about the points that are not belong to X?

Outline

Constraint Violation

Consider the simple case:
$m = m_e > 0$ (Equality Constrained Optimization)

A point is feasible if and only if $\|c(x)\|_2 = 0$

$\|c(x)\|_2$ is a *Constraint Violation*.

A **penalty term** is a non-negative term that is zero if and only if the variable is feasible.

A penalty function normally consists of the objective function and a penalty term. For example

$$P_\sigma(x) = f(x) + \sigma\|c(x)\|_2^2$$

The above function is called the **Courant Penalty Function**

Properties of the Courant Penalty Function

Assume that there exists a $\sigma_0 > 0$ such that $P_{\sigma_0}(x)$ is bounded below (we call that the constrained problem can be *well-penalized*.)

Let $x(\sigma)$ be the solution of

$$\min_{x \in \Re^n} P_\sigma(x) = f(x) + \sigma \|c(x)\|_2^2.$$

Theorem *Let $0 < \sigma_1 < \sigma_2$. Then*

$$f(x(\sigma_1)) \leq f(x(\sigma_2))$$

$$\|c(x(\sigma_1))\|_2 \geq \|c(x(\sigma_2))\|_2$$

Theorem *Let $\delta = \|c(x(\sigma))\|_2$. $x(\sigma)$ is also the solution of*

$$\min_{x \in \Re^n} f(x)$$

subject to $\|c(x)\|_2 \leq \delta$.

Theorem

$$\lim_{\sigma\to\infty} \|c(x(\sigma))\|_2 = \min_{x\in\Re^n} \|c(x)\|_2.$$

Proof If the theorem is not true, there exists $\hat{x}$ such that

$$\|c(x(\sigma))\|_2 \geq \gamma > \|c(\hat{x})\|_2 > \min_{x\in\Re^n} \|c(x)\|_2.$$

holds for all $\sigma > \sigma_0$.
Therefore, for all $\sigma > \sigma_0$, we can show that

$$\begin{aligned} f(\hat{x}) + \sigma\|c(\hat{x})\|_2^2 &\geq f(x(\sigma)) + \sigma\|c(x(\sigma))\|_2^2 \\ &= \left[f(x(\sigma)) + \sigma_0\|c(x(\sigma))\|_2^2\right] + (\sigma - \sigma_0)\|c(x(\sigma))\|_2^2 \\ &\geq \left[f(x(\sigma_0)) + \sigma_0\|c(x(\sigma_0))\|_2^2\right] + (\sigma - \sigma_0)\gamma^2 \end{aligned}$$

Dividing both sides by σ and letting $\sigma \to \infty$ we obtain a contradiction.

A Penalty Function Method Based on Courant PF

Algorithm (A penalty function method)

Step 1 Given $x_0 \in \Re^n$, $\sigma_1 > 0$, $\epsilon \geq 0$, $k := 1$

Step 2 Solve(starting from x_k):

$$\min_{x \in \Re^n} P_{\sigma_k}(x)$$

obtaining $x(\sigma_k)$.

Step 3 If $\|c(x(\sigma_k)\|_2 \leq \epsilon$ then stop.
Set $x_{k+1} := x(\sigma_k)$, $\sigma_{k+1} := 10\sigma_k$;
$k := k + 1$, go to Step 2.

Convergence of Penalty Function Method

Theorem *If* $\epsilon > \min \|c(x)\|_2$, *the penalty function method will terminate after finitely many iterations.*

Theorem *If* $\sigma_k \to \infty$, *any accumulation point* x^* *of* x_k *is a solution of*

$$\min_{x \in \Re^n} f(x)$$

subject to $\|c(x)\|_2 = \min_{y \in \Re^n} \|c(y)\|_2$.

Assume that $\|c(x^*)\| = 0$ (feasible), we have

- $\|c(x_{k+1})\| = O(1/\sigma_k)$.
- $\|x_{k+1} - x^*\| = O(1/\sigma_k)$.
- $\|\lambda_{k+1} - \lambda^*\| = O(1/\sigma_k)$.

Outline

Consider the inequality constrained problem:

$$\min_{x \in \Re^n} f(x)$$
$$s.t. \qquad c_i(x) \geq 0, \quad i = 1, ..., m.$$

A **barrier term** is a term that approaches to positive infinity if the variable approaches to the boundary of the feasible region.

A Barrier Function normally consists the objective function and a barrier term.

$$P(x) = f(x) + \frac{1}{\sigma} h(c(x))$$

Some Barrier Functions

- Inverse barrier function

$$P_\mu(x) = f(x) + \mu \sum_{i=1}^{m} \frac{1}{c_i(x)}$$

- Logarithmic barrier function

$$P_\mu(x) = f(x) + \mu \sum_{i=1}^{m} \log(\frac{1}{c_i(x)})$$

Homework

NLP-HW4: Could you give another barrier function, and compare it with the Log-barrier function?

Outline

How to Avoid Infinite Penalty Parameter?

Another look of the simple penalty function: $P_\sigma(x) = f(x) + \sigma\|c(x)\|_2^2$:

$$\nabla f(x(\sigma)) + \sigma \sum_{i=1}^{m} c_i(x(\sigma))\nabla c_i(x(\sigma)) = 0$$

when $x(\sigma) \to x^*$, $-\sigma c_i(x(\sigma)) \to \lambda_i^*$.

If $\lambda_i = 0$ for all i, we may not require $\sigma \to \infty$!

Consider the equivalent problem:

$$\min f(x) - \sum_{i=1}^{m} \lambda_i^* c_i(x)$$

$$s.t. \quad c(x) = 0.$$

The Lagrange multipliers for the above problem are all zero!

Augmented Lagrangian Function

The multiplier penalty function:

$$P(x,\lambda,\sigma) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x) + \frac{1}{2}\sum_{i=1}^{m} \sigma_i c_i(x)^2$$

Powell's modification to the simple penalty function: Compare

$$\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0 \qquad (KKT)$$

$$\nabla f(x(\sigma)) + \sigma \sum_{i=1}^{m} c_i(x(\sigma)) \nabla c_i(x(\sigma)) = 0 \qquad (Simple\ Penalty)$$

Gradient is correct, but the function value is not $\rightarrow$ shifting $c_i(x)$!

$$f(x) + \sigma \sum_{i=1}^{m} (c_i(x) - \theta_i)^2$$

which also leads to the Augmented Lagrangian Function.

Question: How to choose proper multipliers λ_i?

Given $\lambda^{(k)}$ and $\sigma^{(k)}$, obtain x_{k+1} by minimizing the above function.

$$\nabla f(x_{k+1}) - \sum_{i=1}^{m}(\lambda_i^{(k)} - \sigma_i^{(k)} c_i(x_{k+1}))\nabla c_i(x_{k+1}) = 0$$

Thus, it is natural to let

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} - \sigma_i^{(k)} c_i(x_{k+1}).$$

Fletcher's Differentiable Penalty Function

Idea: multipliers depends on x instead of as parameters.
what we wish:

$$\nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla c_i(x) = 0$$

Least Square Solution:

$$\lambda(x) = \text{argmin}_{\lambda \in \Re^m} \|\nabla f(x) - A(x)\lambda\|_2$$

$A(x) = \nabla c(x)^T$: Jacobian of $c(x)$

Fletcher's differentiable penalty function:

$$P(x, \sigma) = f(x) - \lambda(x)^T c(x) + \sigma \|c(x)\|_2^2$$

Exact Penalty Function!

Outline

Nonsmooth Exact Penalty Functions

Two commonly used exact penalty functions

- L_1 penalty function:

$$P_1(x) = f(x) + \sigma \|c(x)\|_1$$

- L_∞ penalty function:

$$P_\infty(x) = f(x) + \sigma \|c(x)\|_\infty$$

Advantages: 1) Simple; 2) Exact.

Generalized to Inequality constraints:
$c_i^{(-)} = c_i(x)(i = 1, ..., m); \quad c_i^{(-)} = \min\{0, c_i(x)\}(i = m_e + 1, ..., m)$

Outline

Lagrange-Newton Method

Consider the equality constrained problem:

$$\min_{x \in \Re^n} \quad f(x)$$
$$\text{s. t.} \quad c(x) = 0,$$

KKT conditions (Stationary point to the Lagrangian):

$$\begin{aligned} \nabla f(x) - \nabla c(x)^T \lambda &= 0, \\ -c(x) &= 0. \end{aligned}$$

Apply Newton's Method

$$\begin{pmatrix} W(x_k, \lambda_k) & -A(x_k) \\ -A(x_k)^T & 0 \end{pmatrix} \begin{pmatrix} (\delta x)_k \\ (\delta \lambda)_k \end{pmatrix} = - \begin{pmatrix} \nabla f(x_k) - A(x_k)\lambda_k \\ -c(x_k) \end{pmatrix},$$

where $A(x) = \nabla c(x)^T$, $W(x, \lambda) = \nabla^2 f(x) - \sum_{i=1}^m (\lambda_k)_i \nabla^2 c_i(x_k)$.

Lagrange-Newton Step as a QP step

$$\begin{pmatrix} W(x_k,\lambda_k) & -A(x_k) \\ -A(x_k)^T & 0 \end{pmatrix} \begin{pmatrix} d \\ \bar{\lambda} \end{pmatrix} = - \begin{pmatrix} \nabla f(x_k) \\ -c(x_k) \end{pmatrix},$$

Thus d is the solution of

$$\min \; d^T \nabla f(x_k) + \frac{1}{2} d^T W_k d$$

$$s.t. \qquad c_k + A_k^T d = 0$$

with $\bar{\lambda}$ being the corresponding multipliers.

Outline

SQP method

For general NLP, at each iteration, solve the following QP:

$$\begin{aligned} \min_{d \in \Re^n} \quad & g_k^T d + \frac{1}{2} d^T B_k d, \\ \text{s. t.} \quad & a_i(x_k)^T d + c_i(x_k) = 0, \quad i = 1, ..., m_e, \\ & a_i(x_k)^T d + c_i(x_k) \geq 0, \quad i = m_e + 1, ..., m. \end{aligned}$$

Let d_k be a solution of the above QP, then

$$\begin{aligned} g_k + B_k d_k - A(x_k)\lambda_k &= 0, & \\ (\lambda_k)_i &\geq 0, & i = m_e + 1, ..., m; \\ (\lambda_k)_i [c_i(x_k) + a_i(x_k)^T d_k] &= 0, & i = m_e + 1, ..., m. \end{aligned}$$

Note: d_k is a descent direction of many penalty functions!

Lemma Let d_k be a K-T point of the QP subproblem and λ_k be the corresponding Lagrange multiplier. Consider the L_1 penalty function

$$P(x,\sigma) = f(x) + \sigma \|c^{(-)}(x)\|_1,$$

we have that

$$P'_\alpha(x_k + \alpha d_k, \sigma)\big|_{\alpha=0} \le -d_k^T B_k d_k - \sigma \|c^{(-)}(x_k)\|_1 + \lambda_k^T c(x_k).$$

If $d_k^T B_k d_k > 0$ and $\sigma \ge \|\lambda_k\|_\infty$, then d_k is a descent direction of the penalty function $P(x,\sigma)$.

Wilson-Han-Powell Method

Step 1. Given $x_1 \in \Re^n$, $\sigma > 0$, $\delta > 0$, $B_1 \in \Re^{n \times n}$, $\epsilon \geq 0$, $k := 1$;

Step 2. Solve QP subproblem giving d_k;
if $\|d_k\| \leq \epsilon$ then stop;
find $\alpha_k \in [0, \delta]$ such that

$$P(x_k + \alpha_k d_k, \sigma) \leq \min_{0 \leq \alpha \leq \delta} P(x_k + \alpha d_k, \sigma) + \epsilon_k.$$

Step 3. $x_{k+1} = x_k + \alpha_k d_k$; Generate B_{k+1};
$k := k + 1$; go to Step 2.

update of σ

Question: How to ensure

$$\sigma > \|\lambda_k\|_\infty ?$$

Powell(1978) suggested

$$P(x, \sigma_k) = f(x) + \sum_{i=1}^{m} (\sigma_k)_i |c_i^{(-)}(x)|$$

with the update

$$\begin{aligned} (\sigma_1)_i &= (\lambda_1)_i, \\ (\sigma_k)_i &= \max\left\{ |[\lambda_k]_i|, \frac{1}{2}[(\sigma_{k-1})_i + |(\lambda_k)_i|] \right\}, \quad k > 1, \end{aligned}$$

$i = 1, \cdots, m$.

Update of B_k

Because B_k should approximate the Hessian of the Lagrangian, it is reasonable to require $B_{k+1}s_k = y_k$ with

$$\begin{aligned} s_k &= x_{k+1} - x_k\,, \\ y_k &= \nabla f(x_{k+1}) - \nabla f(x_k) - \sum_{i=1}^{m} (\lambda_k)_i [\nabla c_i(x_{k+1}) - \nabla c_i(x_k)]. \end{aligned}$$

$s_k^T y_k > 0$ may not be true. :(

Therefore, y_k is replaced by

$$\bar{y}_k = \begin{cases} y_k\,, & \text{if } s_k^T y_k \geq 0.2 s_k^T B_k s_k\,, \\ \theta_k y_k + (1-\theta_k) B_k s_k\,, & \text{otherwise.} \end{cases}$$

where $\theta_k = (0.8 s_k^T B_k s_k)/(s_k^T B_k s_k - s_k^T y_k)$.

Superlinearly Convergence of SQP method

Theorem d_k *is a superlinearly convergent step, namely*

$$\lim_{k\to\infty} \frac{\|x_k + d_k - x^*\|}{\|x_k - x^*\|} = 0$$

if and only if

$$\lim_{k\to\infty} \frac{\|P_k(B_k - W(x^*, \lambda^*))d_k\|}{\|d_k\|} = 0,$$

where P_k is a projection from $\Re^n$ onto the null space of $A(x_k)^T$:

$$P_k = (I - A(x_k)(A(x_k)^T A(x_k))^{-1} A(x_k)^T).$$

Outline

Maratos Effect

Example:

$$\min_{x=(u,v)\in\Re^2} \quad f(x) = 3v^2 - 2u,$$
$$\text{s. t.} \quad c(x) = u - v^2 = 0.$$

It is easy to see that $x^* = (0,0)^T$ is the unique minimizer. Initial point:

$$\bar{x}(\epsilon) = (u(\epsilon), v(\epsilon))^T = (\epsilon^2, \epsilon)^T$$

where $\epsilon > 0$ is a small parameter.
Let $B = W(x^*, \lambda^*)$, the quadratic programming subproblem is

$$\min_{d\in\Re^2} \quad d^T \begin{pmatrix} -2 \\ 6\epsilon \end{pmatrix} + \frac{1}{2} d^T \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} d,$$
$$\text{s. t.} \quad d^T \begin{pmatrix} 1 \\ -2\epsilon \end{pmatrix} = 0.$$

The solution of the above QP is $\bar{d}(\epsilon) = (-2\epsilon^2, \quad -\epsilon)$

Therefore, we have that

$$\|\bar{x}(\epsilon) + \bar{d}(\epsilon) - x^*\| = O(\|\bar{x}(\epsilon) - x^*\|^2).$$

Thus, $\bar{d}(\epsilon)$ is a superlinearly convergent step. Direct calculations indicates that

$$\begin{aligned} f(\bar{x}(\epsilon) + \bar{d}(\epsilon)) &= 2\epsilon^2, \\ c(\bar{x}(\epsilon) + \bar{d}(\epsilon)) &= -\epsilon^2. \end{aligned}$$

Because $f(\bar{x}(\epsilon)) = \epsilon^2$ and $c(\bar{x}(\epsilon)) = 0$, we have that

$$\begin{aligned} f(\bar{x}(\epsilon) + \bar{d}(\epsilon)) &> f(\bar{x}(\epsilon)), \\ |c(\bar{x}(\epsilon) + \bar{d}(\epsilon))| &> |c(\bar{x}(\epsilon))|. \end{aligned}$$

This example shows that a superlinearly convergent step can not ensure a reduction in the penalty function (Maratos Effect)

Techniques to overcome Maratos Effect

- Watch-dog
  Reducing the Lagrange function instead of the L_1 penalty.
- Second Order Correction Step
  Another step to the feasible set from the failed point.
- Smooth Exact Penalty Functions as merit

How to combine trust region to SQP?

SQP subproblem:

$$\min_{d\in\Re^n} g_k^T d + \frac{1}{2} d^T B_k d = \phi_k(d)$$

subject to

$$c_i(x_k) + d^T \nabla c_i(x_k) = 0 \qquad i = 1, 2, \ldots, m_e$$

$$c_i(x_k) + d^T \nabla c_i(x_k) \geq 0 \qquad i = m_e + 1, \ldots, m$$

How to combine the above QP with Trust Region $\|d\| \leq \Delta_k$?

- Null Space
- Exact penalty function
- Two ball subproblem

Null Space TR Algorithm

$$\min_{d \in \Re^n} g_k^T d + \frac{1}{2} d^T B_k d = \phi_k(d)$$

subject to

$$\theta c_i(x_k) + d^T \nabla c_i(x_k) = 0 \qquad i = 1, 2, \ldots, m_e$$

$$\theta c_i(x_k) + d^T \nabla c_i(x_k) \geq 0 \qquad i = m_e + 1, \ldots, m$$

where $\theta_k \in (0, 1]$

- range space step, Cauchy step
- null step, quasi-Newton step
- geometrically move feasible region towards to the current iterate

$\star$ Quadratic Program,
$\star$ A TRS problem in the Null Space.

Exact Penalty TR algorithm

$$\min_{d\in\Re^n} g_k^T d + \frac{1}{2} d^T B_k d + \sigma_k ||(c_k + A_k^T d)^-||_\infty = \Phi_k(d)$$

$$s.t. \quad ||d||_\infty \le \Delta_k.$$

Advantages:

- trial step closed related to the merit function
- subproblem always feasible
- automatically update penalty parameter
- no need for approximating Lagrange multipliers

Difficulties: – nonsmooth subproblem

Two Ball TR Subproblem

Celis, Dennis and Tapia(1985):

$$\min_{d \in \Re^n} g_k^T d + \frac{1}{2} d^T B_k d = \phi_k(d)$$

$$s.\,t. \quad ||(c_k + A_k^T d)^- ||_2 \leq \xi_k$$

$$||d||_2 \leq \Delta_k .$$

where $c_k = c(x_k) = (c_1(x), ..., c_m(x))^T$, $A_k = A(x_k) = \nabla c(x_k)^T$, $\xi_k \geq 0$ is a parameter and the superscript “-” means that $v_i^- = v_i (i = 1, ..., m_e)$, $v_i^- = \min[0, v_i](i = m_e + 1, ..., m)$.

liberation of Newton Movement $\longrightarrow$

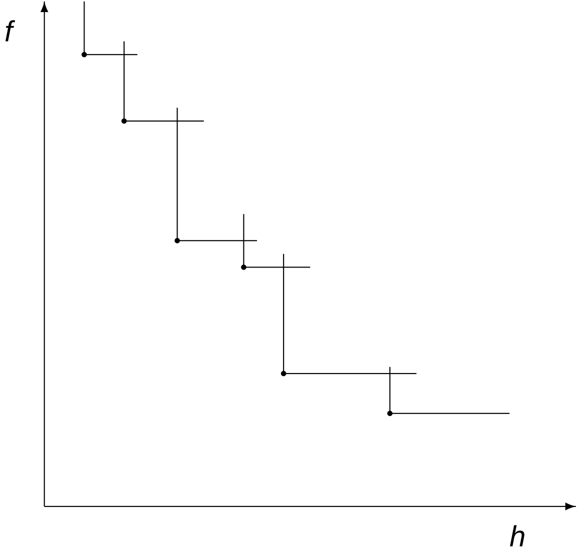
$\alpha_k = 1$ for line search, $\qquad x_{k+1} = x_k + s_k$ for trust region.

Definition of Filter

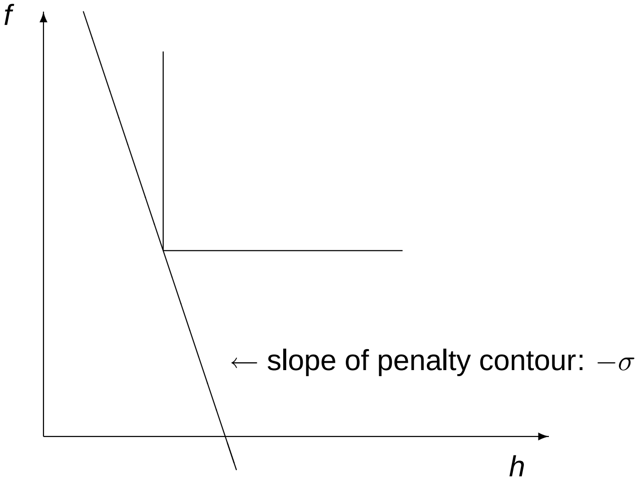
Definition I A pair (f_k, h_k) is said to dominate another pair (f_l, h_l) if and only if both $f_k \leq f_l$ and $h_k \leq h_l$.

Definition II A filter is a list of pairs (f_l, h_l) such that no pair dominates any other. A pair (f_k, h_k) is said to be acceptable for inclusion in the filter if it is not dominated by any pair in the filter.

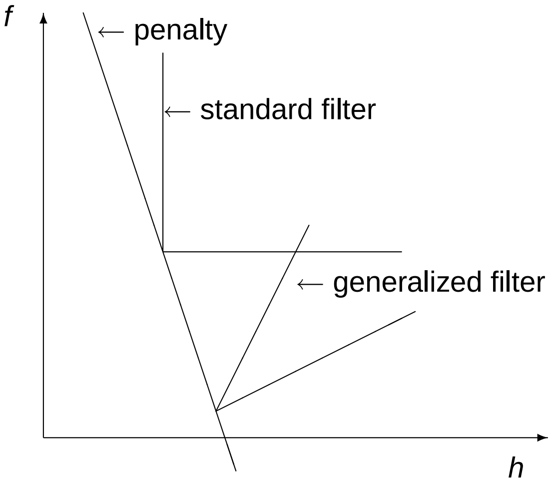
Geometric representation of a filter



Filter vs. Penalty (A filter view of Penalty)



Generalized Filter – Geometric



General Subspace Model for Unconstrained Opt.

At each iteration, a subspace $\mathcal{S}_k$ is available.
Try to construct a quadratic model $Q_k(d) \approx f(x_k + d)$ for $d \in \mathcal{S}_k$
Solve(obtaining d_k):

$$\min_{d \in \mathcal{S}_k} Q_k(d)$$

Do not carry line search – nor trust region
Either continue the process $x_{k+1} - x_k = s_k = d_k$
or modify the model (and the subspace):

$$Q_k(d_k) = f(x_k + d_k), \quad \mathcal{S}_k := \mathcal{S}_k \cup \{d_k\} \backslash \{\text{some old } v\}$$

Choices for Subspaces

Unconstrained Optimization:

- Subspace

$$
\begin{aligned}
Span\{-g_1, -g_2, ..., -g_k\} &= Span\{-g_k, y_{k-1}, ..., y_1\} \\
&= Span\{-g_k, s_{k-1}, ..., s_1\}
\end{aligned}
$$

- Subspace $Span\{-g_k, s_{k-1}, ..., s_{k-m}\}$
- Subspace $Span\{-g_k, y_{k-1}, ..., y_{k-m}\}$

Constrained Optimization:
A possible choice:

$$
S_k = Span\{-g_k, s_1, ..., s_k, -\nabla c_{k_i}\}
$$

Adding (a few) random directions to the subspace.

Discussions

Challenges and Opportunities

- Literature Driven / Problem Driven
- No significant advances /Difficult
- Popular / Recover
- Special Features / Special Applications

THANK YOU!