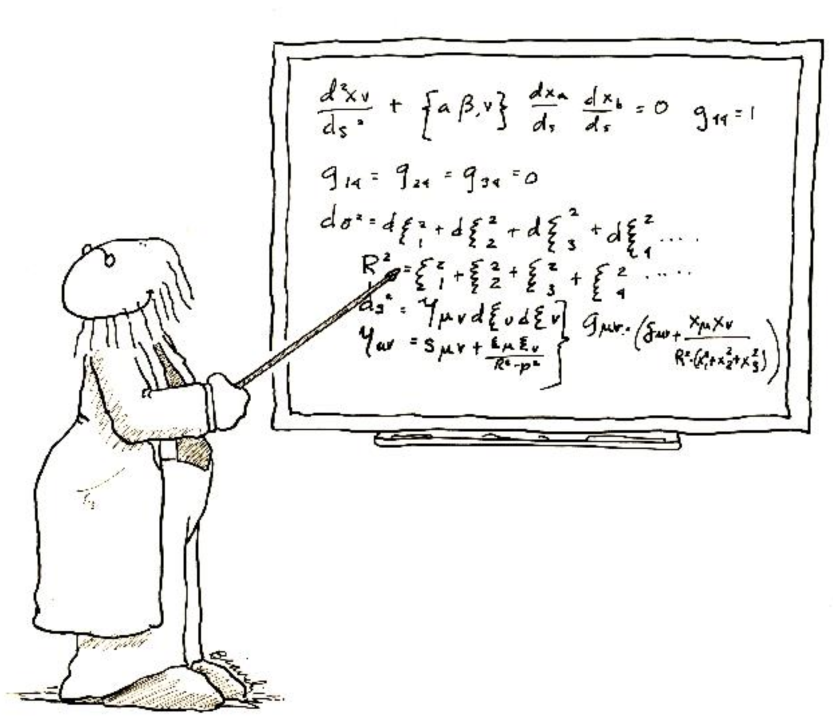
Optimization Theory and Methods



Prof. Xiaoe Ruan

Email: wruanxe@xjtu.edu.cn
Tel:13279321898

2019, Fall

Chapter7. Nonlinearly Constrained Optimizations

- Feature of Methods and Assessment
- Penalty Function Methods
- Lagrange Function Methods
- Feasible Point Method and Generalized Reduced Gradient Methods*
- Sequential Quadratic Programming Method*

Nonlinearly Constrained Optimization

$$\begin{cases} \min f(x), \quad \text{s. t.} \\ c_i(x) = 0, i \in E = \{1,2,\cdots,m_e\}, \\ c_i(x) \geq 0, i \in I = \{m_e+1,\cdots,m\}. \end{cases}$$

At least one of functions $c_i(x) \in R, i = 1,\cdots,m,$ **is nonlinear.**

$$D = \{x | c_i(x) = 0, i \in E, c_i(x) \geq 0, i \in I\}$$

--Constraint Set, Constraint Domain or Feasible Domain.

Optimization Strategy:

Construct iterative sequence $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$. **s.t.**

(1) $f\left(x^{(k)} + \alpha_k d^{(k)}\right) < f\left(x^{(k)}\right)$. **(2)** $x^{(k)} + \alpha_k d^{(k)} \in D$.

1、Features of Methods and Assessment

(1) Elimination Method

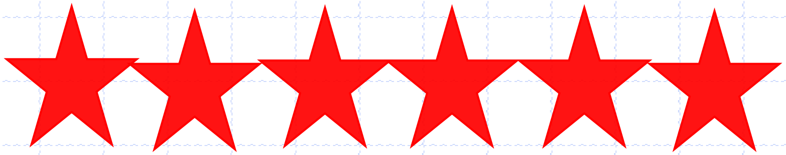
If $c_i(x) > 0 \Rightarrow y_i = \ln c_i(x)$ **Unconstrained ！**

$$\begin{cases} c_i(x_1, \cdots, x_n) = 0, \\ i = 1, 2, \cdots, m. \end{cases} \Rightarrow \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} \bar{c}_1(x_{m+1}, \cdots, x_n) \\ \vdots \\ \bar{c}_m(x_{m+1}, \cdots, x_n) \end{bmatrix}$$ **Then**

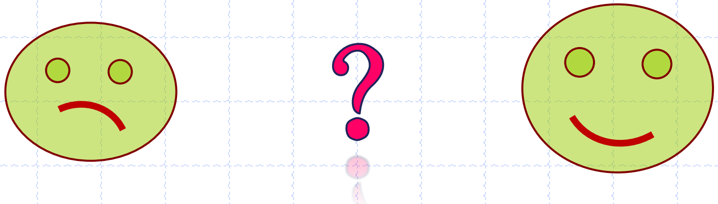
$$\min_{x \in D} f(x) \Rightarrow \min_{\hat{x} \in R^{n-m}} F(\hat{x})$$ **Unconstrained optimization**

Assessment:

Expectation index:



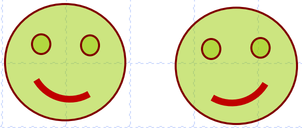
Feasibility:



(2) Grid Method, Random Test Method

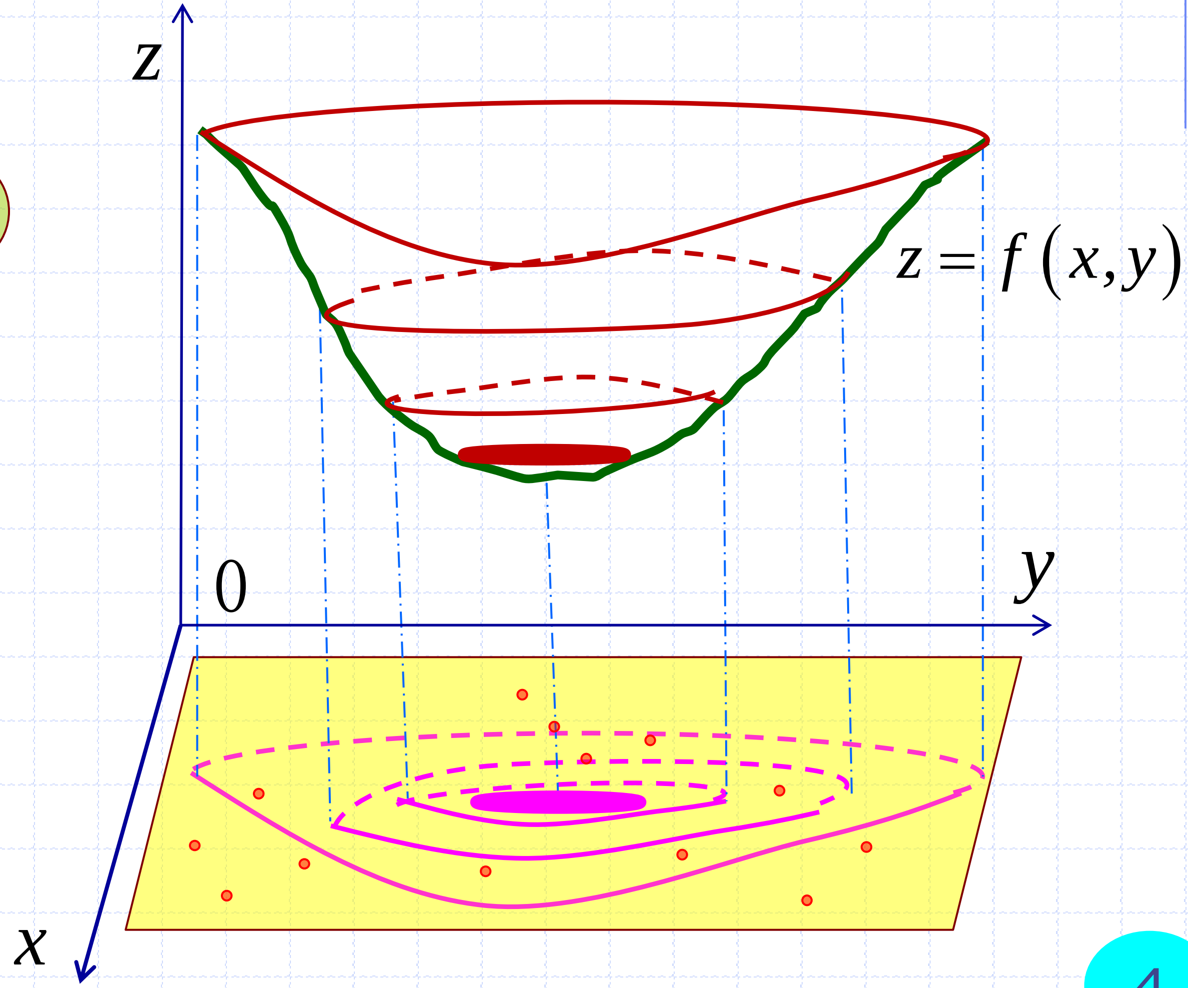
Generate test points and compare values of the objective function at the test points and testify constraint.

Assessment:

Feasibility: 

Accuracy: 

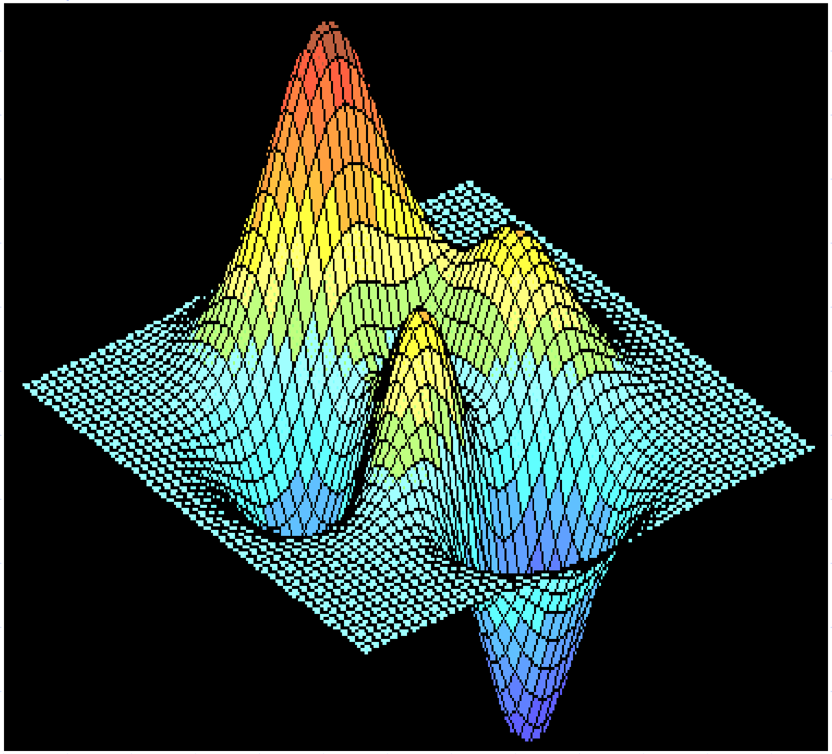
Convergence: ?

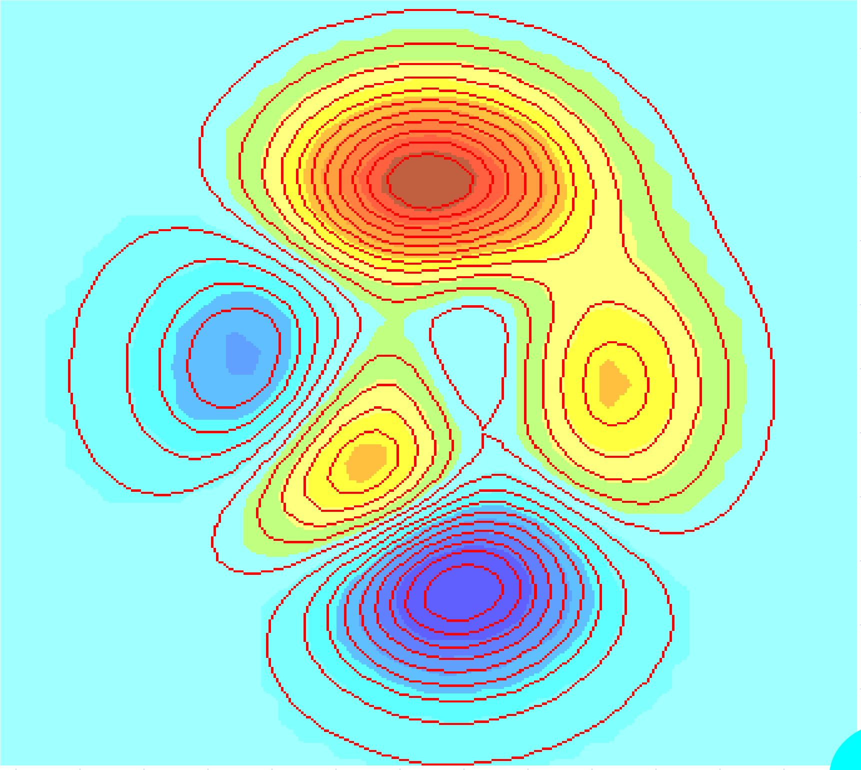


Motivational Problem:

"Maximize the following "peaks" function"

$$z = f(x,y)$$
$$= 3(1-x)^2 e^{-x^2-(y+1)^2} - 10\left(\frac{x}{5} - x^3 - y^5\right)e^{-x^2-y^2} - \frac{1}{3}e^{-(x+1)^2-y^2}$$

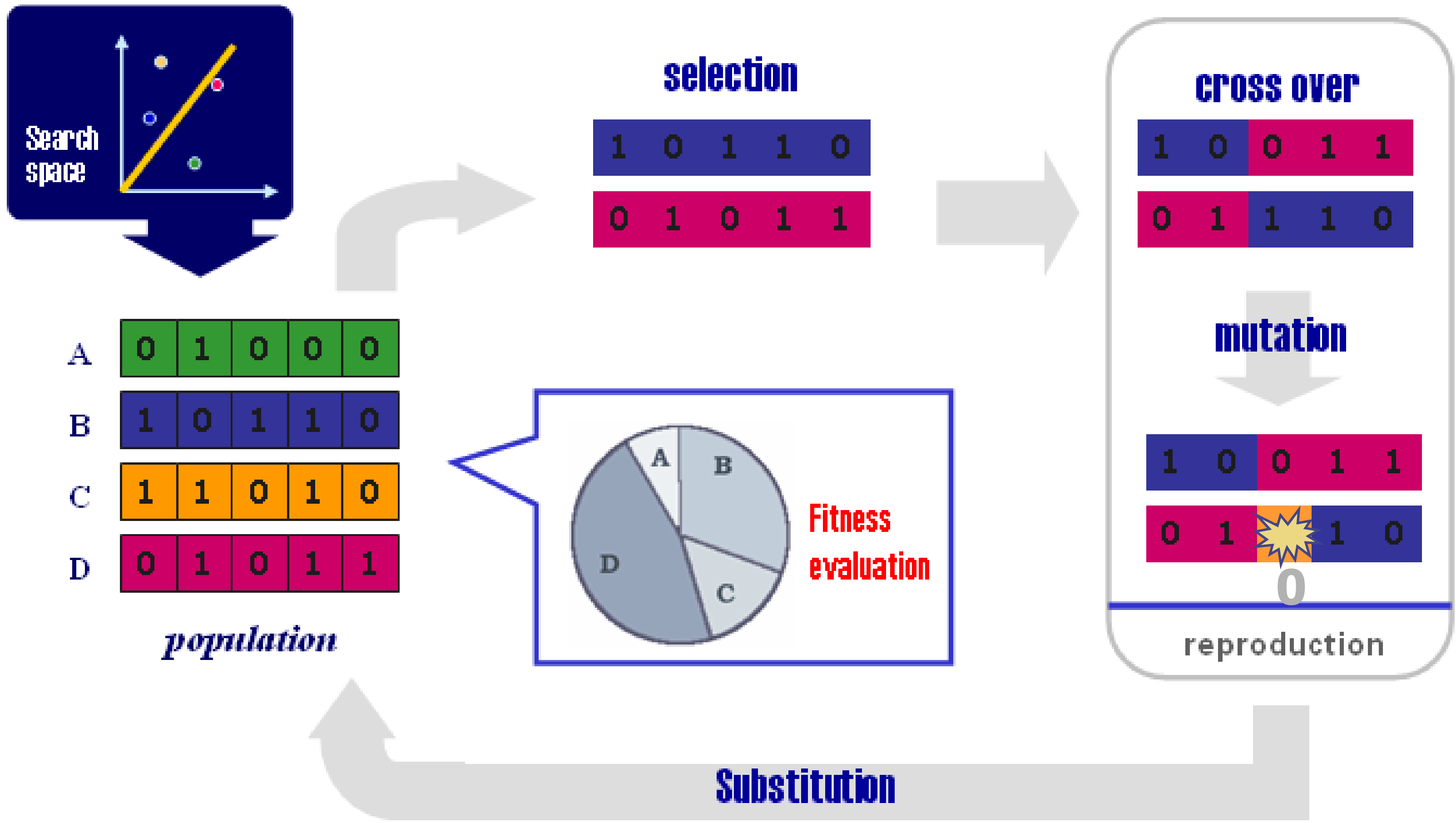




- **Derivatives of the "peaks" function**
  - dz/dx = -6*(1-x)*exp(-x^2-(y+1)^2) - 6*(1-x)^2*x*exp(-x^2-(y+1)^2) - 10*(1/5-3*x^2)*exp(-x^2-y^2) + 20*(1/5*x-x^3-y^5)*x*exp(-x^2-y^2) - 1/3*(-2*x-2)*exp(-(x+1)^2-y^2)
  - dz/dy = 3*(1-x)^2*(-2*y-2)*exp(-x^2-(y+1)^2) + 50*y^4*exp(-x^2-y^2) + 20*(1/5*x-x^3-y^5)*y*exp(-x^2-y^2) + 2/3*y*exp(-(x+1)^2-y^2)
  - d(dz/dx)/dx = 36*x*exp(-x^2-(y+1)^2) - 18*x^2*exp(-x^2-(y+1)^2) - 24*x^3*exp(-x^2-(y+1)^2) + 12*x^4*exp(-x^2-(y+1)^2) + 72*x*exp(-x^2-y^2) - 148*x^3*exp(-x^2-y^2) - 20*y^5*exp(-x^2-y^2) + 40*x^5*exp(-x^2-y^2) + 40*x^2*exp(-x^2-y^2)*y^5 -2/3*exp(-(x+1)^2-y^2) - 4/3*exp(-(x+1)^2-y^2)*x^2 -8/3*exp(-(x+1)^2-y^2)*x
  - d(dz/dy)/dy = -6*(1-x)^2*exp(-x^2-(y+1)^2) + 3*(1-x)^2*(-2*y-2)^2*exp(-x^2-(y+1)^2) + 200*y^3*exp(-x^2-y^2)-200*y^5*exp(-x^2-y^2) + 20*(1/5*x-x^3-y^5)*exp(-x^2-y^2) - 40*(1/5*x-x^3-y^5)*y^2*exp(-x^2-y^2) + 2/3*exp(-(x+1)^2-y^2)-4/3*y^2*exp(-(x+1)^2-y^2)

→ **An analytic solution is not easily found in a reasonable time span.**

Genetic Algorithms



GA: Operators Example

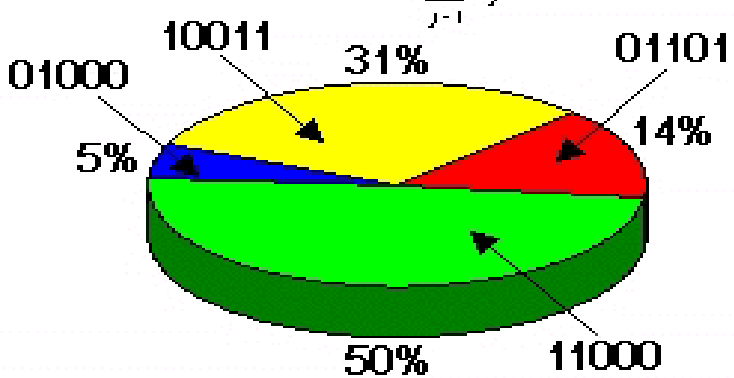
Operators Example

The Problem is to Maximize $f(x) = x^2$

Number	String	Fitness	% of the Total
1	01101	169	14.4
2	11000	576	49.2
3	01000	64	5.5
4	10011	361	30.9
Total		1170	100.0

1.- *Roulette Wheel Selection*

$$P_s = \frac{f_i}{\sum_{j=1}^{4} f_j}$$



2.- *One-Point Crossover (SPX)*

$P_c \in [0.6 .. 1.0]$

parents	offspring
01|101 (169)	01000 (64)
11|000 (576)	11101 (841)

3.- *Mutation*

$P_m \in [0.001 .. 0.1]$

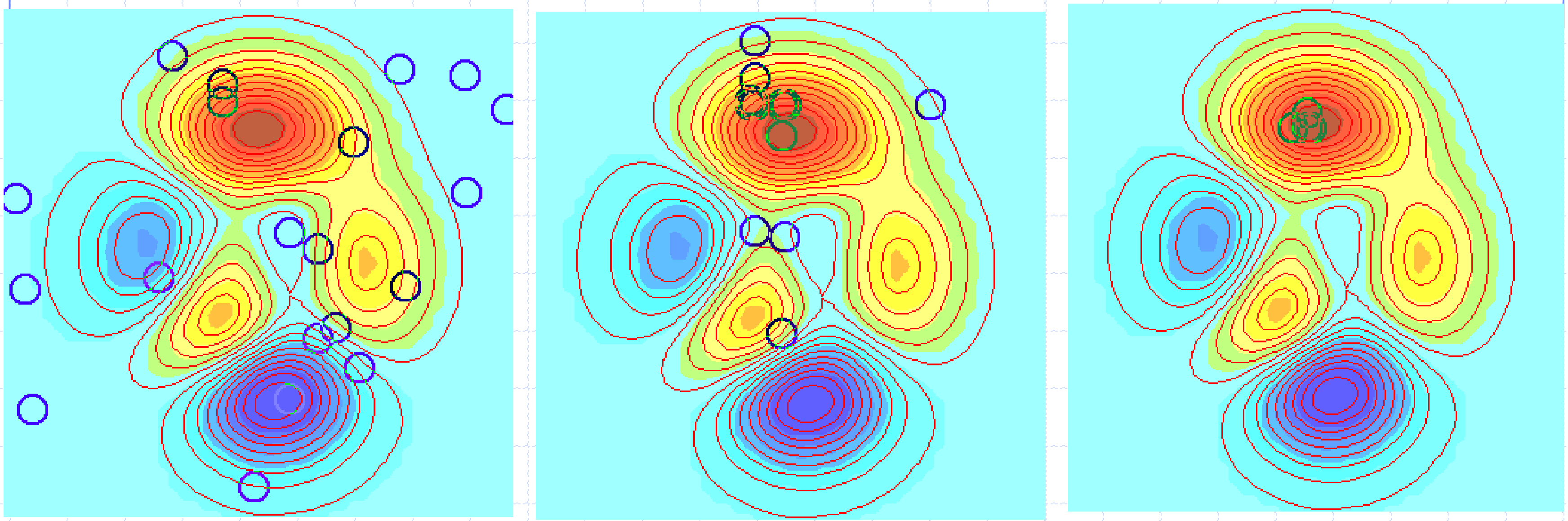
Mutate the first allele

0	1	0	0	0		1	1	0	0	0
4	3	2	1	0		4	3	2	1	0

fitness=64 ⇨ fitness=576

Apply a Genetic Algorithm

GA process: Start with multiple feasible solutions and apply GA repeatedly to obtain a solution.



Initial population **5th generation** **10th generation**

(3) Generalized Reduced Gradient Method
Rosen Gradient Method

$$\begin{cases} \min\ f(x), \quad \text{s.t.} \\ c_i(x) = 0, i \in E = \{1, 2, \cdots, m_e\}, \\ c_i(x) \geq 0, i \in I = \{m_e + 1, \cdots, m\}. \end{cases}$$

$$\begin{cases} \min\ \sigma, \quad \text{s.t.} \\ \nabla f\left(x^{(k)}\right)^{\mathrm{T}} d < \sigma, \\ \nabla c_i\left(x^{(k)}\right) d \geq -\sigma, 0, i \in I\left(x^{(k)}\right), \\ d^{\mathrm{T}} d \leq 1. \end{cases}$$

Assessment:

Feasibility:

Accuracy:

Convergence:

- **Elimination Method**
- **Reduced Gradient Method**
- **Projected Gradient Method**
- **Active Set Method**

(4) Penalty Function Method

$$\begin{cases} \min f(x), \quad \text{s. t.} \\ c_i(x) = 0, i \in E = \{1, 2, \cdots, m_e\}, \\ c_i(x) \geq 0, i \in I = \{m_e + 1, \cdots, m\}. \end{cases}$$

Assessment:

Feasibility:

Accuracy:

Convergence:

$$\min_{x \in R^n} p(x, \sigma) = f(x) + \sigma \bar{p}(x),$$

where

$$\bar{p}(x) = \sum_{i=1}^{m_e} |c_i(x)|^2 + \sum_{i=m_e+1}^{m} |\min(0, c_i(x))|^2$$

--Penalty Function

(5) Sequential Linear Programming Method, Sequential Quadratic Programming Method

$$
\begin{cases}
\min\ f(x), \quad \text{s.t.} \\
c_i(x) = 0, i \in E = \{1, 2, \cdots, m_e\}, \\
c_i(x) \geq 0, i \in I = \{m_e + 1, \cdots, m\}.
\end{cases}
$$

$$
\begin{cases}
\min\ d^{\mathrm{T}} \nabla f\left(x^{(k)}\right) + \frac{1}{2} d^{\mathrm{T}} W\left(x^{(k)}, \lambda^{(k)}\right) d, \quad \text{s.t.} \\
c_i\left(x^{(k)}\right) + d^{\mathrm{T}} \nabla c_i\left(x^{(k)}\right) = 0, i \in E = \{1, 2, \cdots, m_e\}, \\
c_i\left(x^{(k)}\right) + d^{\mathrm{T}} \nabla c_i\left(x^{(k)}\right) \geq 0, i \in I = \{m_e + 1, \cdots, m\}.
\end{cases}
$$

Assessment:

Feasibility:

Accuracy:

Convergence:

2. Penalty Function Method

NP: $$\begin{cases} \min\ f(x), \quad \text{s.t.} \\ c_i(x)=0, i\in E=\{1,2,\cdots,m_e\}, \\ c_i(x)\ge 0, i\in I=\{m_e+1,\cdots,m\}. \end{cases}$$

If x **is feasible point, then** $c_i(x)=0, i\in E,$ $c_i(x)\ge 0, i\in I.$

Otherwise, there exists at least an index i **or** j,

s.t. $c_i(x)\ne 0, i\in E,$ **or** $c_j(x)<0, j\in I.$

Denote $$\bar{p}(x)=\sum_{i=1}^{m_e}\left|c_i(x)\right|^2+\sum_{i=m_e+1}^{m}\left|\min\left(0,c_i(x)\right)\right|^2.$$

--Degree of constraint violation (Constraint Violation Function)

Target: **With the processing of iteration, it forces the iterative point approaching to feasible domain and eventually satisfies constraints and thus achieves the minimizer.**

Let $p(x,\sigma)=f(x)+\sigma\bar{p}(x)$ **--Penalty Fcn,** σ **--Penalty factor**

For $\sigma >> 1$, **if** $p(x^*(\sigma))=\min\limits_{x\in R^n} p(x,\sigma)$ **and** $x^*(\sigma)\in D$, **then** $\bar{p}(x^*(\sigma))=0$. **Thus** $x^*(\sigma)$ **is the minimizer of the original pbm.**

i.e. $f(x^*(\sigma))=p(x^*(\sigma))=\min\limits_{x\in R^n} p(x,\sigma)\le f(x), \quad \forall x\in D.$

In usual, choose $\sigma_1<\sigma_2<\cdots<\sigma_k\to\infty.$

Solve $\min\limits_{x\in R^n} p(x,\sigma_k)=f(x)+\sigma_k\bar{p}(x).$

---Sequential Unconstrained Minimization Technique（SUMT）

As $\sigma_k\bar{p}(x)\to\infty,(\bar{p}(x)\neq 0,k\to\infty)$, **thus**

if $p(x^*(\sigma_k))=\min\limits_{x\in R^n} p(x,\sigma_k)$, **then** $\bar{p}(x^*(\sigma_k))\to 0,(k\to\infty).$

This implies $\lim\limits_{k_i\to\infty} x^*(\sigma_{k_i})=x^*$ **is the minimizer of the original pbm.**

Th.1. **Given** $p(x,\sigma)=f(x)+\sigma\bar{p}(x)$ **and** $\sigma_{k+1}>\sigma_k>0.$

If $p\left(x^*(\sigma_k)\right)=\min\limits_{x\in R^n} p(x,\sigma_k),\ p\left(x^*(\sigma_{k+1})\right)=\min\limits_{x\in R^n} p(x,\sigma_{k+1}),$

then $p\left(x^*(\sigma_{k+1}),\sigma_{k+1}\right)\geq p\left(x^*(\sigma_k),\sigma_k\right)$ **--Penalty fcn increasing**

$\bar{p}\left(x^*(\sigma_{k+1})\right)\leq \bar{p}\left(x^*(\sigma_k)\right)$ **--Constraint violation descending**

$f\left(x^*(\sigma_{k+1})\right)\geq f\left(x^*(\sigma_k)\right)$ **--Objective fcn increasing**

From discussion above, if $\bar{p}\left(x^{(0)}\right)>0$ **then** $x^*(\sigma_k)\notin D,$

but $\left\|x^*(\sigma_k)-D\right\|\to 0,$ **and** $\lim\limits_{k_i\to\infty} x^*\left(\sigma_{k_i}\right)=x^*.$ **Therefore**

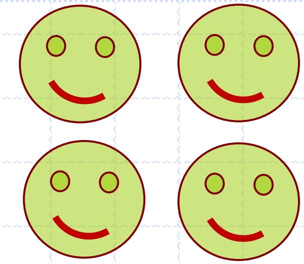
$p(x,\sigma)=f(x)+\sigma\bar{p}(x)$ **---Exterior Point Penalty Fcn**

SUMT is Exterior Point Penalty SUMT **or**

Exterior Point Penalty Function Method

Comment on Exterior Point SUMT

Feasibility:



Accuracy:

$\lim\limits_{k_i \to \infty} x^*\left(\sigma_{k_i}\right) = x^*$ **is theoretically the minimizer of the original pbm.**

But for finite iteration index,

$x^*\left(\sigma_k\right)$ **dose not rigorously guarantee the constraints.**

Convergence: **Linearly convergent.**

Drawbacks: **1. Optimization of unconstrained penalty fcn method possibly has no bounded minimizer.**

2. While σ_k is increasing, the unconstrained optimization may become ill.

Generalization: $\bar{p}(x) = \sum_{i=1}^{m_e} \left|c_i(x)\right|^{\alpha} + \sum_{i=m_e+1}^{m} \left|\min\left(0, c_i(x)\right)\right|^{\beta}, \quad \alpha \geq 1, \beta \geq 1.$

Forms of violation fcns and SUMT are diverse.

Interior Point Penalty Function Method

$$\min_{x\in D} f(x), \quad D=\{x \mid c_i(x)\ge 0, i=1,\cdots,m\}.$$ **Denote**

$$D^0=\{x \mid c_i(x)>0, i=1,\cdots,m\}$$ **--Interior of Constraint Domain**

Interior Point Penalty fcn: $p_I(x,\sigma)=f(x)+\sigma(t)\,\overline{p}_I(x),$

satisfying **(1)** $\overline{p}_I(x), x\in D^0$ **is continuous.**

(2) If $\{x^{(k)}\}\in D^0$, $\lim\limits_{k\to\infty} x^{(k)}=x_B$ **and at least there exists** i,

s.t. $c_i(x_B)=0$ **makes** $\overline{p}_I(x^{(k)})\to+\infty\,(k\to\infty).$

(3) $\sigma(t)$ **is strictly monotonously increasing and**

$\lim\limits_{k\to\infty} t_k=0$ **induces** $\lim\limits_{k\to\infty}\sigma(t_k)=0.$

e.g. $p_I(x,\sigma)=f(x)+\sigma\sum_{i=1}^{m}\ln c_i(x).$ **or** $p_I(x,\sigma)=f(x)+\sigma\sum_{i=1}^{m}\frac{1}{c_i(x)}$

IPP-SUMT:

$$D=\left\{x\middle|c_i\left(x\right)\ge 0, i=1,\cdots,m\right\}$$

$$D^0=\left\{x\middle|c_i\left(x\right)> 0, i=1,\cdots,m\right\}$$

$$\forall x^{(0)}\in D^0, t_1>0,\quad x^*\left(t_1\right)=arg\min_{x\in D} p_I\left(x,\sigma\left(t_1\right)\right)$$

$$0<t_2<t_1, x^{(0)}=x^*\left(t_1\right),\quad x^*\left(t_2\right)=arg\min_{x\in D} p_I\left(x,\sigma\left(t_2\right)\right)$$

$$0<t_{k+1}<t_k, x^{(0)}=x^*\left(t_k\right),\quad x^*\left(t_{k+1}\right)=arg\min_{x\in D} p_I\left(x,\sigma\left(t_{k+1}\right)\right)$$

$$\lim_{k_i\to\infty} x^*\left(t_{k_i}\right)=x^*, s.t. f\left(x^*\right)=\min_{x\in D} f\left(x\right)$$

Th.2 **Assume** (a) $f(x),\ c_i(x), i=1,\cdots,m.$ **continuous;**

(b) **IPP fcn:** $p_I(x,\sigma(t))=f(x)+\sigma(t)\bar{p}_I(x),$

(c) $A^*=\left\{x^* \,\middle|\, f(x^*)=\min\limits_{x\in D} f(x)\right\}\neq\varnothing$ **is isolately compact;**

(d) $A^*\bigcap D\neq\varnothing.$ (e) $0\leftarrow t_{k+1}<t_k<\cdots<t_1.$ **Then**

(1) **Exists a compact set** S **s.t.** $A^*\subset S^0,\ \forall x\in D\bigcap S\setminus A^*,\ f(x)>f(x^*).$

For properly small t_k, $x^*(t_k)=\arg\min\limits_{x\in S^0\bigcap D^0} p_I(x,\sigma(t_k)),$

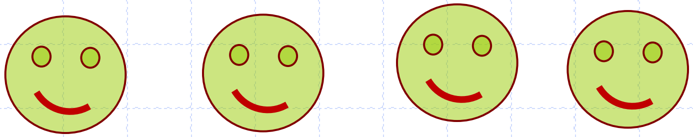
$\lim\limits_{k_i\to\infty} x^*(t_{k_i})=x^*\in A^*.$ **If** A^* **bounded, then** $\lim\limits_{k\to\infty} x^*(t_k)=x^*\in A^*.$

(2) $\lim\limits_{k\to\infty}\sigma(t_k)\bar{p}_I(x^*(t_k))=0.$ (3) $\lim\limits_{k\to\infty} f(x^*(t_k))=f(x^*).$

(4) $\lim\limits_{k\to\infty} p_I(x^*(t_k),\sigma(t_k))=f(x^*).$

(5) $f(x^*(t_{k+1}))\leq f(x^*(t_k)).$ (6) $\bar{p}_I(x^*(t_{k+1}))\geq\bar{p}_I(x^*(t_k)).$

Comments on IPP SUMT:

Feasibility: 

Accuracy: Sequence by IPP SUMT belongs to feasible domain.

Thus $\lim\limits_{k_i \to \infty} x^*\left(\sigma_{k_i}\right) = x^*$ is the minimizer.

Convergence: Linearly convergent.

Drawbacks: 1.While $\sigma\left(t_k\right) \to 0$ optimization possibly becomes ill.

2. Unfeasible for equality constraint optimization.

Solvable strategies:

Hybrid Algorithm of combining IPP-SUMT with EPP-SUMT or Lagrange Fcn Method, etc.

3.Lagrange Function Method **Recall**

NOEC:
$$\begin{cases} \min f(x), \text{ s. t.} \\ c_i(x) = 0, i = 1, \cdots, m. \end{cases}$$

Let
$$L(x,\lambda) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x)$$

1-st-order Kuhn-Tucker Necessity Conditions:

Given x^* **is the local minimizer of NOEC.** $c_i(x)(i = 1, \cdots, m)$ **and** $f(x)$ **are 1st-order continuously differentiable at** x^*

If $SFD(x^*, d) = LFD(x^*, d)$, **Then there exists**

$\lambda^* = (\lambda_1^*, \lambda_2^*, \cdots, \lambda_m^*)^T$, **s.t.**

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0, i = 1, \cdots, m.$$

Question: $L(x^*, \lambda^*) \overset{?}{=} \min_{x \in R^n} L(x, \lambda^*)$

If $\nabla_{xx}^2 L(x^*,\lambda^*) > 0$, **then** $L(x^*,\lambda^*) = \min_{x\in R^n} L(x,\lambda^*)$.

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

Penalty Lagrange Fcn or Augmented Lagrange Fcn.

If x^* **is the local minimizer of NOEC,** **then**

$$\nabla_x L(x^*,\lambda^*) = \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0, \quad c_i(x^*) = 0, \; i = 1,\cdots,m.$$

Thus $$\nabla_x p(x^*,\lambda^*,\sigma) = \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) + \sigma \sum_{i=1}^{m} c_i(x^*) \nabla c_i(x^*) = 0.$$

$$\begin{aligned}\nabla_{xx}^2 p(x^*,\lambda^*,\sigma) &= \nabla^2 f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla^2 c_i(x^*) + \sigma A(x^*) A(x^*)^{\mathrm{T}} \\ &= \nabla_{xx}^2 L(x^*,\lambda^*) + \sigma A(x^*) A(x^*)^{\mathrm{T}}\end{aligned}$$

where $A(x^*) = \left[\nabla c_1(x^*) \,|\, \nabla c_2(x^*) \,|\cdots|\, \nabla c_m(x^*)\right]$.

$$\nabla_x p(x^*,\lambda^*,\sigma)=0, \quad \nabla^2_{xx} p(x^*,\lambda^*,\sigma)=\nabla^2_{xx} L(x^*,\lambda^*)+\sigma A(x^*)A(x^*)^{\mathrm{T}}.$$

(1) If $\forall 0\neq d\in R^n,\ A(x^*)^{\mathrm{T}} d\neq 0.$ **then** $A(x^*)A(x^*)^{\mathrm{T}}$ **PD.**

Thus $\exists \sigma^*>0,$ **s.t. for** $\forall \sigma\geq\sigma^*,\ \nabla^2_{xx} p(x^*,\lambda^*,\sigma)$ **PD.**

(2) If $0\neq d\in R^n,\ A(x^*)^{\mathrm{T}} d=0$ **and** $d^{\mathrm{T}}\nabla^2_{xx} L(x^*,\lambda^*)d>0,$ **then**

$\nabla^2_{xx} p(x^*,\lambda^*,\sigma)$ **PD.**

From 2nd-order sufficiency , (1) and (2) guarantees

x^* **is the minimizer of multiplier penalty fcn** $p(x,\lambda,\sigma)$

and
$$\begin{cases} p(x^*,\lambda^*,\sigma)=\min\limits_{x\in R^n} p(x,\lambda,\sigma) \\ c_i(x^*)=0, i=1,\cdots,m. \end{cases} \Rightarrow f(x^*)=\min_{x\in D} f(x)$$

$$\nabla L(x^*,\lambda^*)=0, \nabla^2_{xx} L(x^*,\lambda^*)>0; \Rightarrow \begin{cases} \nabla_x p(x^*,\lambda^*,\sigma)=0, \\ \nabla^2_{xx} p(x^*,\lambda^*,\sigma)>0, \sigma\geq\sigma^*. \end{cases}$$

For NOIC: $\min f(x), \text{s. t. } c_i(x) \geq 0, i = 1, \cdots, m.$

$\Longrightarrow$ $\min f(x), \text{s. t. } c_i(x) - z_i^2 = 0, i = 1, \cdots, m.$

Let $\tilde{p}(x,\lambda,z,\sigma) = f(x) - \sum_{i=1}^{m} \lambda_i \left(c_i(x) - z_i^2\right) + \frac{1}{2}\sigma \sum_{i=1}^{m} \left(c_i(x) - z_i^2\right)^2$

$$\frac{\partial \tilde{p}(x,\lambda,z,\sigma)}{\partial z_i} = 0 \Longrightarrow z_i^2 = \frac{1}{\sigma} max\left\{0, \sigma c_i(x) - \lambda_i\right\}.$$

Augmented Lagrange Function

$$p_I(x,\lambda,\sigma) = f(x) + \frac{1}{2\sigma} \sum_{i=1}^{m} \left[\left(max\left\{0, \sigma c_i(x) - \lambda_i\right\}\right)^2 - \lambda_i^2\right]$$

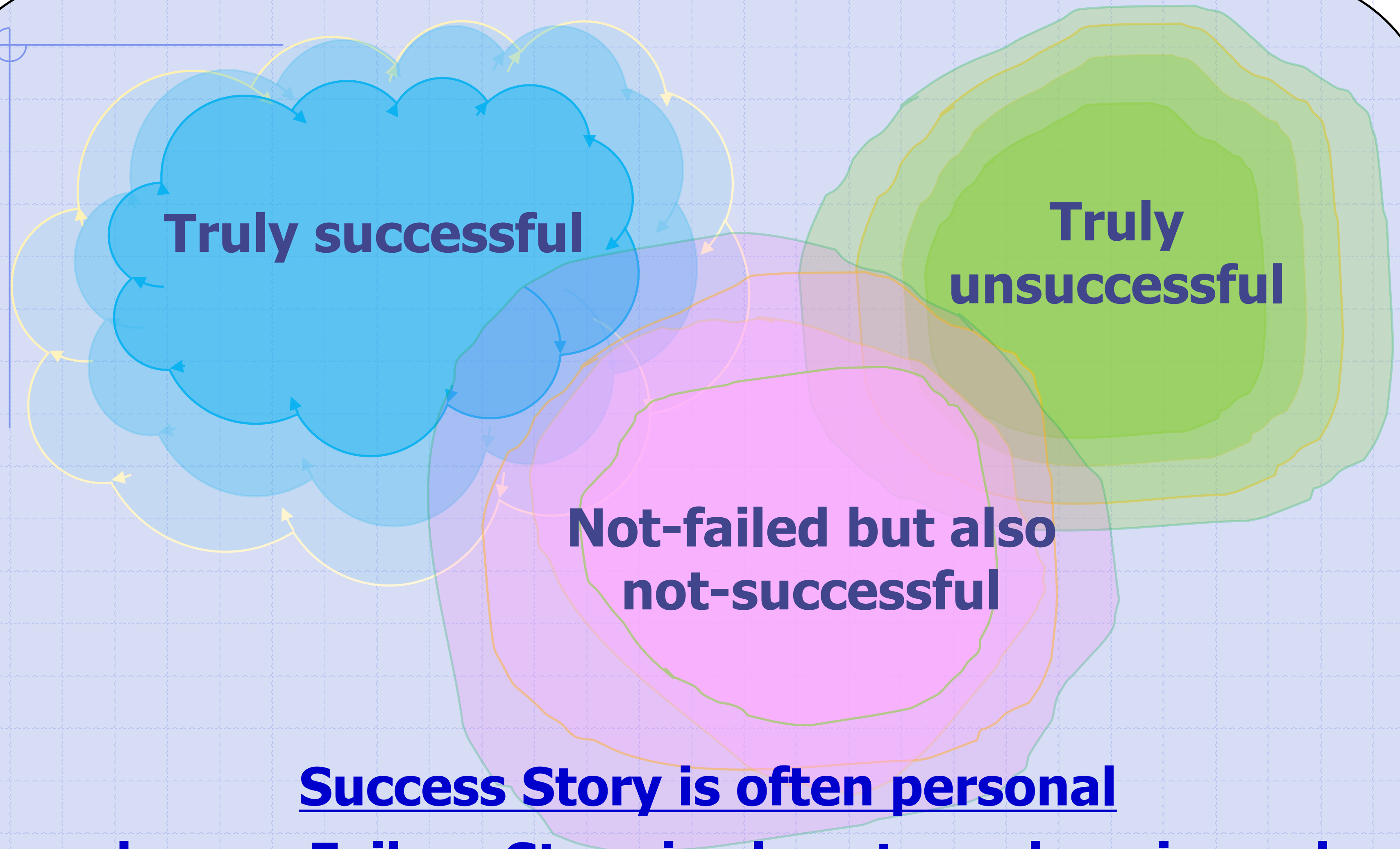
For NOEIC: $\min f(x), \text{s. t. } c_i(x) = 0, i \in E, c_i(x) \geq 0, i \in I.$

Augmented Lagrange Function

$$p_I(x,\lambda,\sigma) = f(x) - \sum_{i=1}^{m_e} \lambda_i c_i(x) + \frac{1}{2}\sigma \sum_{i=1}^{m} c_i(x)^2 + \frac{1}{2\sigma} \sum_{i=m_e+1}^{m} \left[\left(max\left\{0, \sigma c_i(x) - \lambda_i\right\}\right)^2 - \lambda_i^2\right]$$


success
Another way
Fence
Cave
Cliff
Fence

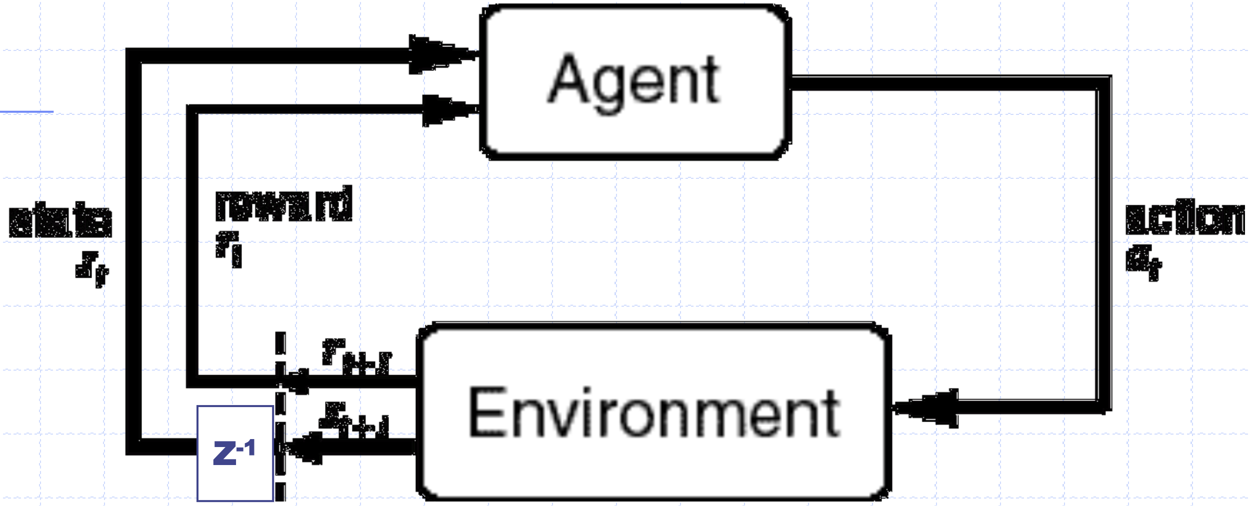
Set of Graduate Students



Success Story is often personal

whereas Failure Story is almost surely universal.

Block Diagram of Reinforcement Learning System



- Immediate rewards: pleasure, fun or pain
- Long-term return: goal achievement=long term accumulation of rewards(+ or -)
- Value Function(state), optimal control policy

THANK YOU FOR ATTENDING