Nonlinear Optimization (NLP)

Mihai Anitescu
Argonne National Laboratory
June 20, 2005

Outline

- What is nonlinear optimization optimization?
- Why? Example of optimization based on a subjective criterion.
- Why? Example of optimization problems derived from variational principles in physics.
- How? Modeling environments and algorithms.
- Thanks – Sven Leyffer.

Nonlinear Optimization-Nonlinear Programming

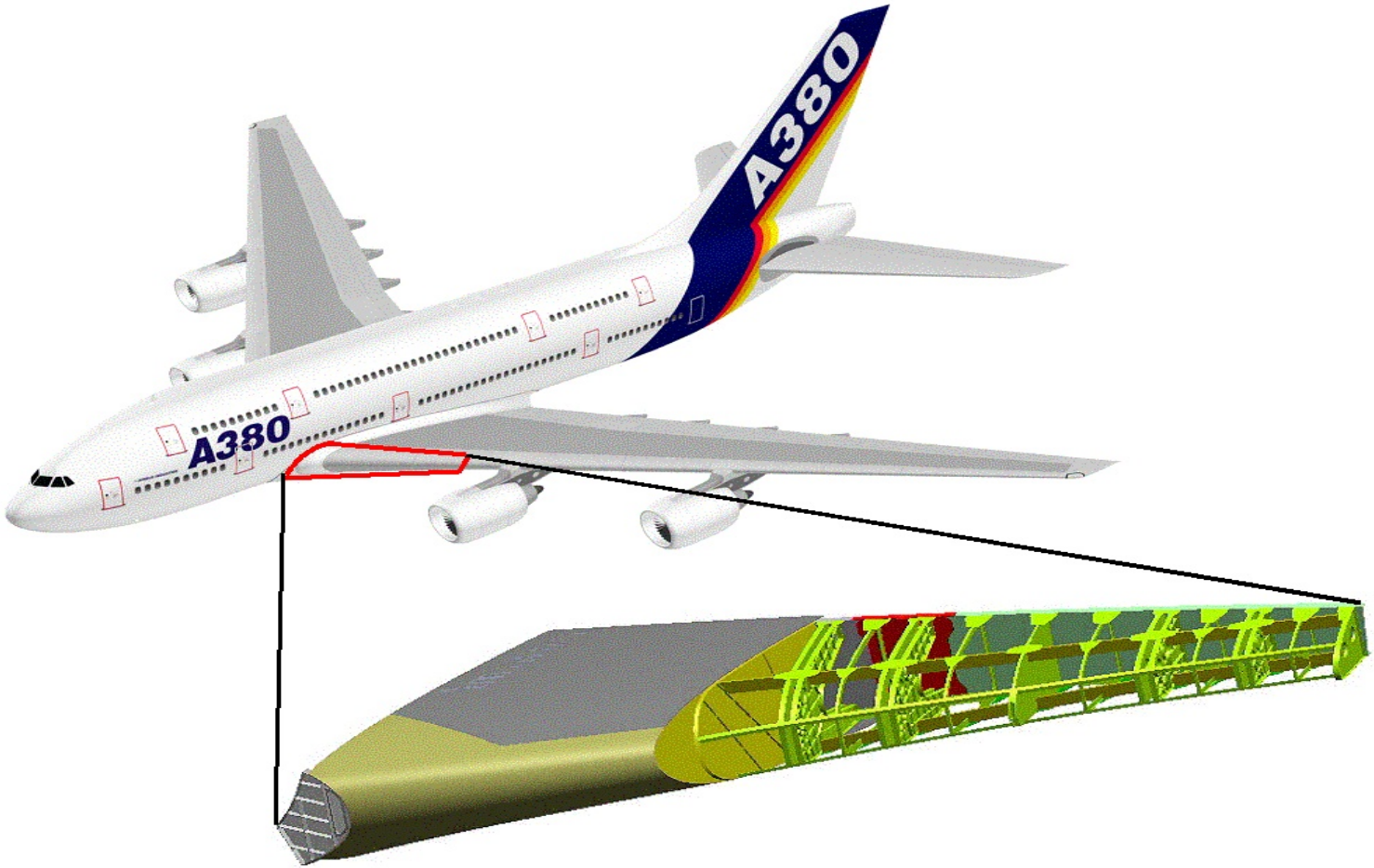
$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & c(x)=0,\ h(x)\le 0 \end{array} \quad \text{(or)} \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & c(x)=0,\ h(x)+y=0 \\ & y\ge 0 \end{array} \quad \text{(or)} \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & c(x)=0 \\ & x\in K \end{array}$$

- The variables *y* are called slacks.
- In the latter case, the "data" functions *f,c,* are not identical with the 2 preceding cases.
- The problem is called **nonlinear** when either *f* or *(c,h)* or both are nonlinear.
- The set K may include integrality constraints, MINLP.
- **The above is a powerful modeling paradigm, in which many problems may be rephrased or approximated, though it is important to exploit the particularities of the problem – the "structure".**

“Subjective” criteria of optimization.

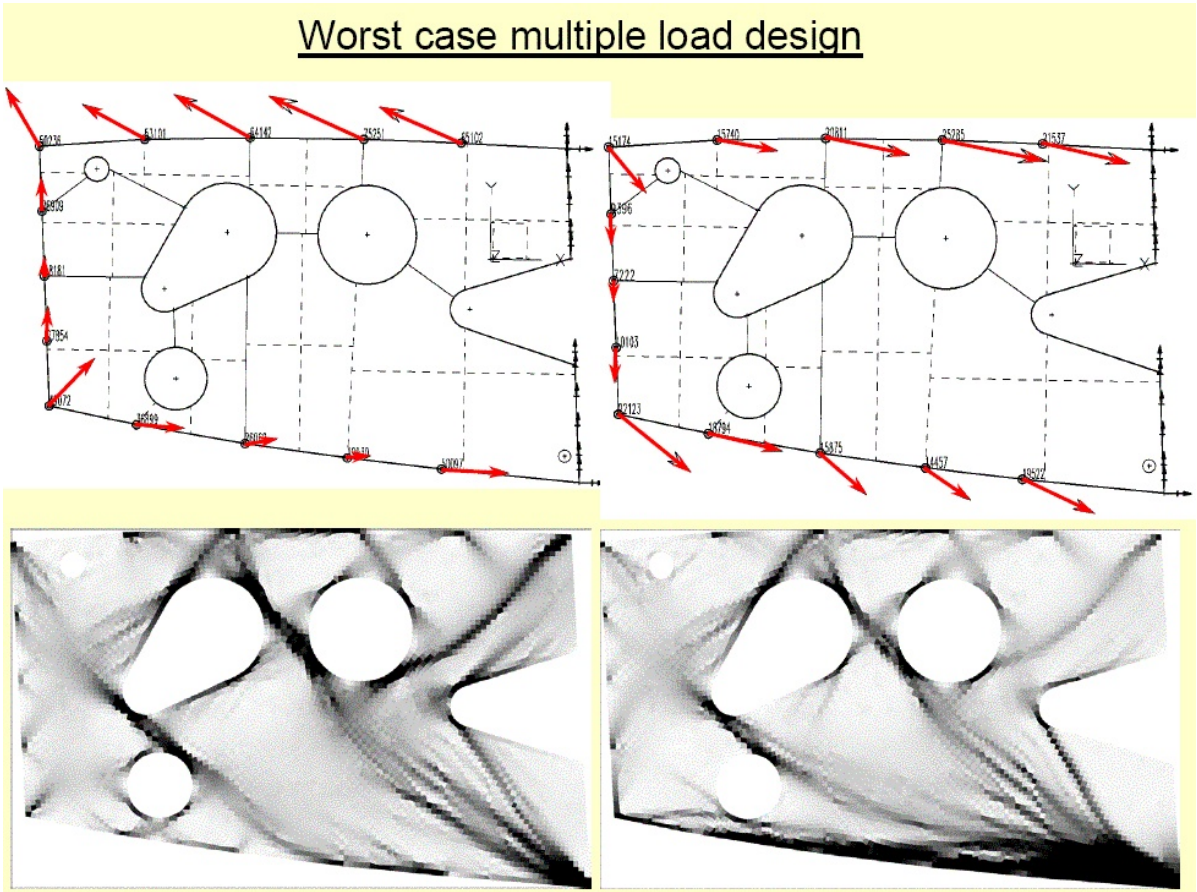
- Trying to optimize an user defined criterion: time to accomplish a task, or cost, or matching criterion ….
- Generally a specialist makes a judgment and defines the criterion based on domain-specific knowledge.
- Optimal behavior and design of Engineering, Financial, Management applications.
- Examples: structural design and phase problems in crystallography.

The Airbus wing



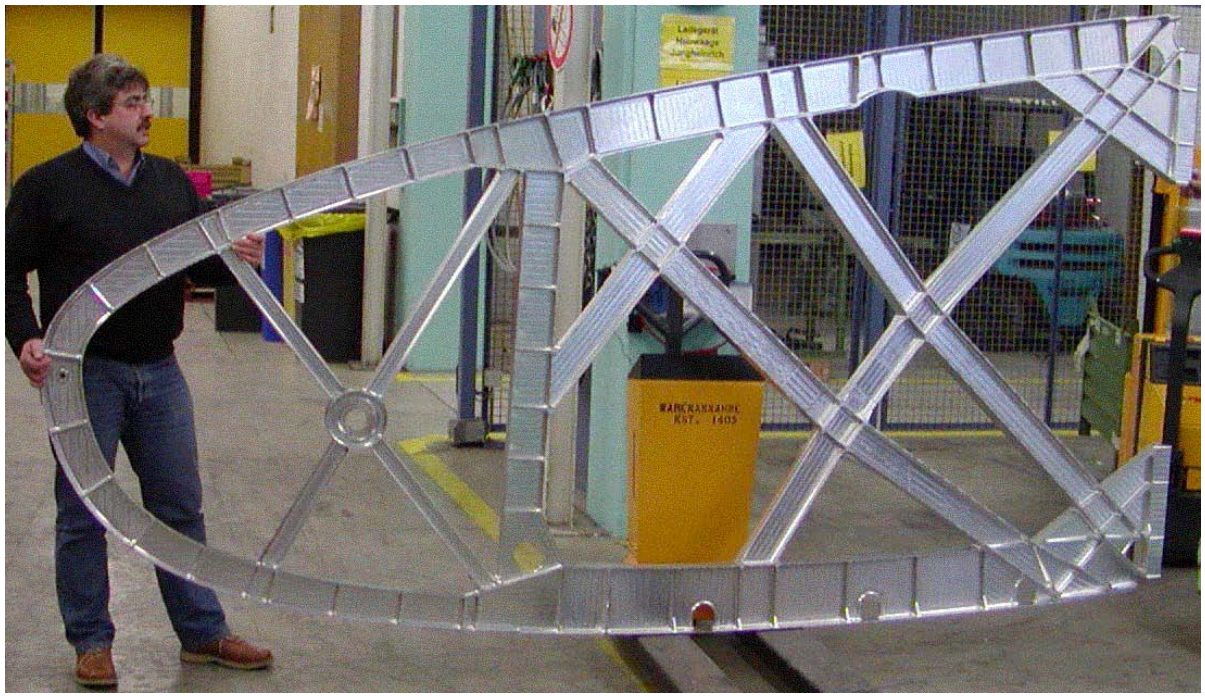
(From Sven Leyffer): Optimizing the inboard inner leading edge ribs.

Design considerations (Kocvara et al.)

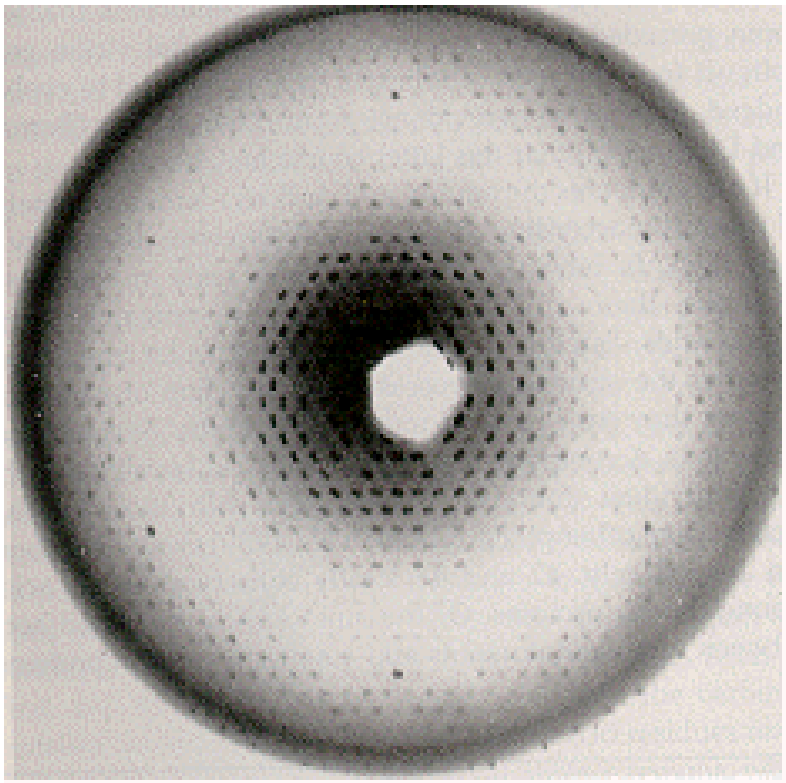


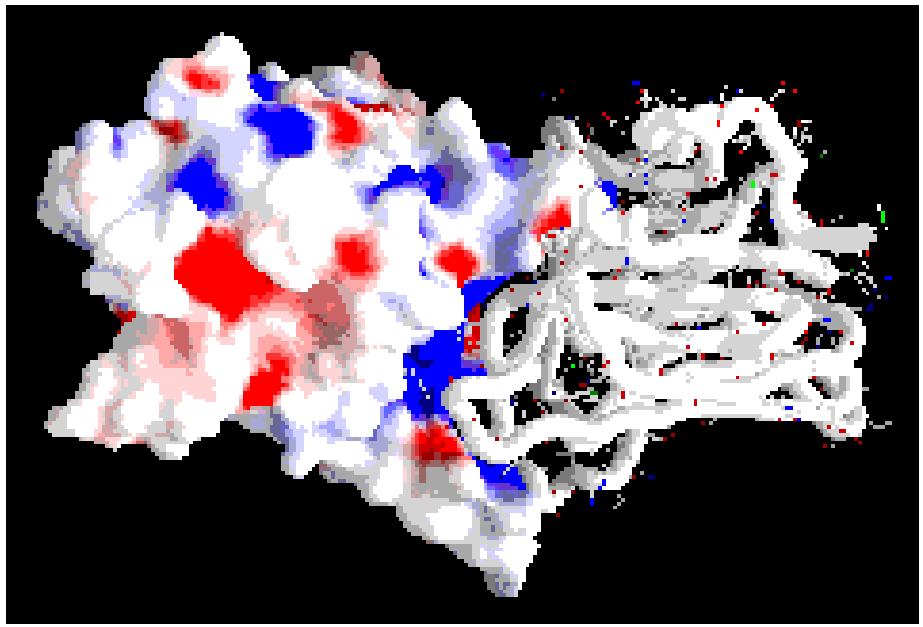
Minimize weight of the structure subject to load and design restrictions

Final Design—”Truss topology”



Xray crystallography





•How do we obtain a 3D structure (right) from its diffraction pattern (left)?

•It is essentially the unique high resolution approach to detect protein structure irrespective of size.

•Problem: Find the atomic distributions that minimizes “discrepancy”.

Phase problem-centrosymmetric (Sahinidis et al.)

Indices

m	index used for reflections ($m = 1, \ldots, M$)
t	index used for triplet invariants ($t = 1, \ldots, T$)

Variables

ϕ_m	phase of the mth reflection
φ_m	normalized phase of the mth reflection equal
ω_t	triplet invariant defined by $\omega_t = \phi_{m_t} + \phi_{m'_t} +$ where $\mathbf{h}_{m_t} + \mathbf{h}_{m'_t} + \mathbf{h}_{m''_t} = \mathbf{0}$
α_t	binary decision variable
β_t	binary decision variable equal to $(1 - \cos\omega_t)$

Parameters

M	number of reflections
n	number of atoms in the unit cell
T	number of invariants
$\lvert E_m \rvert$	structure factor amplitude associated with reflection $\mathbf{h}_m$
A_t	constant equal to $2n^{-1/2}\lvert E_{m_t}\rvert\lvert E_{m'_t}\rvert\lvert E_{m''_t}\rvert$
$\overline{\omega}_t$	conditional expected value of the cosine of the triplet invariant, equal to $I_1(A_t)/I_0(A_t)$ (Germain *et al.*, 1970)

Model M1

$$\begin{aligned} \min \quad & f(\boldsymbol{\beta}) = \frac{\sum_{t=1}^{T} A_t \left(4\beta_t\overline{\omega}_t + (1 + \overline{\omega}_t{}^2 - 2\overline{\omega}_t)\right)}{\sum_{t=1}^{T} A_t} \\ \text{s.t.} \quad & \varphi_{m_t} + \varphi_{m'_t} + \varphi_{m''_t} = 2\alpha_t + \beta_t, \quad t = 1, \ldots, T \quad (1) \\ & \varphi_m \in \{0, 1\}, \qquad m = 1, \ldots, M \\ & \alpha_t, \beta_t \in \{0, 1\}, \qquad t = 1, \ldots, T. \end{aligned}$$

Other applications.

- Data assimilation in weather forecasting → PDE constraint
- Image reconstruction from acoustic wave data → PDE constraint
- Crew scheduling, vehicle routing → integer variables
- Reactor core reloading nonlinear → integer variables
- Radio therapy treatment planning → nonlinear integer
- Oil field infrastructure design → PDE c/s & integer var
- Simulation of competition in electricity markets equilibrium c/s.

Variational Description of Phenomena in Physical Sciences

- In these problems, the “state” variable is the solution of an optimization problem, which is formulated based on a law of physics, rather than a subjective criterion.
- In electronic structure computation: the electronic density.
- In complex fluids the (density and species) distributions at thermodynamic equilibrium.
- In Hamiltonian systems, the trajectory is the solution of an optimization problem
- Fermat's variational principle states that a signal in anisotropic media propagates between two points along a curve which renders Fermat's functional $I(l)$ stationary
- ….

(Thomas-Fermi) Density Functional Theory

Problem: For a given atomic configuration, determine electronic density from using the variational principle.

$$\min_{\rho} E\left[\rho, \left\{\mathbf{R}_A\right\}\right]$$

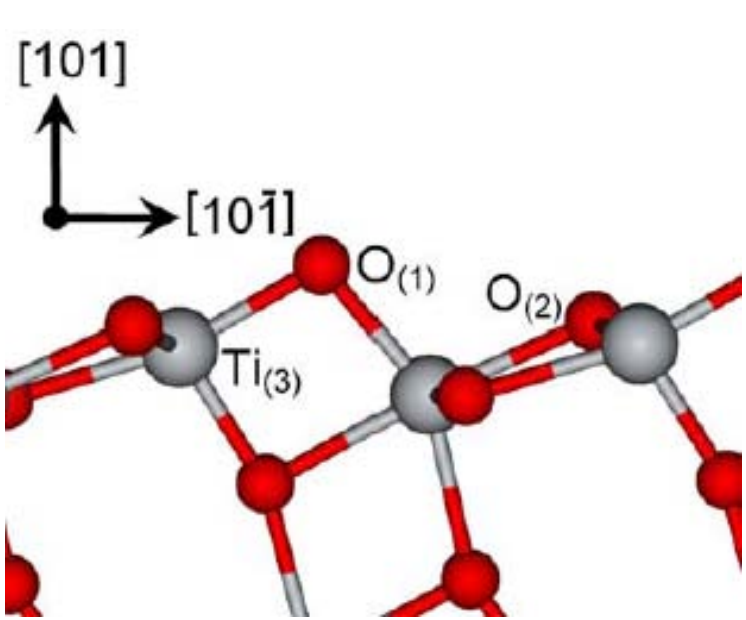
$$\text{s.t.} \int \rho = N_e$$

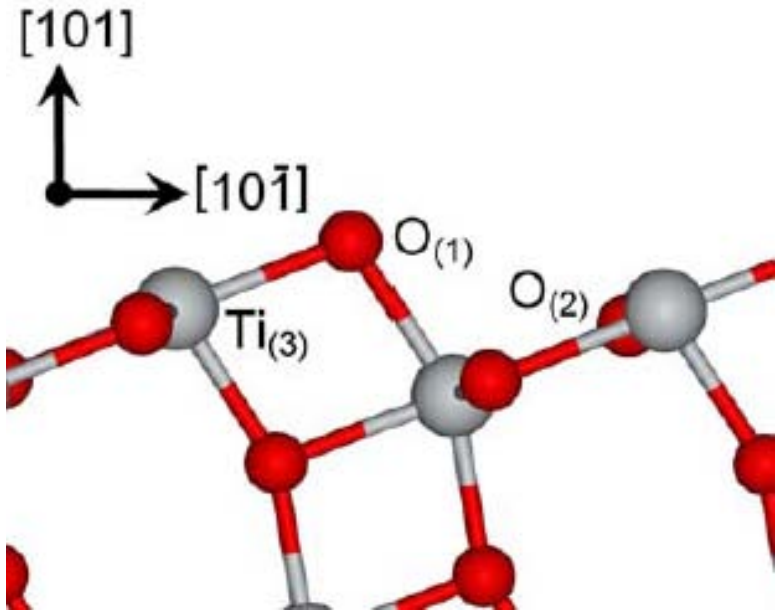
$$E\left[\rho, \left\{\mathbf{R}_A\right\}\right] = E_{ne}\left[\rho, \left\{\mathbf{R}_A\right\}\right] + J\left[\rho\right] + K\left[\rho\right] + T\left[\rho\right] + V_{nn}\left(\left\{\mathbf{R}_A\right\}\right)$$

$$T\left[\rho\right] = C_F \int \rho^{\frac{5}{3}(\mathbf{r})}\, d\mathbf{r}, \quad K\left[\rho\right] = -C_x \int \rho^{\frac{4}{3}(\mathbf{r})}\, d\mathbf{r}$$

$$E_{ne}\left[\rho, \left\{\mathbf{R}_A\right\}\right] = \sum_{A=1}^{M} \int \frac{Z_A\, \rho(\mathbf{r})}{\left\|\mathbf{R}_A - \mathbf{r}\right\|} d\mathbf{r}, \quad J\left[\rho\right] = \frac{1}{2} \int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{\left\|\mathbf{r} - \mathbf{r}'\right\|} d\mathbf{r}\, d\mathbf{r}'$$

Surface structure of the TiO2 nanoparticle





Computations carried out by Peter Zapol et al. from MSD-Argonne, using Kohn-Sham DFT.

Time-dependent variational problems

- We approximate them by a well-chosen numerical scheme.
- For Hamiltonian Systems in molecular dynamics, used by Schlick et al.
- Or the evolution is sufficiently slow that the problem can be considered quasistatic – in effect, a homotopy.

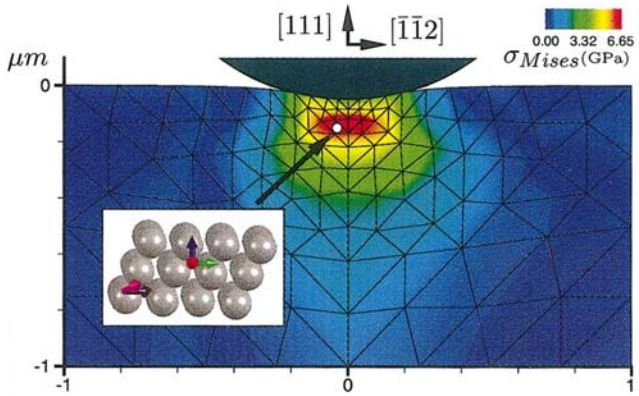
$$\dot{u} = F(u),$$

$$u_{k+1} = u_k + h\tilde{F}(u_k, u_{k+1}) = \nabla_{u_{k+1}} \Phi(u_k, u_{k+1})$$

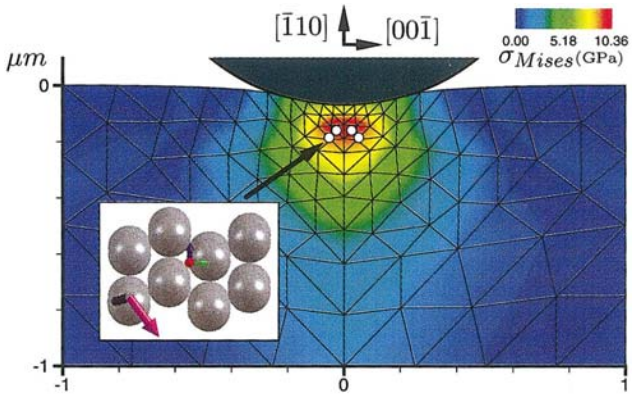
$$\rho(t) = \min_\rho E\left[\rho, \{\mathbf{R}_A\}(t), L(t)\right]$$

$$\text{s.t.} \int \rho(t) = N_e$$

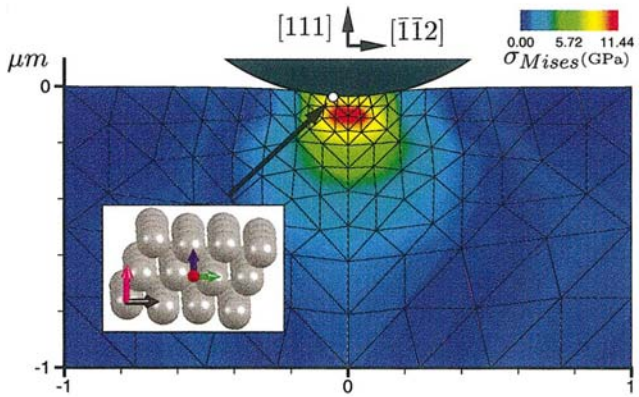
Nano-indentation



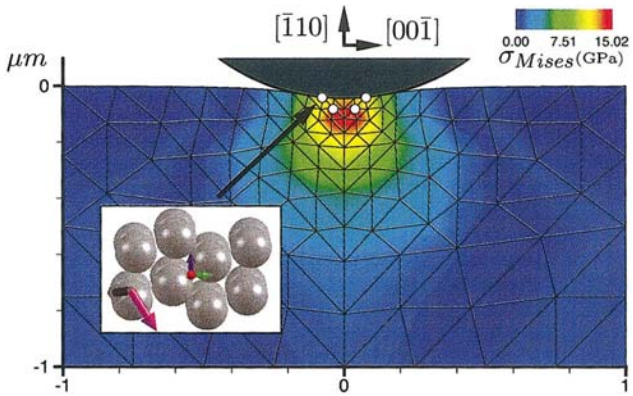
(a)



(a)



(b)



(b)

- One of the “hot pursuits” in mechanical engineering: Simulating complex phenomena starting from first principles, as opposed to empirical potentials.
- Density Functional Theory based defect nucleation (Carter, Ortiz, et al.)

Multi-rigid-body dynamics with contact and friction

- A subject pursued by the author of this presentation for some time.
- Essential in the study of robotics, granular materials, pharmaceutical drug processing (powders).
- The velocity of the system at the next step is the solution of the minimum energy problem subject to nonpenetration and frictional constraints.
- Example: the study of size-based segregation in granular materials.

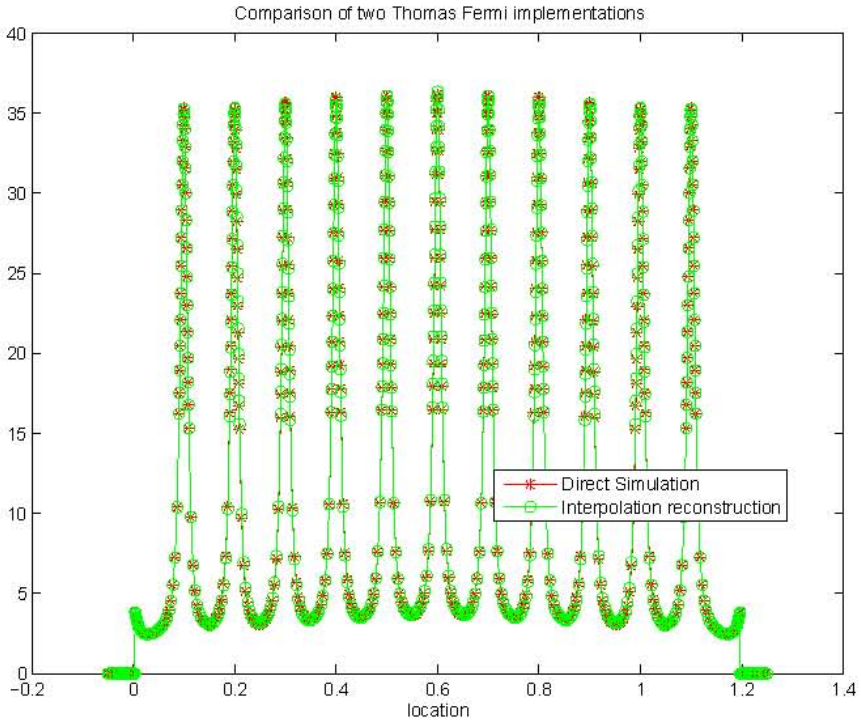
$$
\begin{aligned}
v^{(l+1)} = \operatorname{argmin}_{\widehat{v}} \; & \frac{1}{2}\widehat{v}^T M \widehat{v} + k^{(l)^T}\widehat{v} \\
\text{subject to } & \frac{1}{h}\Phi^{(j)}(q^{(l)}) + \nabla\Phi^{(j)^T}\widehat{v} + \mu^{(j)} d_k^{(j)^T}\widehat{v} \geq 0, \\
& j \in \mathcal{A}(q^{(l)}, \epsilon), \quad k = 1, 2, \ldots, m^{(j)},
\end{aligned}
$$

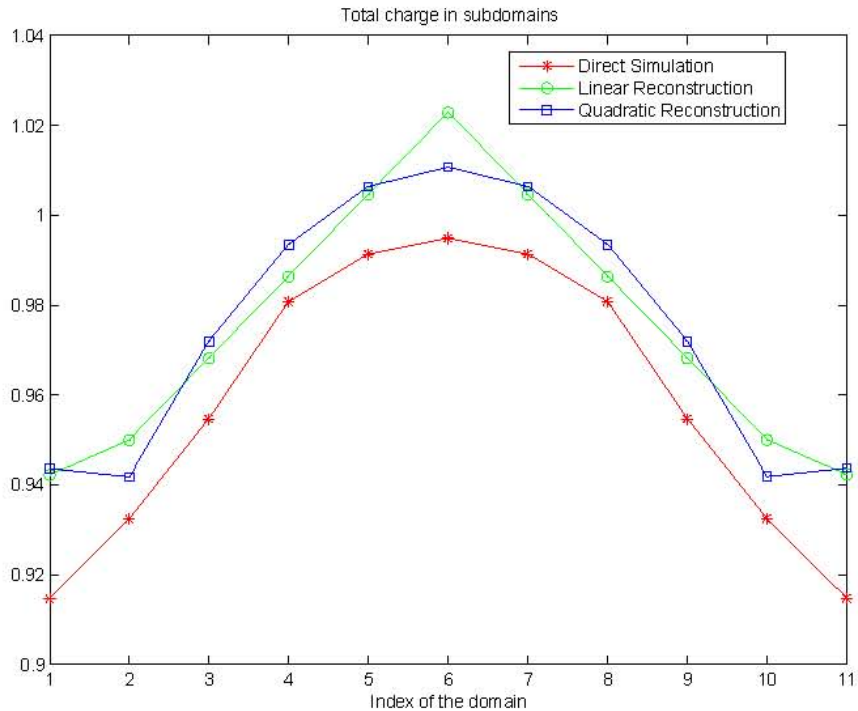
Modeling (Nonlinear) Optimization Problems

AMPL & GAMS

- high level languages for nonlinear optimization
- interpret problem description, interface to solvers & returns results
- details of solver, derivatives & pre-solve are hidden from user
- modeling language (e.g. var, minimize, subject to, ...)
- programming language (e.g. while, if, ...)

AMPL example





- One-dimensional Thomas-Fermi problem.
- Once you have created the model, you can even run it over the internet with the NEOS server.

AMPL for Thomas Fermi DFT

```
#user-defined parameters
param n integer;
param ma integer;
param mg integer;
param dist;
param ratioGap;
param delta;
param cutoff integer;
param pi;
param Z;

#parameters of Thomas Fermi model
param CF:=0.3*(3*pi*pi)^(2/3);
param CX:=0.75*(3/pi)^(1/3);
param indexCutoff=2*(ma+mg)*cutoff;
#total number of nodes
param N:=(n+2)*(2*ma+2*mg);
param ZA{i in 1..n}=Z;

param Nelec:=sum{i in 1..n} ZA[i];

param atomicPosition{i in 1..n}:=dist*i; #atomic positions
param xloc{i in 1..(ma+mg)}:= if i <= ma then (i-0.5)/ma*ratioGap*dist*0.5 else (ma-0.5)/ma*ratioGap*dist*0.5 + 0.5*(1-ratioGap)*dist *(i-ma)/mg;
param xi{i in 1..N}:= if (i-2*(ma+mg)*floor((i-1)/(2*(ma+mg)))-ma-mg-0.5) < 0
            then floor((i-1)/(2*(ma+mg)))*dist-xloc[abs(i-2*(ma+mg)*floor((i-1)/(2*(ma+mg)))-ma-mg-1)]
            else  floor((i-1)/(2*(ma+mg)))*dist+xloc[i-2*(ma+mg)*floor((i-1)/(2*(ma+mg)))-ma-mg];
var x;
var rho{i in 1..N} >=0;
```

Objective function

minimize obj: sum{i in 1..(N-1)} 0.5*(CF*(rho[i]+delta)^(5/3)-CX*(rho[i]+delta)^(4/3) + CF*(rho[i+1]+delta)^(5/3)-CX*(rho[i+1]+delta)^(4/3))*(xi[i+1]-xi[i])

#Kinetic and Exchange

- sum{i in 1..(N-1), j in 1..n} ZA[j]*(rho[i+1]/sqrt((xi[i+1]-atomicPosition[j])^2+delta)+rho[i]/sqrt((xi[i]-atomicPosition[j])^2+delta))*0.5*(xi[i+1]-xi[i])

Electron-Nucleus

+ 0.5 * 0.25*sum{i in 1..(N-1), j in max(1,i-indexCutoff)..min(i+indexCutoff,N-1)} (xi[i+1]-xi[i])*(xi[j+1]-xi[j])*(rho[i]*rho[j]/sqrt((xi[i]-xi[j])^2+delta) +

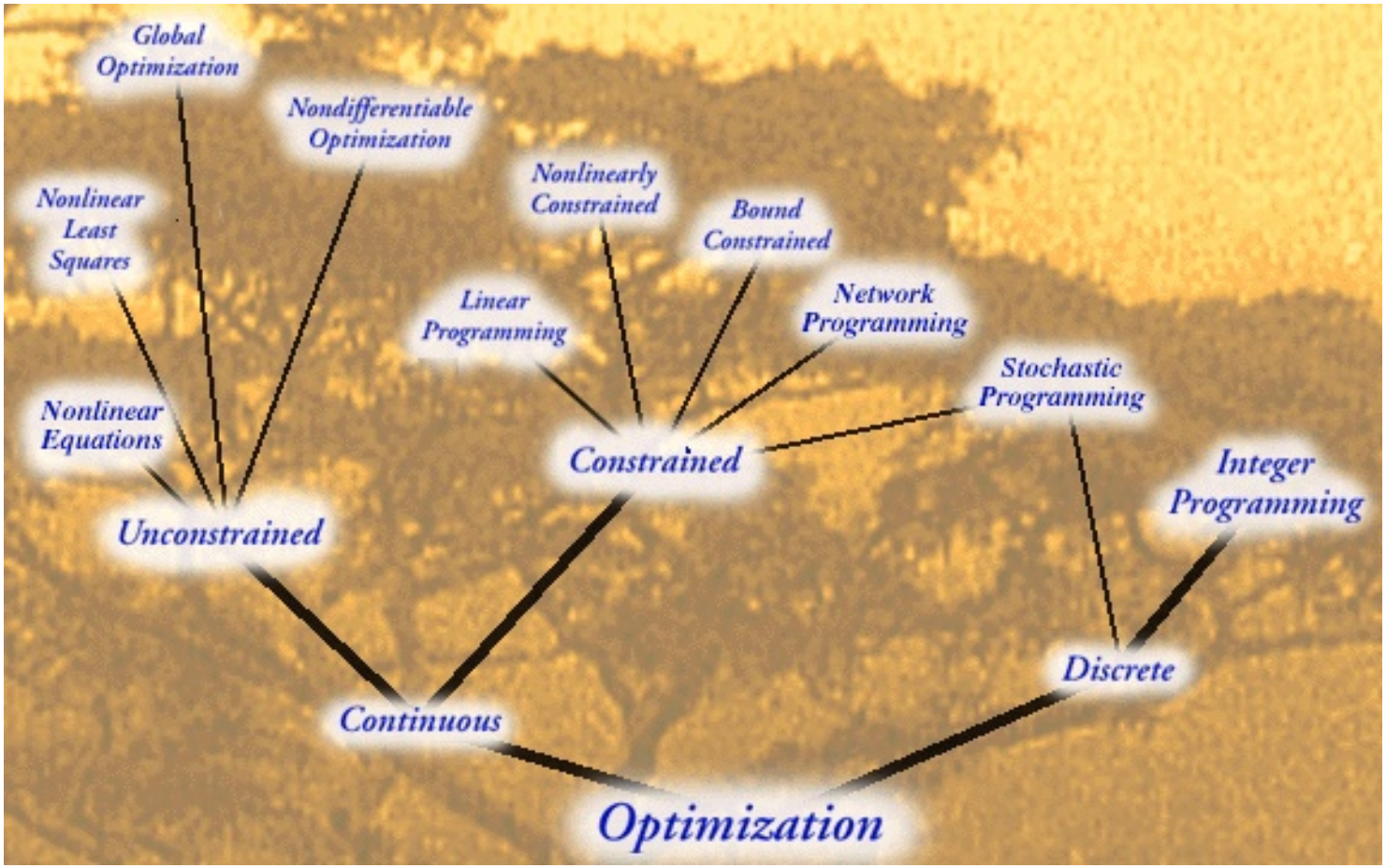
rho[i]*rho[j+1]/sqrt((xi[i]-xi[j+1])^2+delta) +rho[i+1]*rho[j]/sqrt((xi[i+1]-xi[j])^2+delta) + rho[i+1]*rho[j+1]/sqrt((xi[i+1]-xi[j+1])^2+delta));

Electron-electron

subject to

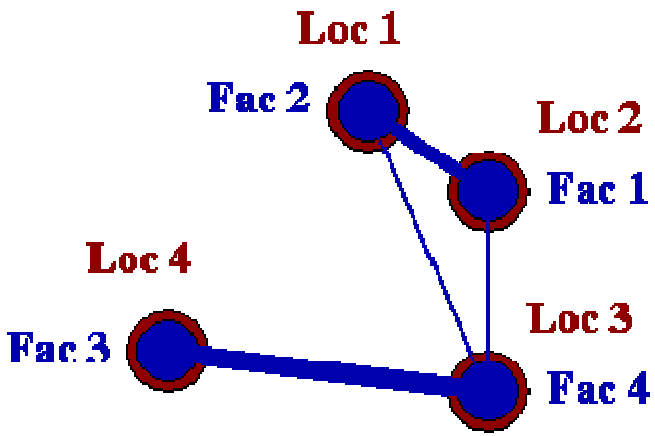
constr: sum{i in 1..(N-1)} 0.5*(xi[i+1]-xi[i])*(rho[i]+rho[i+1])=Nelec;

Simple Taxonomy



Isn’t any problem reducible to NLP? Sure, but it is very efficient to recognize and exploit structure.

The Quadratic Assignment (facility location) Problem



$$\min_{p \in \Pi} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} d_{p(i)p(j)}$$

- How does one place n facilities at n location such that the total cost is minimized?
- Solved by the METANEOS team (Linderoth, Goux, Wright) and friends.
- The branch-and-bound procedure IDEAL for distributed computing.
- The solution of the NUG30 – a 30 year open problem made the headlines.

NUG30 statistics

Table 2: nug30 Run Statistics	
Average number of available workers	652.7
Maximum number of available workers	1009
Running wall clock time (sec)	597,872
Total cpu time (sec)	346,640,860
Average machine speed	0.56
Minimum machine speed	0.045
Maximum machine speed	1.074
Equivalent CPU time (sec) on an HP-C3000	218,823,577
Parallel Efficiency	93%
Number of times a machine joined the computation	19,063

What we (optimizers) still need to figure out

- Global NONCONVEX (including integer) optimization problems are still far from being systematically solved on the large scale (many times we dance of joy when we solve N=30 ☺).
- Iterative solvers for constrained optimization beyond PDE constraints --- the preconditioning problems.
- There always appear new classes of problems that do not satisfy the usual assumptions – for example complementarity constraints.