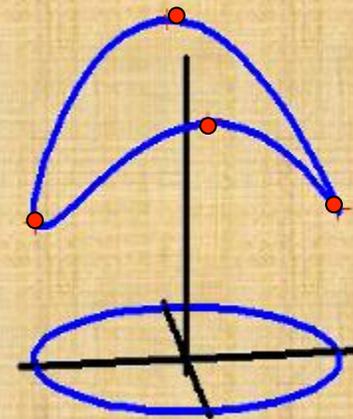
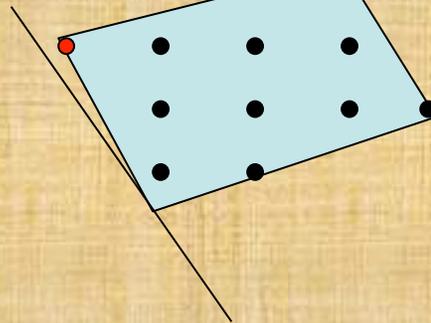
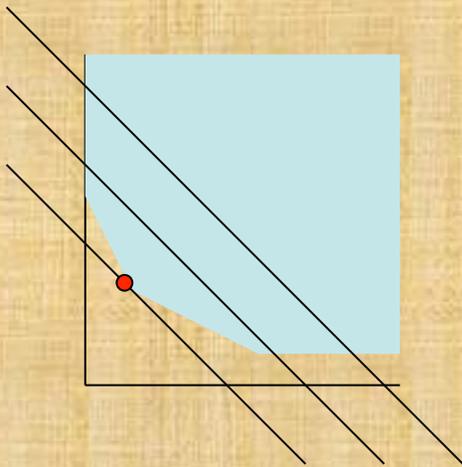


USNA Math Colloquium

The Algebra of Optimization

Will Traves

Thursday 06 NOV 2003 at 1345 in Chauvenet 220

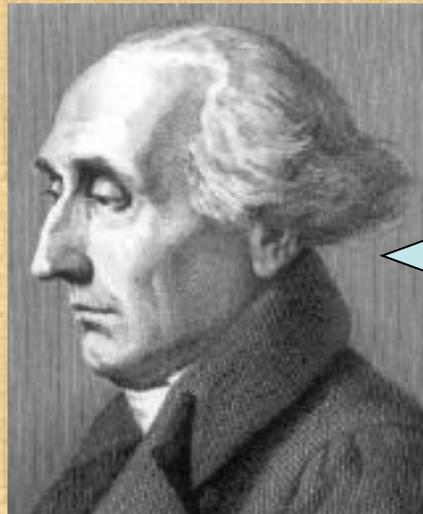


Optimization Using Lagrange Multipliers

A typical problem in SM223 asks for the maximum of the function $f(x,y,z) = x^3 + 2xyz - z^2$ subject to the constraint $g(x,y,z) = x^2 + y^2 + z^2 - 1 = 0$.

J.-L. Lagrange showed that this can be done by solving the associated system of equations $\nabla f = \lambda \nabla g, g = 0$:

$$\begin{cases} 3x^2 + 2yz = 2x\lambda \\ 2xz = 2y\lambda \\ 2xy - 2z = 2z\lambda \\ x^2 + y^2 + z^2 = 1 \end{cases}$$



Those look hard.

I need Gröbner Bases to solve the system!

Gröbner Basics

A Gröbner Basis is a representation of a system of equations that may contain redundant equations, but that encapsulates a lot of information about the system and its solutions.



1. Ideals
2. Monomial Orderings
3. Division
4. Leading Term Ideal
5. Elimination Theory
6. Solving Lagrange's Equations

Bruno Buchberger named Gröbner Bases after his Ph.D. advisor, Wolfgang Gröbner.

Gröbner Basics

A Gröbner Basis is a representation of a system of equations that may contain redundant equations, but that encapsulates a lot of information about the system and its solutions.

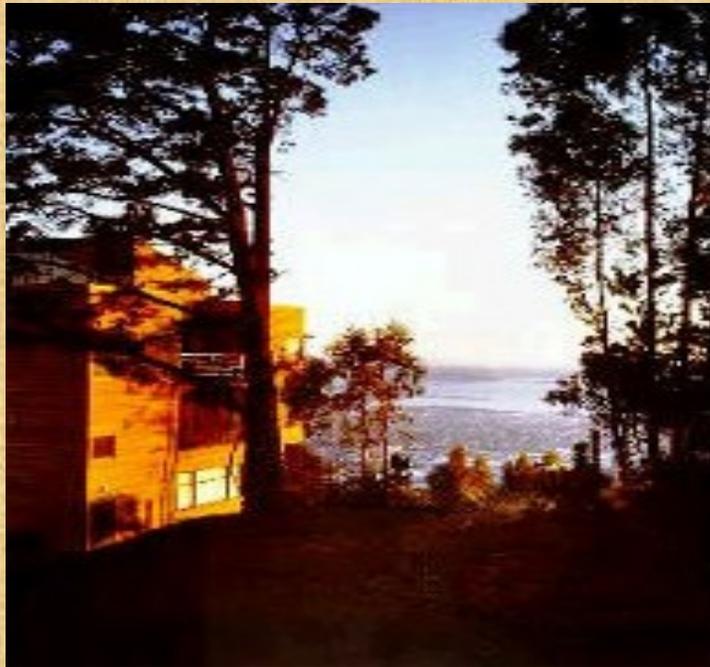


1. Ideals
2. Monomial Orderings
3. Division
4. Leading Term Ideal
5. Elimination Theory
6. Solving Lagrange's Equations

RISC is located in a castle in Austria. It was donated by the Austrian government to recognize industrial applications of mathematics.

Gröbner Basics

A Gröbner Basis is a representation of a system of equations that may contain redundant equations, but that encapsulates a lot of information about the system and its solutions.



1. Ideals
2. Monomial Orderings
3. Division
4. Leading Term Ideal
5. Elimination Theory
6. Solving Lagrange's Equations

MSRI (Berkeley) held a year-long program in Commutative Algebra and its applications in 2002.

Ideals

Given polynomials f_1, \dots, f_s we consider the system of equations

$$f_1(x_1, \dots, x_n) = 0, \dots, f_s(x_1, \dots, x_n) = 0. \quad (*)$$

A **polynomial consequence** is a polynomial $g_1 f_1 + g_2 f_2 + \dots + g_s f_s$.

Solving (*) is equivalent to solving the system

$$\{g=0: g \text{ is a polynomial consequence of } f_1, \dots, f_s\}.$$

The **ideal** (f_1, \dots, f_s) is the set of all polynomial consequences of the f_i 's.



The f_i 's are called basis elements for the ideal.

Example: The two ideals $(xy+1, y^2-1)$ and $(xy+1, y^2-1, -x-y)$ are equal. The second set of generators is more desirable – it is a Gröbner Basis.

Monomial Orders

In order to define Gröbner Bases we introduce an ordering on monomials $x^a y^b z^c$ in $\mathbf{C}[x,y,z]$. A monomial ordering is an order that is:

TOTAL: for every pair of monomials m_1 and m_2 we have either $m_1 \geq m_2$ or $m_2 \leq m_1$.

ARTINIAN: every set of monomials has a smallest monomial.

MULTIPLICATIVE: if $m_1 \geq m_2$ then $m \cdot m_1 \geq m \cdot m_2$ for any monomial m .

Example: Order monomials x^n by degree $x^n > x^m$ if $n > m$.

Lex Order

We say that

$$x^a y^b z^c > x^d y^e z^f$$

if

$$a > d \text{ or}$$

$$a=d \text{ and } b > e \text{ or}$$

$$a=d \text{ and } b=e \text{ and } c > f.$$

This says that the bigger monomial is determined by degree in the left-most variables (eg. $x^2 > xyz$).

Definition: The **leading term** of a polynomial is the largest monomial in the sum:

$$\text{LT}(5xy+3x^2) = x^2.$$

Division Process

Can use Leading Terms to “divide” $f = xy^2 - x$ by $xy+1$ and y^2-1 .

```
r=0
While f != 0
  if LT(fi) divides LT(f)
    for some fi then
      subtract a multiple of
      fi from f to kill LT(f).
  else
    r=r+LT(f)
    f=f-LT(f)
  endif
End
Return(r)
```

LT($xy+1$) divides LT(xy^2-x)

so change $f = xy^2-x$ to

$$(xy^2-x) - y(xy+1) = -x-y.$$

Since LT($-x-y$) is not divisible by LT($xy+1$) or LT(y^2-1), add $-x$ to r .

Now $f = -y$.

Since LT($-y$) is not divisible by LT($xy+1$) or LT(y^2-1), add $-y$ to r .

Now $f = 0$. Output $r = -x-y$.

Division Process

Can use Leading Terms to “divide” $f = xy^2 - x$ by $xy+1$ and y^2-1 .

But $f = xy^2 - x = x(y^2 - 1)$ is in the ideal $(xy + 1, y^2 - 1)$.

It would be nice if the remainder was zero when f is in the ideal.

We can correct this by using a “good” set of generators for our ideal – a Gröbner Basis.

$LT(xy+1)$ divides $LT(xy^2-x)$
so change $f = xy^2-x$ to
 $(xy^2-x) - y(xy+1) = -x-y$.

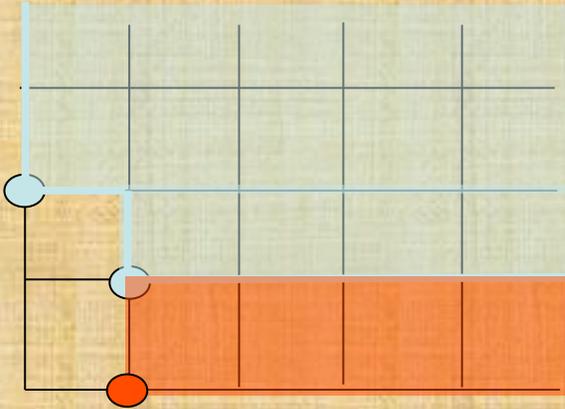
Since $LT(-x-y)$ is not divisible by $LT(xy+1)$ or $LT(y^2-1)$, add $-x$ to r .
Now $f = -y$.

Since $LT(-y)$ is not divisible by $LT(xy+1)$ or $LT(y^2-1)$, add $-y$ to r .
Now $f = 0$. Output $r = -x-y$.

Leading Term Ideal

The **leading term ideal** of an ideal I is $LT(I) = (LT(g) : g \text{ in } I)$.

Example: $LT((xy + 1, y^2 - 1))$ is not just (xy, y^2) because $-x-y$ is in $(xy+1, y^2-1)$ so x is in $LT((xy + 1, y^2 - 1))$.



Definition: a **Gröbner Basis** for the ideal $I=(h_1, h_2, \dots, h_s)$ is a collection of polynomials g_1, g_2, \dots, g_t such that $(g_1, g_2, \dots, g_t) = I$ and $(LT(g_1), LT(g_2), \dots, LT(g_t)) = LT(I)$.

Example: A Gröbner Basis for the ideal $(xy+1, y^2-1)$ is given by $[xy+1, y^2-1, -x-y]$. In fact, this is redundant. Another Gröbner Basis is given by $[y^2-1, -x-y]$.

Remainders are Well-Defined

Theorem: If $G=[g_1, g_2, \dots, g_t]$ is a Gröbner Basis for I and $f \in \mathbf{C}[x_1, \dots, x_n]$ then the remainder $\text{Rem}(f, G)$ of f upon division by G is **well-defined** and $f \in I$ if and only if $\text{Rem}(f, G) = 0$.

Example: Dividing $xy^2 - x$ by the Gröbner Basis $[y^2 - 1, -x - y]$ for the ideal $(xy+1, y^2-1)$, we obtain

$$\begin{aligned} xy^2 - x &= -y^2(-x-y) + (-y^3 - xy) \\ &= -y^2(-x-y) + y(-x-y) + (-y^3 + y^2) \\ &= -y^2(-x-y) + y(-x-y) + (-y)(y^2 - 1) + 0. \end{aligned}$$

The remainder is 0.

Elimination Theory

Solving the system
$$\begin{cases} 2y - x = 0 \\ xy - 3y - 2 = 0 \end{cases}$$

Can be accomplished by **eliminating** x ,

$$xy - 3y - 2 = 2y^2 - 3y - 2 = 0,$$

and solving for y .

The polynomial $2y^2 - 3y - 2$ is an element of

$$(2y - x, xy - 3y - 2) \cap \mathbf{C}[y],$$

the ideal of all polynomials obtained by eliminating x .

We call this ideal **Elim** $((2y - x, xy - 3y - 2), x)$.

Elimination with Gröbner Bases

Elimination Theorem: If $[g_1, \dots, g_t]$ is a Gröbner Basis for I in **Lex order** then a Gröbner Basis for $I \cap C[x_k, x_{k+1}, \dots, x_n]$ is just given by those g_i that only involve the variables x_k, \dots, x_n .

Elimination and back substitution can then be used to solve the associated system of equations

$$\begin{cases} g_1(x_1, \dots, x_n) = 0 \\ \vdots \\ g_t(x_1, \dots, x_n) = 0. \end{cases}$$

A Hard Example

$$\left\{ \begin{array}{l} 3x^2 + 2xy = 2x\lambda \\ 2xz = 2y\lambda \\ 2xy - 2z = 2z\lambda \\ x^2 + y^2 + z^2 = 1 \end{array} \right. \Rightarrow GB = \left\{ \begin{array}{l} \lambda - \frac{3}{2}x - \frac{3}{2}y - \frac{167616}{3835}z^6 - \frac{36716}{590}z^4 - \frac{134419}{7670}z^2 = 0 \\ x^2 + y^2 + z^2 - 1 = 0 \\ xy - \frac{19584}{3835}z^5 + \frac{1999}{295}z^3 - \frac{6403}{3835}z = 0 \\ xz + yz^2 - \frac{1152}{3835}z^5 + \frac{108}{295}z^3 - \frac{2556}{3835}z = 0 \\ y^3 + yz^2 - y - \frac{9216}{3835}z^5 + \frac{906}{295}z^3 - \frac{2562}{3835}z = 0 \\ y^2z - \frac{6912}{3835}z^5 + \frac{827}{295}z^3 - \frac{3839}{3835}z = 0 \\ yz^3 - yz - \frac{576}{59}z^6 + \frac{1605}{118}z^4 - \frac{453}{118}z^2 = 0 \\ z^7 - \frac{1763}{1152}z^5 + \frac{655}{1152}z^3 - \frac{11}{288}z = 0 \end{array} \right.$$

$$z^7 - \frac{1763}{1152}z^5 + \frac{655}{1152}z^3 - \frac{11}{288}z = z(z^2 - 1)(z^2 - 4/9)(z^2 - 11/128) = 0.$$

Implicitization

$$x(t) = \frac{2t}{t^2 + 1}$$

$$y(t) = \frac{t^2 - 1}{t^2 + 1}$$

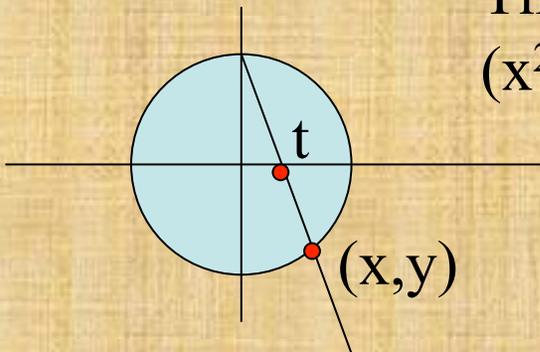
What is the equation (in x and y) of the curve?

Compute the Gröbner basis for the ideal

$$((t^2 + 1)x - 2t, (t^2 + 1)y - (t^2 - 1))$$

in the ring $k[t, x, y]$ with lex order $t > x > y$.

The intersection of this ideal with $k[x, y]$ is just $(x^2 + y^2 - 1)$. This shows that we have a circle.



Linear Programming

Linear programming solves optimization problems with:

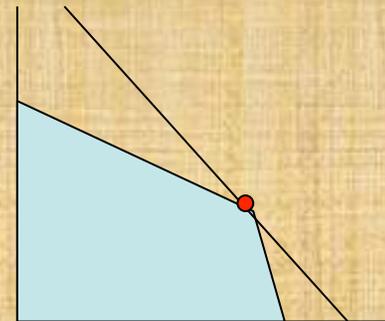
- linear constraints $a_1 u_1 + a_2 u_2 + \cdots + a_n u_n \leq b$
- linear objective functions $c_1 u_1 + c_2 u_2 + \cdots + c_n u_n$
- non-negativity conditions $u_i \geq 0$.

Maximize $u_1 + u_2$ subject to

$$3u_1 + u_2 \leq 6$$

$$10u_1 + 2u_2 \leq 32$$

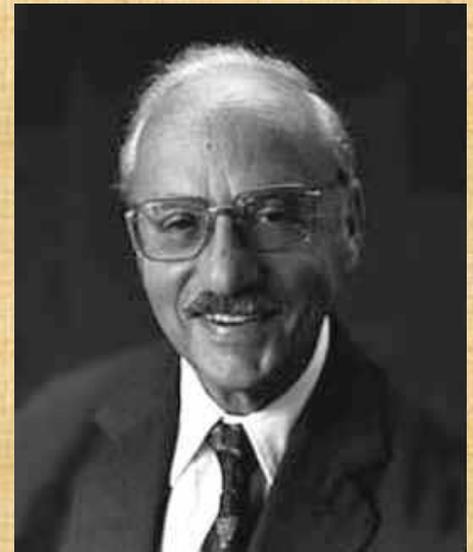
$$u_1 \geq 0, u_2 \geq 0.$$



Simplex Algorithm

FACT: In linear programming problems the optimum is always achieved at a “corner” of the feasible region.

The **Simplex Algorithm** (Dantzig, '49) constructs a path that moves from corner vertices to neighboring corner vertices in such a way that the objective function never decreases. Thus, we meander about the outside of the feasible region.



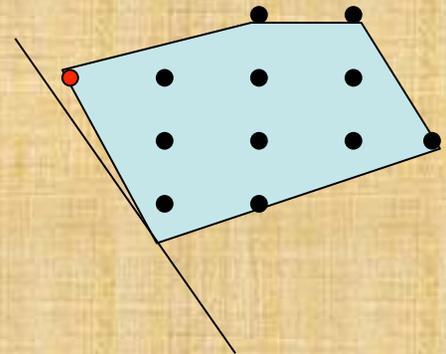
Interior point methods (Karmarkar) have recently been shown to be faster. In these more complicated methods, the path is allowed to traverse the interior of the feasible region.

Integer Programming

Integer programming solves optimization problems similar to those in linear programming, except that we require the solutions to be integer-valued

- linear constraints $a_1 u_1 + a_2 u_2 + \dots + a_n u_n \leq b$
- linear objective functions $c_1 u_1 + c_2 u_2 + \dots + c_n u_n$
- non-negativity conditions $u_i \geq 0$
- **each u_i is an integer**

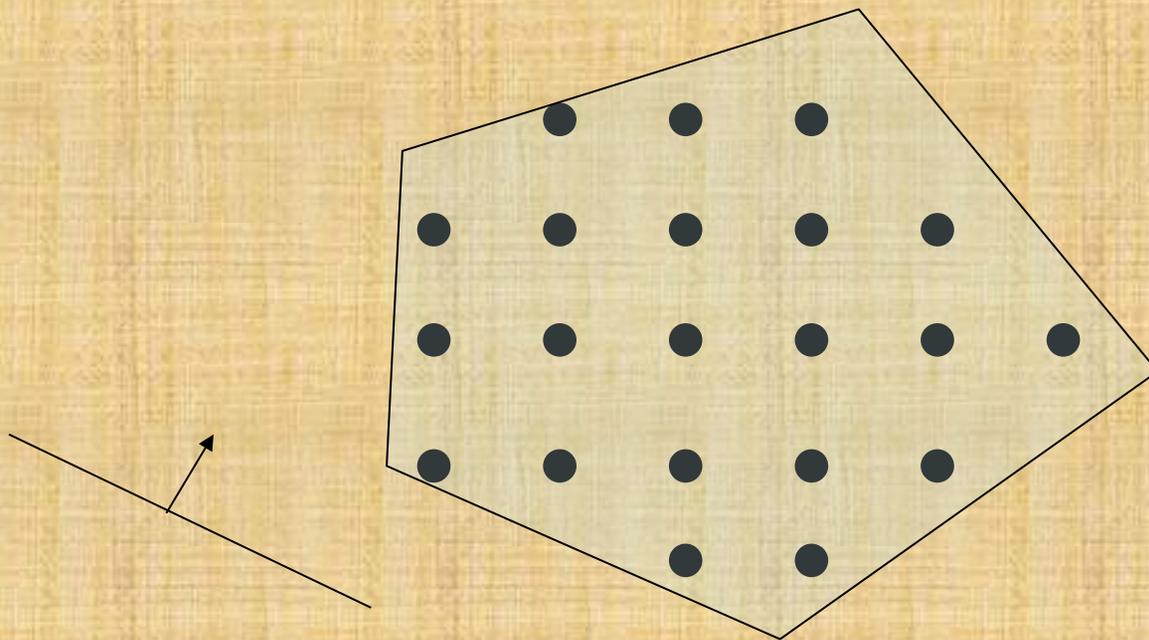
The simplex method (and interior point methods) no longer work for IP problems.



Heuristics

IP problems are NP-complete so heuristic techniques are generally required to solve them.

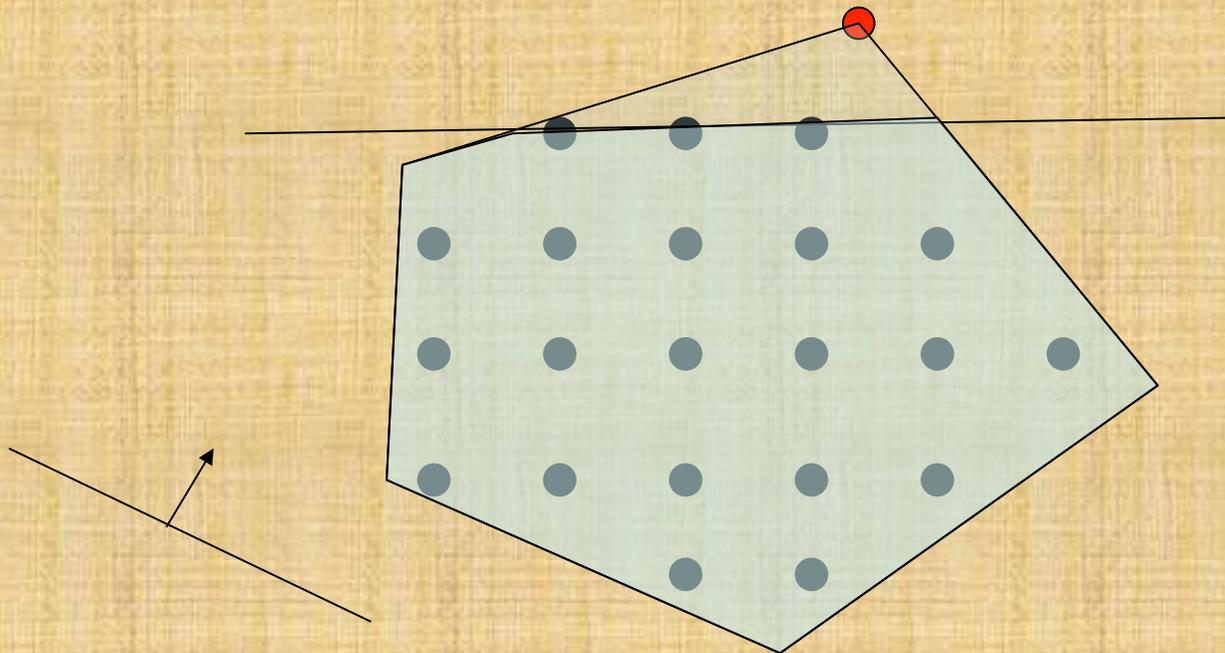
Gomory's Cutting Plane Method: isolate non-integer solutions from the integer solutions with new constraint equations.



Heuristics

IP problems are NP-complete so heuristic techniques are generally required to solve them.

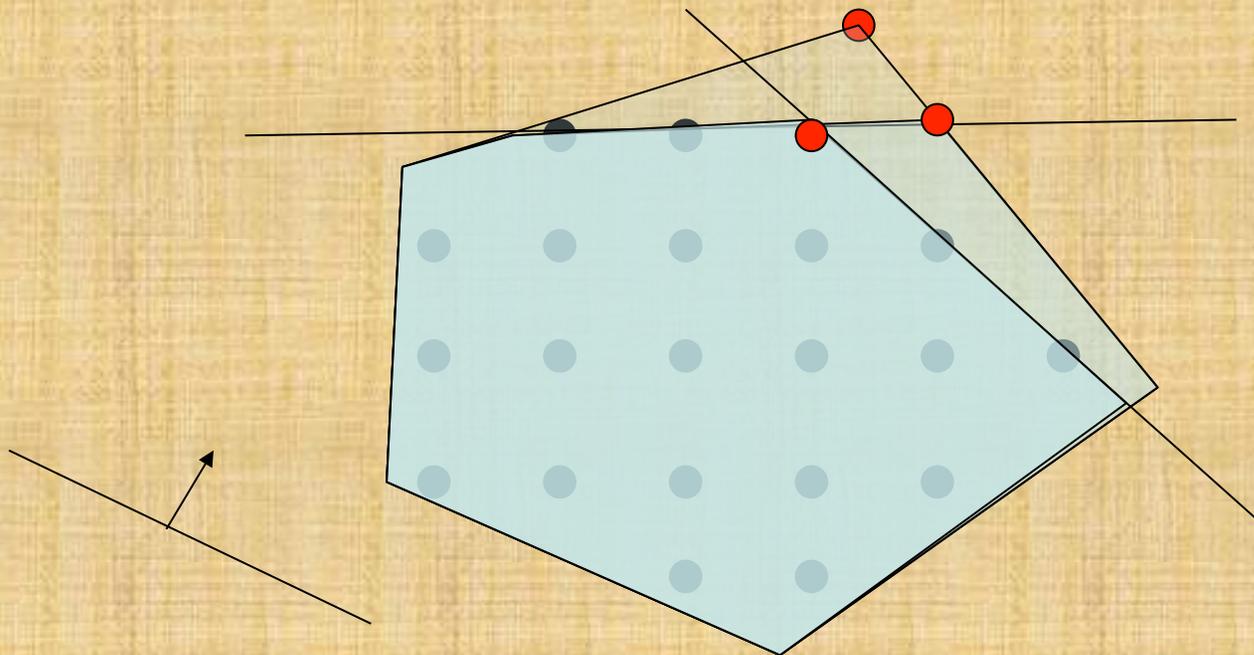
Gomory's Cutting Plane Method: isolate non-integer solutions from the integer solutions with new constraint equations.



Heuristics

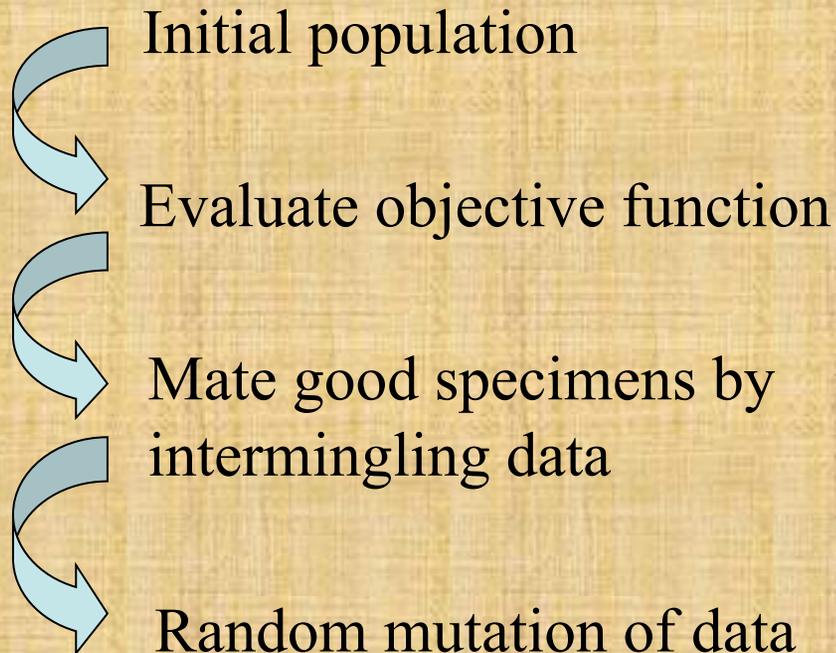
IP problems are NP-complete so heuristic techniques are generally required to solve them.

Gomory's Cutting Plane Method: isolate non-integer solutions from the integer solutions with new constraint equations.



Genetic Algorithms

Genetic Algorithms: consider a population of feasible solutions and, through random mutation and mating of good solutions, produce a new generation of feasible solutions. Because mating tends to preserve good characteristics of the solution, after many generations a very good solution will be found.



Natural selection: remove worst specimens

ENS Zane, MIDN Platt

Reduction to Standard Form

We first note a few reductions that standardize the optimization problem:

Objective Function

Our goal is to **maximize or minimize** the objective function $f(u_1, \dots, u_n) = c_1 u_1 + c_2 u_2 + \dots + c_n u_n$ on the feasible region.

maximizing $f(u_1, \dots, u_n)$ \iff minimizing $-f(u_1, \dots, u_n)$

We may assume that the problem is a **minimization problem**.

Reduction to Standard Form (2)

Replacing Inequality Constraints

(1) Inequalities $\mathbf{a}_1\mathbf{u}_1 + \cdots + \mathbf{a}_n\mathbf{u}_n \geq \mathbf{b}$ can be replaced by
 $-\mathbf{a}_1\mathbf{u}_1 - \cdots - \mathbf{a}_n\mathbf{u}_n \leq -\mathbf{b}.$

(2) Inequalities $\mathbf{a}_1\mathbf{u}_1 + \cdots + \mathbf{a}_n\mathbf{u}_n \leq \mathbf{b}$ can be replaced by
an equality constraint by adding a **slack** variable \mathbf{s}

$$\mathbf{a}_1\mathbf{u}_1 + \cdots + \mathbf{a}_n\mathbf{u}_n + \mathbf{s} = \mathbf{b}$$

$$\mathbf{s} \geq \mathbf{0}.$$

Standard Form

Minimize the objective function

$$f(\mathbf{u}_1, \dots, \mathbf{u}_n) = \mathbf{c}_1 \mathbf{u}_1 + \dots + \mathbf{c}_n \mathbf{u}_n$$

subject to

$$\mathbf{a}_{11} \mathbf{u}_1 + \dots + \mathbf{a}_{1n} \mathbf{u}_n = \mathbf{b}_1$$

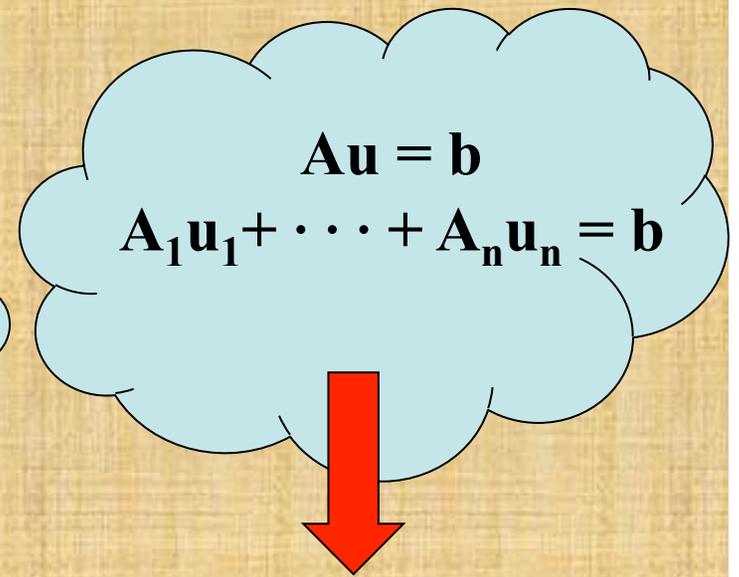
$$\mathbf{a}_{21} \mathbf{u}_1 + \dots + \mathbf{a}_{2n} \mathbf{u}_n = \mathbf{b}_2$$

⋮

$$\mathbf{a}_{m1} \mathbf{u}_1 + \dots + \mathbf{a}_{mn} \mathbf{u}_n = \mathbf{b}_m$$

and

$$\mathbf{u}_1 \geq 0, \dots, \mathbf{u}_n \geq 0 \text{ integers.}$$



$$t^{\mathbf{A}_1 \mathbf{u}_1} t^{\mathbf{A}_2 \mathbf{u}_2} \dots t^{\mathbf{A}_n \mathbf{u}_n} = t^{\mathbf{b}}$$

Standard Form

$$\varphi : \mathbb{k}[x_1, \dots, x_n] \rightarrow \mathbb{k}[t_1^{\pm 1}, \dots, t_m^{\pm 1}]$$

$$x_i \rightarrow t^{A_i} = t_1^{a_{1i}} t_2^{a_{2i}} \cdots t_m^{a_{mi}}$$

$$\begin{aligned} \varphi(x^u) &= \varphi(x_1^{u_1} x_2^{u_2} \cdots x_m^{u_m}) \\ &= t^{A_1 u_1 + \cdots + A_n u_n} \end{aligned}$$

So \mathbf{u} is in the feasible region
if and only if $\varphi(x^u) = t^b$.

Kernel of φ

The map φ contains a lot of information about the IP problem.

How do we find $\text{Ker}(\varphi)$, an ideal we call I_A ?

If all the coefficients in the constraint equations are positive then we can use a Gröbner Basis computation in $k[x_1, \dots, x_n, t_1, \dots, t_m]$.

Theorem: To find $I_A = \text{Ker}(\varphi)$, set

$$J = (x_1 - t^{A_1}, x_2 - t^{A_2}, \dots, x_n - t^{A_n})$$

and compute $J \cap k[x_1, \dots, x_n]$ using a lex order with $t_1 > \dots > t_m > x_1 > \dots > x_n$ to eliminate the t 's. The resulting generating set consists of monomials $x^u - x^v$ with $Au = Av$.

$$(x_1 - t^{A_1}, x_2 - t^{A_2}, \dots, x_n - t^{A_n}) \cap k[x_1, \dots, x_n] = \text{Ker}(\varphi)$$

Take $f(x_1, \dots, x_n) \in \text{LHS}$. Then $f(x) = \sum P(x, t)(x_i - t^{A_i})$ so

$$\varphi(f)(x) = f(\varphi(x)) = \sum P(\varphi(x), t)(\varphi(x_i) - t^{A_i}) = \sum P(\varphi(x), t)(t^{A_i} - t^{A_i}) = 0,$$

that is, $f \in \text{Ker}(\varphi) = \text{RHS}$.

Similarly, if $f \in \text{Ker}(\varphi)$ then $f \in k[x_1, \dots, x_n]$ and

$$\begin{aligned} f(x_1, \dots, x_n) &= f((x_1 - t^{A_1}) + t^{A_1}, \dots, (x_n - t^{A_n}) + t^{A_n}) \\ &\in f(t^{A_1}, \dots, t^{A_n}) + J \\ &= \varphi(f(x)) + J = J. \end{aligned}$$

$$\text{Ker}(\varphi) = (x^u - x^v : Au = Av)$$

$$\text{If } f = \sum_u c_u x^u = \sum_b \sum_{Au=b} c_u x^u \in \text{Ker}(\varphi),$$

$$\text{then } f(t^{A_1}, \dots, t^{A_n}) = \sum_b \sum_{Au=b} c_u t^b = 0.$$

It follows that $\sum_{Au=b} c_u = 0$ for each b .

If $Av = b$ then

$$f = \sum_b \sum_{Au=b} c_u (x^u - x^v) + c_u x^v = \sum_b \sum_{Au=b} c_u (x^u - x^v).$$

Solving the IP

We've seen that $\mathbf{I}_A = \mathbf{Ker}(\varphi) = (\mathbf{x}^u - \mathbf{x}^v : \mathbf{A}\mathbf{u} = \mathbf{A}\mathbf{v})$. It turns out that the Gröbner basis (under any monomial order) for \mathbf{I}_A is also generated by binomials $\mathbf{x}^u - \mathbf{x}^v$ with $\mathbf{A}\mathbf{u} = \mathbf{A}\mathbf{v}$.

Define a monomial order $>_c$ that depends on the cost vector \mathbf{c} :

$$\mathbf{x}^u >_c \mathbf{x}^v \text{ if } \mathbf{c} \cdot \mathbf{u} > \mathbf{c} \cdot \mathbf{v} \text{ or if } \mathbf{c} \cdot \mathbf{u} = \mathbf{c} \cdot \mathbf{v} \text{ and } \mathbf{x}^u >_{\text{lex}} \mathbf{x}^v$$

Let \mathbf{u} be any feasible solution ($\mathbf{A}\mathbf{u}=\mathbf{b}$) to the IP and consider \mathbf{x}^u . Compute the remainder of \mathbf{x}^u divided by the Gröbner basis for \mathbf{I}_A . At each stage in the reduction process we replace \mathbf{x}^u by a monomial \mathbf{x}^v with smaller cost $\mathbf{c} \cdot \mathbf{v}$ and $\mathbf{A}\mathbf{v} = \mathbf{A}\mathbf{u} = \mathbf{b}$. Thus, **the remainder is an optimal solution to the IP.**

$$\mathbf{x}^u \longrightarrow \mathbf{x}^u - (\mathbf{x}^u - \mathbf{x}^v) = \mathbf{x}^v \quad \mathbf{x}^u >_c \mathbf{x}^v \implies \mathbf{c} \cdot \mathbf{u} \geq \mathbf{c} \cdot \mathbf{v}$$

An Example

Minimize $-11u_1 - 15u_2$ subject to

$$4u_1 + 5u_2 + u_3 = 37$$

$$2u_1 + 3u_2 + u_4 = 20$$

$$u_1, u_2, u_3, u_4 \geq 0.$$

Form the ideal

$$I = (x_1 - t_1^4 t_2^2, x_2 - t_1^5 t_2^3, x_3 - t_1, x_4 - t_4)$$

and compute its Gröbner basis:

$$GB = \{t_1 - x_3,$$

$$t_2 - x_4,$$

$$x_4^2 x_3^4 - x_1,$$

$$x_4 x_3^3 x_2 - x_1^2,$$

$$x_4 x_3 x_1 - x_2,$$

$$x_4 x_1^4 - x_3 x_2^3,$$

$$x_3^2 x_2^2 - x_1^3\}$$

The point $(0,0,37,20)$ is feasible but not optimal. We find the remainder of $x_3^{37} x_4^{20}$ under division by the Gröbner basis. This gives $x_1^4 x_2^4 x_3^1$.

The optimal solution to the IP is $(4,4,1,0)$ with value -104 .

Negative Coefficients

Our computations so far required $a_{ij} > 0$.

To deal with negative values, we still try to set

$$\varphi(x_i) = t^{A_i} = t^{A_i}/t^{B_i}$$

and compute the kernel of φ . Instead of considering the ideal

$$(x_1 - t^{A_1}, \dots, x_n - t^{A_n}) \cap k[x_1, \dots, x_n],$$

we look at the ideal

$$(t^{B_1}x_1 - t^{A_1}, \dots, t^{B_n}x_n - t^{A_n}, t_0t_1 \cdots t_m - 1)$$

in $k[x_1, \dots, x_n, t_0, t_1, \dots, t_m]/(t_0t_1 \cdots t_m - 1)$ and intersect it with the subring $k[x_1, \dots, x_n]$.

We get the same result: $\mathbf{I}_A = (\mathbf{x}^u - \mathbf{x}^v : \mathbf{A}\mathbf{u} = \mathbf{A}\mathbf{v})$.

Conti-Traverso Algorithm

To solve: Minimize $\mathbf{c} \cdot \mathbf{u}$ subject to $\mathbf{A}\mathbf{u}=\mathbf{b}$ and $\mathbf{u} \geq \mathbf{0}$.

1. Construct Gröbner basis for the ideal

$$(t^{B_1} x_1 - t^{A_1}, \dots, t^{B_n} x_n - t^{A_n}, t_0 t_1 \cdots t_m - 1)$$

with respect the monomial order that orders the monomials $\mathbf{x}^{\mathbf{u}}$ via $>_c$ and orders any monomial involving the t 's ahead of any monomial only involving the \mathbf{x} 's.

2. Find the remainder $\mathbf{t}_0^{\mathbf{q}} \mathbf{x}^{\mathbf{v}}$ of $t^{\mathbf{b}} = t_0^{\mathbf{p}} t^{b+p(e_1+\cdots+e_m)}$ where $\mathbf{p} = \max\{-\mathbf{b}_i : \mathbf{b}_i < 0\}$.

3. If $\mathbf{q} \neq \mathbf{0}$ then the problem has no feasible solution, otherwise \mathbf{v} is the optimal solution to the IP.

Effectiveness of CT-method

The Conti-Traverso algorithm is implemented in **GRIN**, a software package developed by Serkan Hosten.

The method is competitive with industrial software (**CPLEX**) for dense matrices filled with random entries.

The method is particularly effective when the IP problem needs to be solved for many different values of \mathbf{b} . In this case, the Gröbner basis can be pre-computed and finding the remainder (solving) is extremely fast.