Solution Techniques

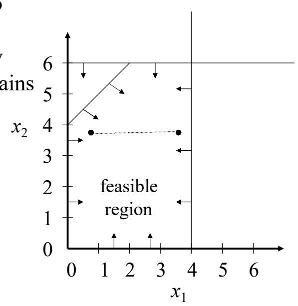
- Good news: many large optimization problems can be solved, even with combinatorial explosion
- Three categories of solution methods:
  - Heuristic methods: quick methods to discount many infeasible solutions
  - Stochastic methods: start with randomly generated solutions, move around "feasibility space" until good enough solution is found (genetic algorithms, simulated annealing)
  - Deterministic methods: sequentially prove that large numbers of solutions are non-optimal, prove mathematically that a final solution is optimal without complete enumeration
- Which is best? Depends on your time, resources and how critical it is to find <u>the</u> optimal solution

CPE 778, Lect. 1

1

Convex Regions

• This feasible region to an LP is <u>convex</u> since a line drawn between any two feasible points remains completely within the region

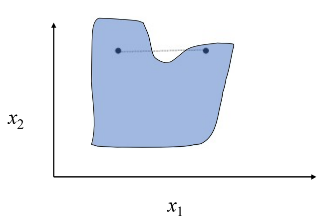


CPE 778, Lect. 4

2

A Nonconvex Region

- If the feasible region looks like



then the constraint set forms a nonconvex region. The problem is clear – even for a simple objective function like $f(x)=x_2$, we see one local and one global maximum

CPE 778, Lect. 4

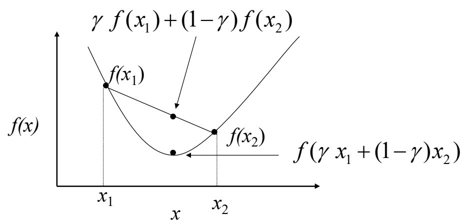
3

A Mathematical Definition of Convexity

- A function $f(x)$ is convex iff:

$$f(\gamma\, x_1 + (1-\gamma)x_2) \le \gamma\, f(x_1) + (1-\gamma) f(x_2)$$

for any two values x_1 and x_2 and for any $0 \le \gamma \le 1$



CPE 778, Lect. 4

4

The Use of Convexity

- Take the following optimization problem:

$$\begin{aligned} &\min && f(\bar{x}) \\ &\text{s.t.} && \bar{g}(\bar{x}) \le 0 \end{aligned}$$

- Regardless of problem type, if all the $\bar{g}(\bar{x})$ are convex, then the feasible region will be a convex set.
- If the objective function $f(x)$ is also convex, then we have a <u>convex programming problem</u>
- This guarantees us that the problem has only one minimum, the global one
- This also works for maxima – just multiply the objective function and constraints by -1

CPE 778, Lect. 4

5

Proving Convexity of Functions

- Using the definition of convexity to prove that a function is convex is difficult
- Here's an easier procedure:
  1. Construct the Hessian matrix $H(x)$
  2. Compute its eigenvalues, check their signs
  3. Refer to the chart to judge convexity
- Let's look at each step via an example

CPE 778, Lect. 4

6

The Hessian Matrix

- It's just the matrix of second partial derivatives:

$$H(\bar{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{bmatrix}$$

Computing the Eigenvalues of the Hessian

Ex: $f(x) = 5x_1^2 + 3x_1x_2 + 2x_2^2 + 7x_1 + 8x_2 + 12$

Now find the Hessian:

$$\frac{\partial f}{\partial x_1} = 10x_1 + 3x_2 + 7 \qquad \frac{\partial f}{\partial x_2} = 3x_1 + 4x_2 + 8$$

$$\frac{\partial^2 f}{\partial x_1^2} = 10 \quad \frac{\partial^2 f}{\partial x_2^2} = 4 \qquad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 3$$

$$H = \begin{bmatrix} 10 & 3 \\ 3 & 4 \end{bmatrix}$$

Computing the Eigenvalues of the Hessian

Ex: $f(x) = 5x_1^2 + 3x_1x_2 + 2x_2^2 + 7x_1 + 8x_2 + 12$

Now find the eigenvalues via det(λI-H)=0 :

$$\lambda I - H = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} 10 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} \lambda\text{-}10 & 3 \\ 3 & \lambda - 4 \end{bmatrix}$$

$$\det[\lambda I - H] = (\lambda - 10)(\lambda - 4) - 9 = \lambda^2 - 14\lambda + 31$$

Using the quadratic formula gives λ= 11.24, 2.76

Eigenvalues and Positive Definiteness

- Here's how we relate the signs of the eigenvalues to convexity:

$f(x)$	$H(x)$	All Eigenvalues
strictly convex	positive definite	>0
convex	positive semidefinite	≥ 0
neither	indefinite	some ≥0, some ≤0
concave	negative semidefinite	≤0
strictly concave	negative definite	<0

CPE 778, Lect. 4

10

Conclusions

- Since λ= 11.24, 2.76 for our example, we know that $H(x)$ is positive definite and that f is a convex function
- Note the positive definiteness has nothing to do with the signs of the elements of a matrix, but only with the signs of the eigenvalues
- Being able to restructure an optimization problem as a convex programming problem can make things much easier!

CPE 778, Lect. 4

11

Elementary Operations

- These three basic operations can be used to find a matrix inverse, scale the elements to help avoid round-off error, or even find new information to help understand an optimization solution.
- The three elementary operations we can perform are:
  - Any row (equation) may be multiplied by a scalar constant
  - The order of the equations and variables may be changed freely
  - Any equation may be replaced by a linear combination of itself and other equations (note we cannot add two equations and replace a third one!)

C&PE 601

12

Gaussian Elimination

- Gaussian Elimination is the process of using the elementary row operations to partially invert, and then solve, a matrix to find unknown values.
- We aim to convert A to be upper triangular (all zeroes below the diagonal). Then, we can find x via backsubstitution
- This is much more efficient that reducing the matrix to diagonal form, or to the identity matrix
- How do we create zeroes below the diagonal?
- We replace each row below the diagonal with a linear combination of rows such that elements are "zeroed out"

C&PE 601

13

Example

- Let's solve the following 3x3 matrix:

$$\begin{bmatrix} 1 & 7 & 5 \\ 4 & 2 & 4 \\ 3 & -6 & -1 \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = \begin{bmatrix} 8 \\ -8 \\ 12 \end{bmatrix}$$

- Start by generating zeroes in the first column, noting we must also operate on the RHS vector:

$$\left[\begin{array}{ccc|c} 1 & 7 & 5 & 8 \\ 4 & 2 & 4 & -8 \\ 3 & -6 & -1 & 12 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} 1 & 7 & 5 & 8 \\ 0 & -26 & -16 & -40 \\ 0 & -27 & -16 & -12 \end{array}\right]$$

- The circled value is called the <u>pivot element</u>

C&PE 601

14

Example, continued

- Now we complete the process by generating a zero at 3,2:

$$\left[\begin{array}{ccc|c} 1 & 7 & 5 & 8 \\ 0 & -26 & -16 & -40 \\ 0 & -27 & -16 & -12 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} 1 & 7 & 5 & 8 \\ 0 & -26 & -16 & -40 \\ 0 & 0 & 0.615 & 29.52 \end{array}\right]$$

- This can now we solved by backsubstitution:

$$T_3 = 29.52 / 0.615 = 48$$

$$T_2 = \frac{1}{-26}\left[-40 + 16T_3\right] = -28$$

$$T_1 = \frac{1}{1}\left[8 - 7T_2 - 5T_3\right] = -36$$

- What do we do if the pivot element is zero, or nearly zero?

C&PE 601

15

Partial Pivoting

- A matrix with a zero on the diagonal still has a solution, in general
- Since we can rearrange the rows and columns freely, we just move things around (called reordering) to avoid having a zero on the diagonal
- In fact, for accuracy reasons, it's often good to have the largest element in any row on the diagonal
- Rearranging the columns of the matrix to get the largest element on the diagonal in each step is called partial pivoting
- This can be expensive for a large matrix, so often we scale the elements in a row and then pick one of the larger ones

C&PE 601

16

Singular Matrices

- Often, when we model a large process, we end up writing more equations than are needed
- Assume we have more equations than variables. What do we do with the extra equations?
- Use them to check our answer. If the result of solving the square system also satisfies the extra equations, then they were redundant. If they are not satisfied, the system was inconsistent, and no solution exists
- What if we have redundant equations in our square system?
- Again, there is no solution, and we call the matrix singular

C&PE 601

17

Determinants

- The determinant of a matrix is a scalar, denoted det(A)
- For a 2 x 2 matrix, it is computed via:

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}$$

- We've all seen determinants before. What is the most efficient way to compute the determinant?
- Depends on the matrix structure. For sparse matrices, use cofactor expansion:

Choose a row or column with many zeroes. For each element in that row or column, multiply the element by the determinant of the minor (or cofactor) to find the determinant of the entire matrix

C&PE 601

18

Determinants via Triangularization

- One more result is helpful here:
  The determinant of a triangular or diagonal matrix is the product of its diagonal elements
- So if we can use the elementary row transformations to give us a triangular form for A, we can get the determinant easily. How do we do this?
- It's just regular Gaussian Elimination
- This is in fact the fastest way to find a determinant, since the steps required for triangularization are $O(N^2)$, and for cofactor expansion are O(N!) .

C&PE 601

19

Eigenvalues and Eigenvectors

- Consider the problem Ax=b as one of mapping a 1 x N vector b into a 1 x N vector x
- There are a set of nonzero vectors which are parallel to Ax, such that

$$A\vec{x} = \lambda\,\vec{x}$$

where $\vec{x}$ is an eigenvector of A,

and λ is an eigenvalue of A

- Eigenvalues and eigenvectors are important in classifying PDEs, data analysis, convergence analysis for linear and nonlinear systems, and in optimization

C&PE 601

20

Computing Eigenvectors and Eigenvalues

- The simplest way to find these is to start with the definition and solve for the λ's:

$$A\vec{x} = \lambda\,\vec{x} \Rightarrow (A - \lambda\, I)\vec{x} = 0$$

$$\text{so}\ \det(A - \lambda\, I) = 0 = p(\lambda)$$

- p(λ) is called the characteristic polynomial of A
- So what kind of problem must we solve to find the eigenvalues?
  A root-finding problem for an Nth-order polynomial
- Once we have the eigenvalues, what type of problem must we solve to find the eigenvectors?
  A set of linear systems. Note we usually have arbitrary constants in eigenvectors, due to the zero vector on the RHS

C&PE 601

21

Eigenvalues and Positive Definiteness

- Eigenvalues allow us to compute positive definiteness, which helps us understand whether an optimal point is local or global. For now, let's just use these definitions:

A	All Eigenvalues
positive definite	>0
positive semidefinite	≥0
indefinite	some ≥0, some ≤0
negative semidefinite	≤0
negative definite	<0

C&PE 601

22

Bracketing Algorithms

- Here's a completely different numerical solution plan:

1. Choose a set of discrete points along the function
2. Evaluate the function at each point
3. Pick the best one!

- This can be slow and tedious, but it guarantees you the determination of all extrema
- Many algorithms are based on this idea: golden section, interval methods, etc.

CPE 778, Lect. 5

23

Evaluating Algorithms

- Now we know which problems can be solved by unconstrained methods
- How do we decide which method is best?
  - See which one finds the best (or a better) answer (reliability or robustness)
  - See how many iterations each one takes to find the same answer, or the optimal solution (convergence)
- The simplest type of convergence is linear:

$$\frac{\left\|\bar{x}^{k+1}-\bar{x}^{*}\right\|}{\left\|\bar{x}^{k}-\bar{x}^{*}\right\|} \le C \quad 0 \le C \le 1 \quad k \text{ "large"}$$

CPE 778, Lect. 5

24

Evaluating Algorithms, cont.

- Usually, algorithms which converge linearly are too slow for industrial-size problems
- We hope to get polynomial convergence:

$$\frac{\left\|\bar{x}^{k+1}-\bar{x}^*\right\|}{\left\|\bar{x}^{k}-\bar{x}^*\right\|^P} \le C \quad 0 \le C \le 1 \quad P \ge 1 \quad k \text{ "large"}$$

- P=2 (quadratic convergence) is usually good enough for modern codes
- Another type of convergence is defined only as $k \to \infty$

$$\lim_{k\to\infty} \frac{\left\|\bar{x}^{k+1}-\bar{x}^*\right\|}{\left\|\bar{x}^{k}-\bar{x}^*\right\|} \to 0 \quad \text{Superlinear convergence}$$

CPE 778, Lect. 5

25

Optimization via Newton’s Method

- Now that we have tools to evaluate algorithms, let’s look at a simple one: Newton’s Method
- Basic idea: solve $\nabla f(\bar{x}) = 0$ numerically
- To find a root near an initial point x^k, write a Taylor series for $f(x^k)$:

$$f(x) \cong f(x^k) + (x - x^k) f'(x^k)$$

- Now set $x=x^*$, the root. Then $f(x)$=0 and

$$-f(x^k) \cong (x^* - x^k) f'(x^k)$$

- Rearrange and create an iteration formula to get

$$x^{k+1} = x^k - f(x^k) \big/ f'(x^k)$$

CPE 778, Lect. 5

26

Optimization via Newton’s Method

- To use this for optimization, just work with the first derivative instead of the original function:

$$x^{k+1} = x^k - f'(x^k) \big/ f''(x^k)$$

- This procedure has advantages and disadvantages
- What is the convergence rate?

CPE 778, Lect. 5

27

Convergence of Newton's Method

- Start with a Taylor series for $f(x)$, keeping the quadratic term:

$$f(x) \cong f(x^k) + (x - x^k) f'(x^k) + \frac{1}{2}(x - x^k)^2 f''(x^k)$$

- Use this to find $f'(x)$, the extreme point. Take the derivative of both sides, set =0

$$f'(x) \cong f'(x^k) + \frac{1}{2}(x - x^k) f''(x^k) \cdot 2 = 0$$

- Rearrange to get

$$x^{k+1} = x^k - f'(x^k) \big/ f''(x^k)$$

- Which means we have quadratic convergence – good!

CPE 778, Lect. 5

28

Disadvantages of Newton's Method

- The method requires first and second derivative information (or approximations to these) which may be tough to get
- If $f''(x^k) \to 0$ at some point, then the size of each step will be large, and convergence will be slow
- The Taylor series approximation of the function may be poor for functions whose slope changes quickly, causing the method to fail.

CPE 778, Lect. 5

29

Approximate Newton Methods

- What if we can't take a derivative, because we don't have a closed form for the objective function?
- Called a <u>black box</u> model

$$\bar{x} \longrightarrow \boxed{\ ?\ } \longrightarrow f(\bar{x})$$

- Examples:
  - variables are temperatures and pressures, function is total operating cost from simulator
  - variables are locations of oil wells, function is production rate from simulator

CPE 778, Lect. 6

30

Approximate Newton Methods, cont.

- Now we can't take any derivatives, but we can evaluate the function at any point
- So approximate $f'(x^k) / f''(x^k)$
- Use a <u>finite difference</u> approach
- One way to do that is

$$x^{k+1} = x^k - \frac{\left[f(x^k+h) - f(x^k-h)\right] / 2h}{\left[f(x^k+h) - 2f(x^k) + f(x^k-h)\right] / h^2}$$

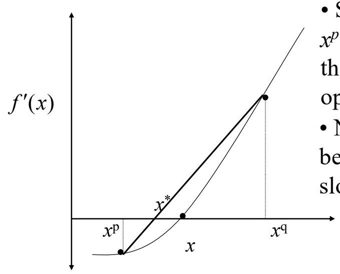
- Just choose a step size h, and evaluate the function around the current x value

CPE 778, Lect. 6

31

Quasi-Newton Methods (Secant method)

- Remember, we want to solve $f'(x) = 0$



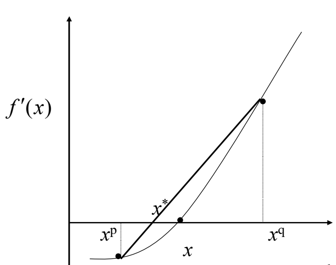
- Start with two points, x^p and x^q, for which the derivative is opposite in sign
- Now, draw a line between the points, with slope

$$m = \frac{f'(x^q) - f'(x^p)}{x^q - x^p}$$

CPE 778, Lect. 6

32

Quasi-Newton Methods (Secant method)



- Put the slope of that line in the iteration formula:

$$x^{k+1} = x^k - f'(x^k) / m$$

- Set x^q as our initial point, substitute in for m to get:

$$x^{k+1} = x^k - \frac{f'(x^k)}{\left[f'(x^k) - f'(x^p)\right] / (x^k - x^p)}$$

CPE 778, Lect. 6

33

Search Directions

- Idea: look for a series of points for which $f(\bar{x}^{k+1}) < f(\bar{x}^k)$
- This will eventually find a local minimum, or get stuck in a saddle point
- So what direction do we have to move in to get from $\bar{x}^k$ to $\bar{x}^{k+1}$?

CPE 778, Lect. 7

34

Search Directions, cont.

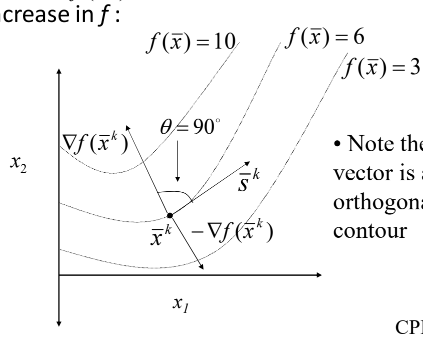
- The key is to show that any direction $\bar{s}$ which satisfies $\nabla^T f(\bar{x})\bar{s} < 0$ will lead to an improvement in $\bar{x}$
- Note this is vector-vector multiplication, between the gradient of the function and the search direction

CPE 778, Lect. 7

35

A 2-D Example

- Note $\nabla f(\bar{x}^k)$ is the direction of the greatest increase in f :



- Note the gradient vector is always orthogonal to the contour

CPE 778, Lect. 7

36

Finding the search direction

- How do we get $\nabla^T f(\bar{x}^k)\bar{s} < 0$?
- Use vector calculus to show

$$\nabla^T f(\bar{x}^k)\bar{s} = \left|\nabla f(\bar{x}^k)\right|\left|\bar{s}^k\right|\cos\theta$$

- The magnitudes are positive, so the sign is defined by the cosine term
- In our example, θ=90°, so cos θ=0. This means the search direction is parallel to the contour, and no (local) change in *f* occurs

CPE 778, Lect. 7

37

Finding the search direction, cont.

- If $\theta < 90^\circ$ then $\cos\theta > 0$ so $\nabla^T f(\bar{x}^k)\bar{s} > 0$

That would be a maximization direction

- If $\theta > 90^\circ$ then $\cos\theta < 0$ so $\nabla^T f(\bar{x}^k)\bar{s} < 0$

That would be a minimization direction

- Review the picture to get a feel for search directions...

CPE 778, Lect. 7

38

Two Major Issues

- Now, we know that θ should be > 90°
- But which $\bar{s}^k$ will get us to the minimum the fastest?
- And how far should we go in the $\bar{s}^k$ direction before recomputing $\nabla f(\bar{x}^{k+1})$ and getting a new $\bar{s}^{k+1}$?

CPE 778, Lect. 7

39

Steepest Descent: details

- We just stated that $-\nabla f(\bar{x}^k)$ is the direction of fastest decrease in *f*
- So use that direction as $\bar{s}^k$
- This satisfies our constraint on θ:

$$\nabla^T f(\bar{x}^k)\bar{s}^k = \nabla^T f(\bar{x}^k)\left(-\nabla f(\bar{x}^k)\right)$$
$$= \left|\nabla f(\bar{x}^k)\right|\left|-\nabla f(\bar{x}^k)\right|\cos 180^\circ$$
$$= \left(\left|\nabla f(\bar{x}^k)\right|\right)^2(-1) < 0$$

CPE 778, Lect. 7

40

Steepest Descent: algorithm

- We need a way to get from one point to the next, called an <u>iteration formula</u>
- For Steepest Descent, we get

$$\bar{x}^{k+1} = \bar{x}^k + \Delta\bar{x}^k = \bar{x}^k + \alpha^k \bar{s}^k$$
$$= \bar{x}^k - \alpha^k \nabla f(\bar{x}^k)$$

- α^k is a scalar defining how far in the search direction to move

CPE 778, Lect. 7

41

Ideas – Conjugate Gradient Method

- Goal: minimize $f(\bar{x})$
  - use previous search direction to help find new search direction
  - allows optimal line search to be used without 90° turn at each iteration
  - Increases computation time for each iteration, but requires fewer iterations
  - Unconstrained problems only

CPE 778, Lect. 8

42

Conjugate Search Directions

- Instead of finding search directions which are orthogonal, find a new search direction which is conjugate to the previous one.
- Two directions are conjugate if:

$$\left(\bar{s}^{i}\right)^{T} H\left(\bar{s}^{j}\right)=0$$

where H is the Hessian Matrix

CPE 778, Lect. 8

43

Conjugate Gradient: Algorithm

- Here are the steps:

1. Start with Steepest Descent: pick an $\bar{x}^0$ and set $\bar{s}^0 = -\nabla f(\bar{x}^0)$
2. Compute $\bar{x}^{k+1} = \bar{x}^k + \alpha^k \bar{s}^k$ by choosing α which minimizes f in the direction $\bar{s}^k$
3. Compute $f(\bar{x}^{k+1}), \nabla f(\bar{x}^{k+1})$

continued...

CPE 778, Lect. 8

44

Conjugate Gradient: Algorithm

4. Find $\bar{s}^{k+1}$ using

$$\bar{s}^{k+1} = -\nabla f(\bar{x}^{k+1}) + \bar{s}^k \frac{\nabla^T f(\bar{x}^{k+1}) \nabla f(\bar{x}^{k+1})}{\nabla^T f(\bar{x}^k) \nabla f(\bar{x}^k)}$$

5. Test convergence:

Is $\left\|\nabla f(\bar{x}^k)\right\| < \varepsilon$?

If yes, stop. If no, iterate from step 2

CPE 778, Lect. 8

45

Ideas – Conjugate Gradient Method

- Goal: minimize $f(\bar{x})$
  - use previous search direction to help find new search direction
  - allows optimal line search to be used without 90° turn at each iteration
  - Increases computation time for each iteration, but requires fewer iterations
  - Unconstrained problems only

CPE 778, Lect. 8

46

Conjugate Search Directions

- Instead of finding search directions which are orthogonal, find a new search direction which is <u>conjugate</u> to the previous one.
- Two directions are conjugate if:

$$\left(\bar{s}^{i}\right)^{T} H\left(\bar{s}^{j}\right) = 0$$

where H is the Hessian Matrix

CPE 778, Lect. 8

47

Conjugate Gradient: Algorithm

- Here are the steps:

1. Start with Steepest Descent: pick an $\bar{x}^0$ and set $\bar{s}^0 = -\nabla f(\bar{x}^0)$
2. Compute $\bar{x}^{k+1} = \bar{x}^k + \alpha^k \bar{s}^k$ by choosing α which minimizes f in the direction $\bar{s}^k$
3. Compute $f(\bar{x}^{k+1}), \nabla f(\bar{x}^{k+1})$

continued...

CPE 778, Lect. 8

48

Conjugate Gradient: Algorithm

4. Find $\bar{s}^{k+1}$ using

$$\bar{s}^{k+1} = -\nabla f(\bar{x}^{k+1}) + \bar{s}^k \frac{\nabla^T f(\bar{x}^{k+1})\nabla f(\bar{x}^{k+1})}{\nabla^T f(\bar{x}^k)\nabla f(\bar{x}^k)}$$

5. Test convergence:

$$\text{Is } \left\|\nabla f(\bar{x}^k)\right\| < \varepsilon \,?$$

If yes, stop. If no, iterate from step 2

CPE 778, Lect. 8

49

Ideas – Newton's Method

- Goal: minimize $f(\bar{x})$
  - Neither steepest descent nor conjugate gradient uses curvature (2nd derivative) information
  - If you have such information, why not use it?
  - Newton's method does not use previous search direction information, however

CPE 778, Lect. 8

50

Newton's Method: Derivation

- As before, start with a Taylor series truncated after the quadratic term:

$$f(\bar{x}) \cong f(\bar{x}^k) + \nabla^T f(\bar{x}^k)\Delta\bar{x}^k + \frac{1}{2}(\Delta\bar{x}^k)^T H(\bar{x}^k)(\Delta\bar{x}^k)$$

- Take the gradient, set it equal to zero:

$$\nabla f(\bar{x}) = \nabla f(\bar{x}^k) + H(\bar{x}^k)(\Delta\bar{x}^k) = 0$$

- Solve for the change in x:

$$\bar{x}^{k+1} - \bar{x}^k = \Delta\bar{x}^k = -\left[H(\bar{x}^k)\right]^{-1}\nabla f(\bar{x}^k)$$

CPE 778, Lect. 8

51

Newton's Method: Details

$$\bar{x}^{k+1} - \bar{x}^{k} = \Delta\bar{x}^{k} = \left[H(\bar{x}^{k})\right]^{-1}\nabla f(\bar{x}^{k})$$

- Note this is (*n* x *n)* x *(n* x *1*) = (*n* x 1), so we get all the *x* values by matrix-vector multiplication

- Convergence is greatly improved by including a step length:

$$\bar{x}^{k+1} - \bar{x}^{k} = -\alpha^{k}\left[H(\bar{x}^{k})\right]^{-1}\nabla f(\bar{x}^{k})$$

- Note this is just like steepest descent, with

$$\bar{s}^{k} = -\left[H(\bar{x}^{k})\right]^{-1}\nabla f(\bar{x}^{k})$$

CPE 778, Lect. 8

52

Newton's Method: More Details

$$\bar{x}^{k+1} - \bar{x}^{k} = \Delta\bar{x}^{k} = -\alpha^{k}\left[H(\bar{x}^{k})\right]^{-1}\nabla f(\bar{x}^{k})$$

- Using this directly requires the inversion of the Hessian Matrix, which is computationally difficult
- An easier way is to rearrange the equation:

$$H(\bar{x}^{k})\Delta\bar{x}^{k} = -\alpha^{k}\nabla f(\bar{x}^{k})$$

- What kind of a problem is this?
- A set of linear equations to solve. This can be done with much less round-off error than matrix inversion

CPE 778, Lect. 8

53

Standard Form of an LP

- Reformulating problems in a common form allows one solution algorithm to work for many problems
- The <u>Standard Form</u> of an LP is:

$$\begin{aligned}
\max \quad & z = c_1x_1 + c_2x_2 + \ldots + c_nx_n \\
\text{s.t.} \quad & a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1 \\
& a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2 \\
& \ldots \\
& a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le b_m \\
& x_j \ge 0\,,\ j = 1,\ldots,n
\end{aligned}$$

- Note we need maximization, all ≤ constraints, and nonnegative variables

CPE 778, Lect. 9

54

Canonical Form of an LP

- Another commonly used form in linear programming
- The Canonical Form of an LP is:

$$
\begin{aligned}
\text{max} \quad & z = c_1x_1 + c_2x_2 + \ldots + c_nx_n \\
\text{s.t.} \quad & a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\
& a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\
& \ldots \\
& a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \\
& x_j \geq 0 \,,\; j = 1, \ldots, n
\end{aligned}
$$

- Note we need maximization, all equality constraints, and nonnegative variables

CPE 778, Lect. 9

55

Converting an Equality to an Inequality

- We can replace

$$\sum_{j=1}^{n} a_{ij}x_j = b_j \quad \text{with}$$

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_j \;\text{ and }\; \sum_{j=1}^{n} a_{ij}x_j \geq b_j \;\text{ or }\; -\sum_{j=1}^{n} a_{ij}x_j \leq -b_j$$

This replaces the equality with two inequalities

CPE 778, Lect. 9

56

Converting an Inequality to an Equality

- To do this, we need to invent a new variable to make up the difference between the inequality and the equality, called a slack variable
- So replace

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_j \quad \text{with}$$

$$\sum_{j=1}^{n} a_{ij}x_j + s_j = b_j \quad \text{and } s_j \geq 0$$

- s_j is the new slack variable

CPE 778, Lect. 9

57

Matrix Notation

- Now that we can convert between forms, let's write the problem generally in matrix notation, with

$$\overline{\overline{A}} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & & & \\ a_{m1} & & & a_{mn} \end{bmatrix} \qquad \overline{x} = \begin{bmatrix} x_1 \\ \dots \\ x_n \end{bmatrix}$$

$$\overline{b} = \begin{bmatrix} b_1 \\ \dots \\ b_m \end{bmatrix} \qquad \overline{c} = \begin{bmatrix} c_1 \\ \dots \\ c_n \end{bmatrix}$$

CPE 778, Lect. 9

58

Standard Forms in Matrix Notation

- Using that notation, we can write the canonical form as:

$$\begin{aligned} \max \quad & z = \overline{c}^T \overline{x} \\ \text{s.t.} \quad & \overline{\overline{A}}\overline{x} = \overline{b} \\ & \overline{x} \geq 0 \end{aligned}$$

- The standard form becomes:

$$\begin{aligned} \max \quad & z = \overline{c}^T \overline{x} \\ \text{s.t.} \quad & \overline{\overline{A}}\overline{x} \leq \overline{b} \\ & \overline{x} \geq 0 \end{aligned}$$

- How do we find a feasible solution?

Solve $\overline{\overline{A}}\overline{x} = \overline{b}$!

CPE 778, Lect. 9

59

Three Types of Feasible Regions

- We can write the constraint set as $\overline{\overline{A}}\overline{x} = \overline{b}$ where $\overline{\overline{A}}$ is an n x m matrix
- We get three cases:
  1. $n < m$
     - Overspecified – no solution (null feasible region)
  2. $n = m$
     - Exactly specified – one solution, which is optimal (if all $x \geq 0$) for any objective function!

CPE 778, Lect. 10

60

Case Three

- The important case is $n > m$
  - An infinite number of solutions exist
  - Usually, some are feasible with $x \geq 0$
- An important subset of these solutions are the basic solutions:

A basic solution to an LP with $n > m$ is a solution obtained by setting $n - m$ of the variables to zero, and solving the reduced system to find the other variables

- Note not all basic solutions are feasible, since some variable values may be negative

CPE 778, Lect. 10

61

Major Result #1

- Basic feasible solutions are important because of this result:

  If an LP has an optimal solution, it must belong to the set of basic feasible solutions.

- The proof is based on these ideas:
  - It can be shown that all basic feasible solutions are extreme points of the feasible region
  - All LP's are convex, so the optimal solution for any objective function must lie at an extreme point

CPE 778, Lect. 10

62

Major Result #2

- As previously stated, the simplex method is an efficient way to solve LP's because of the following result:

  The simplex method always moves from a given basic feasible solution to a second basic feasible solution with a larger objective value

- Thus a sequence of extreme points is evaluated, each better than the last
- In practice, this requires orders of magnitude fewer function evaluations to execute than if we look at all basic feasible solutions

CPE 778, Lect. 11

63

Challenges

- There are two major difficulties with the simplex method:
  1. You need a method to obtain a basic feasible solution to begin with
  2. For very large problems (n>10,000) this may still look at too many solutions to be practical
- We'll deal with the first challenge now, and discuss the second one later

CPE 778, Lect. 11

64

Finding an Initial B.F.S.

- Remember, our problem in canonical form is:

$$\begin{aligned} \max \quad & z = \bar{c}^T \bar{x} \\ \text{s.t.} \quad & \bar{\bar{A}}\bar{x} = \bar{b} \\ & \bar{x} \geq 0 \end{aligned}$$

- Assuming all $b_i \geq 0$, the following procedure is used
  1. Make sure your variables are numbered such that the slack variables are at the end
  2. Set the <u>first</u> n-m variables equal to zero. Often these are the non-slack variables.

CPE 778, Lect. 11

65

Finding an Initial B.F.S., cont.

$$\begin{aligned} \max \quad & z = \bar{c}^T \bar{x} \\ \text{s.t.} \quad & \bar{\bar{A}}\bar{x} = \bar{b} \\ & \bar{x} \geq 0 \end{aligned}$$

3. Solve the m x m linear system to find the values of the basic variables. If $b_i \geq 0$, they will all be positive, so the basic solution will be feasible

- What if we don't have $b_i \geq 0$? Then we add <u>artificial variables</u> to the problem, and we still can find an initial basic feasible solution
- This is tricky, but always possible

CPE 778, Lect. 11

66

Simplex Tableaus

- A tableau is just a tabular form of the matrix equations we saw before
- Let's set one up for our example from last time:

$$\begin{aligned} \max \quad & z = 120x_1 + 100x_2 \\ \text{s.t.} \quad & 2x_1 + 2x_2 + x_3 = 8 \\ & 5x_1 + 3x_2 + x_4 = 15 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

- Remember $n - m = 2$, so we set the first two variables to zero, and get [0,0,8,15] as our initial basic feasible solution

CPE 778, Lect. 11

67

The Initial Tableau

- Now take each constraint and the objective function, and put them in a tableau:

variables	x_1	x_2	x_3	x_4	z	constants
constraint 1	2	2	1	0	0	8
constraint 2	5	3	0	1	0	15
objective row	-120	-100	0	0	1	0

- Note we rewrite the objective function as

$$-120x_1 - 100x_2 + z = 0$$

- All constraints are entered in the same manner

CPE 778, Lect. 11

68

Reading a Tableau

variables	x_1	x_2	x_3	x_4	z	constants
constraint 1	2	2	1	0	0	8
constraint 2	5	3	0	1	0	15
objective row	-120	-100	0	0	1	0

- The basic variables (x_3 and x_4) are found in columns with a unit vector, and their values are given in the constant column
- What are the values of the non-basic variables?
  Zero
- The value of the objective function at this point is also listed in the constant column

CPE 778, Lect. 11

69

A Key Question

variables	x_1	x_2	x_3	x_4	z	constants
constraint 1	2	2	1	0	0	8
constraint 2	5	3	0	1	0	15
objective row	-120	-100	0	0	1	0

- Are we at the optimal solution?
- Here's a test: If the values in the objective row of the tableau are zero for all basic variables and positive for all non-basic variables, then we are at the optimum

Why? Look at $-120x_1 - 100x_2 + 0x_3 + 0x_4 + z = 0$

- If we add x_1 or x_2 to the basis, they will increase, and thus z will increase

CPE 778, Lect. 11

70

Moving to the Next B.F.S

variables	x_1	x_2	x_3	x_4	z	constants
constraint 1	2	2	1	0	0	8
constraint 2	5	3	0	1	0	15
objective row	-120	-100	0	0	1	0

- So we have a basic feasible solution, but it's not optimal
- How can we get a better B.F.S.?
- Choose a new variable to become basic, and one to become non-basic
- Which one should "enter" the basis (become nonzero)?
- The one which improves the objective the most, which is the one with the most negative entry in the objective row: x_1

CPE 778, Lect. 11

71

Moving to the Next B.F.S, cont.

variables	x_1	x_2	x_3	x_4	z	constants
constraint 1	2	2	1	0	0	8
constraint 2	5	3	0	1	0	15
objective row	-120	-100	0	0	1	0

- Which variable should leave the basis (become zero)?
- Find one which will guarantee that the new value of the entering variable will be positive, for feasibility
- Done by computing the <u>θ-ratios</u>
- <u>θ-ratios</u> are ratios of the constant column entries over the entering (called pivotal) column entries
- For our example $\theta_3 = \frac{8}{2} = 4$ and $\theta_4 = \frac{15}{5} = 3$

CPE 778, Lect. 11

72

Moving to the Next B.F.S, cont.

$$\theta_3 = \frac{8}{2} = 4 \text{ and } \theta_4 = \frac{15}{5} = 3$$

- Subscripts represent the leaving variable
- Choose the leaving variable with the minimum θ-ratio
- Not doing so often leads to an infeasible basic solution:

Choosing θ_4=3 makes x_1 the entering variable, and x_4 the leaving variable. We looked at this basic solution – it is [3,0,2,0]

Choosing θ_3=4 makes x_1 the entering variable, and x_3 the leaving variable. We also looked at this basic solution – it is [4,0,0,-5], which is infeasible

Moving to the Next B.F.S, cont.

variables	x_1	x_2	x_3	x_4	z	constants
constraint 1	2	2	1	0	0	8
constraint 2	5	3	0	1	0	15
objective row	-120	-100	0	0	1	0

- Now, since x_1 is our entering variable and x_4 is our leaving variable, we want to convert the tableau from basic variables x_3 and x_4 to basic variables x_3 and x_1
- This means we need to make a unit vector in the entering variable's column. How do we do that?
- By pivoting (row reduction)

Pivoting Rules

variables	x_1	x_2	x_3	x_4	z	constants
constraint 1	2	2	1	0	0	8
constraint 2	5	3	0	1	0	15
objective row	-120	-100	0	0	1	0

1. Find the pivot, which is the value at the intersection of the pivotal (entering) column and the pivotal (departing) row
2. Multiply the pivotal row by $1/k$, where k is the pivot element, to get a 1 at the pivot point
3. Add multiples of the new pivot row to the other rows to create zeroes in the pivot column

Pivoting: Division Step

variables	x_1	x_2	x_3	x_4	z	constants
x_3	2	2	1	0	0	8
x_4	5	3	0	1	0	15
objective row	-120	-100	0	0	1	0

⬇

variables	x_1	x_2	x_3	x_4	z	constants
x_3	2	2	1	0	0	8
x_1	1	3/5	0	1/5	0	3
objective row	-120	-100	0	0	1	0

CPE 778, Lect. 11

76

Pivoting: Multiplication Step

variables	x_1	x_2	x_3	x_4	z	constants
x_3	2	2	1	0	0	8
x_1	1	3/5	0	1/5	0	3
objective row	-120	-100	0	0	1	0

⬇

variables	x_1	x_2	x_3	x_4	z	constants
x_3	0	4/5	1	-2/5	0	2
x_1	1	3/5	0	1/5	0	3
objective row	0	-28	0	24	1	360

CPE 778, Lect. 11

77

The New Tableau

variables	x_1	x_2	x_3	x_4	z	constants
x_3	0	4/5	1	-2/5	0	2
x_1	1	3/5	0	1/5	0	3
objective row	0	-28	0	24	1	360

- This corresponds to the basic feasible solution [3,0,2,0] that we found before
- Is it optimal?
  No, negative value in objective row
- But it is certainly better!

CPE 778, Lect. 11

78

A Second Iteration

variables	x_1	x_2	x_3	x_4	z	constants
x_3	0	4/5	1	-2/5	0	2
x_1	1	3/5	0	1/5	0	3
objective row	0	-28	0	24	1	360

- Choose entering variable: x_2
- Choose leaving variable:

$$\theta_3 = \frac{2}{4/5} = 2.5 \text{ and } \theta_1 = \frac{3}{3/5} = 5 \text{ so } x_3 \text{ is leaving}$$

- Find pivot element

CPE 778, Lect. 11

79

Second Iteration, continued

variables	x_1	x_2	x_3	x_4	z	constants
	0	1	5/4	-2/4	0	10/4
	1	3/5	0	1/5	0	3
objective row	0	-28	0	24	1	360

⬇

variables	x_1	x_2	x_3	x_4	z	constants
x_2	0	1	5/4	-2/4	0	10/4
x_1	1	0	-3/4	1/2	0	6/4
objective row	0	0	35	10	1	430

CPE 778, Lect. 11

80

Final Tableau

variables	x_1	x_2	x_3	x_4	z	constants
x_2	0	1	5/4	-2/4	0	10/4
x_1	1	0	-3/4	1/2	0	6/4
objective row	0	0	35	10	1	430

- Now we are optimal, since all entries in the objective row are positive
- Solution is [3/2 , 5/2, 0, 0] as obtained last time
- Next time we'll discuss how this can go wrong...

CPE 778, Lect. 11

81

Degenerate Solutions

- What do we do if two θ-ratios are equal?
- Just choose one of the corresponding variables to leave the basis
- This can cause a zero value to occur for a basic variable, which leads to two basic feasible solutions with the same objective value, called degenerate solutions
- The simplex method will then cycle between the two solutions, and not reach the optimum

CPE 778, Lect. 12

82

Perturbation

- So what do we do about this problem?
- If we slightly change the values of the b_i constants, we'll get slightly differing θ-ratios
- The two objective values will now be slightly different, and no cycling will occur
- The error introduced by this technique is no more than normal computer round-off error (if the perturbation is small enough)

CPE 778, Lect. 12

83

Duality

- The dual of an optimization problem is a closely related problem which can be used to help solve the original problem
- The concept is used in many "decomposition" strategies for solving MILP's
  - Lagrangian Relaxation
  - Bender's Decomposition

CPE 778, Lect. 12

84

The Dual Problem

- Our original problem in standard form is: (called the primal problem)

$$\max \quad z = \bar{c}^T \bar{x}$$
$$\text{s.t.} \quad \bar{\bar{A}}\bar{x} \le \bar{b}$$
$$\bar{x} \ge 0$$

- If we rearrange this, we get the dual problem:

$$\min \quad z = \bar{b}^T \bar{w}$$
$$\text{s.t.} \quad \left(\bar{\bar{A}}\right)^T \bar{w} \ge \bar{c}$$
$$\bar{w} \ge 0$$

- Note the constants in the objective are now on the RHS of the constraints, the RHS vector is now in the objective, and max is now min

CPE 778, Lect. 12

85

The Dual Problem, continued

$$\min \quad z = \bar{b}^T \bar{w}$$
$$\text{s.t.} \quad \left(\bar{\bar{A}}\right)^T \bar{w} \ge \bar{c}$$
$$\bar{w} \ge 0$$

- The new variables (w_i) are called the marginal values of each input x
- They represent how much improvement to the objective function could be made per unit x_i
- For example, they answer questions like: if we increase our pipeline diameter such that we can ship more product, how much profit will be gained by each unit of each product?
- You can view these in GAMS by printing variable.m

CPE 778, Lect. 12

86

Converting from Canonical Form

- The primal in canonical form is:

$$\max \quad z = \bar{c}^T \bar{x}$$
$$\text{s.t.} \quad \bar{\bar{A}}\bar{x} = \bar{b}$$
$$\bar{x} \ge 0$$

- When we have equality constraints, the marginal costs may be negative

- The dual is:

$$\min \quad z = \bar{b}^T \bar{w}$$
$$\text{s.t.} \quad \left(\bar{\bar{A}}\right)^T \bar{w} \ge \bar{c}$$
$$\bar{w} \text{ are unrestricted}$$

CPE 778, Lect. 12

87

Conversion Table

- Each part of the primal LP has a corresponding part in the dual:

Primal	Dual
maximization	minimization
Coefficients of Obj.	RHS of constraints
Coeff. of *i*th constraint	Coeff. of *i*th variable
*i*th constraint is ≤ (max)	*i*th variable is ≥0
*i*th constraint is ≥ (min)	*i*th variable is ≥0
*i*th constraint is =	*i*th variable is unrestricted
No. of variables	No. of constraints

CPE 778, Lect. 12

88

Useful Results about the Dual

- If the dual problem has an unbounded optimal solution, then the primal has no feasible solutions
- Conversely, since the dual of the dual is the primal, if the dual has no feasible solutions, then the primal has an unbounded optimal solution
- The logic is that if the marginal costs are undefined, then we cannot have a feasible solution

CPE 778, Lect. 12

89

The Duality Theorem

- If either the primal or the dual has a feasible solution with a finite optimal objective value, then the other problem has a feasible, optimal solution with the same objective value
- This means

$$z^* = \overline{c}^T \overline{x}^* = \overline{b}^T \overline{w}^* = z'^*$$

- So we can solve either the dual or the primal problem, get the objective value, and back-calculate the other variables

CPE 778, Lect. 12

90

Complimentary Slackness

- The Theorem of Complimentary Slackness states:
  If a constraint in the primal is not satisfied with equality, then the corresponding variable in the dual must be zero at optimality.
- In our example, do we have an inactive constrant at optimality?
- Yes, at the optimum point (6,0), constraint one is inactive

$$\begin{aligned} \max \quad & z = 2x_1 + x_2 \\ \text{s.t.} \quad & x_1 + 2x_2 \le 8 \\ & 3x_1 + 4x_2 \le 18 \\ & x_1, x_2 \ge 0 \end{aligned}$$

CPE 778, Lect. 13

91

Complimentary Slackness, cont.

- This means w_1 should equal zero at the optimum of the dual, and it is: (0, 2/3) is the optimum.
- Does this work in reverse? Do we have any inactive constraints in the dual at optimality?
- Yes, constraint 2 is inactive, so x_2 should be zero at optimality, and it is.

$$\begin{aligned} \min \quad & z' = 8w_1 + 18w_2 \\ \text{s.t.} \quad & w_1 + 3w_2 \ge 2 \\ & 2w_1 + 4w_2 \ge 1 \\ & w_1, w_2 \ge 0 \end{aligned}$$

CPE 778, Lect. 13

92

Complimentary Slackness, cont.

$$\begin{aligned} \max \quad & z = 2x_1 + x_2 \\ \text{s.t.} \quad & x_1 + 2x_2 \le 8 \\ & 3x_1 + 4x_2 \le 18 \\ & x_1, x_2 \ge 0 \end{aligned}$$

- Think in terms of constraints as resource limitations
- If we increase the bound on the first constraint (say increase pipeline diameter), we can't change the objective, since we are limited by the second constraint – thus $w_1 = 0$
- Another way to state this is: the marginal value associated with an inactive constraint is always zero at optimality

CPE 778, Lect. 13

93

Finding the Dual Variables

- We can use the simplex tableau to find the values of the dual variables
- Let's do this for our previous example:

$$\begin{aligned} \text{max} \quad & z = 120x_1 + 100x_2 \\ \text{s.t.} \quad & 2x_1 + 2x_2 + x_3 = 8 \\ & 5x_1 + 3x_2 + x_4 = 15 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

- Remember x_3 and x_4 are slack variables, and the optimal solution was [3/2 , 5/2, 0, 0]
- Write down the dual for this problem

CPE 778, Lect. 13

94

Finding the Dual Variables, cont.

- Here's the initial tableau, for the basic feasible solution [0,0,8,15]:

variables	x_1	x_2	x_3	x_4	z	constants
constraint 1	2	2	1	0	0	8
constraint 2	5	3	0	1	0	15
objective row	-120	-100	0	0	1	0

- The slack variables are basic here, and the columns corresponding to the nonbasic variables are combined to form the B matrix
- The dual variables are related to B via $\overline{w}^* = \overline{c}\,\overline{\overline{B}}^{-1}$

CPE 778, Lect. 13

95

Information from the Final Tableau

variables	x_1	x_2	x_3	x_4	z	constants
x_2	0	1	5/4	-2/4	0	10/4
x_1	1	0	-3/4	1/2	0	6/4
objective row	0	0	35	10	1	430

- Now we are at the optimal solution to the primal
- Solution is [3/2 , 5/2, 0, 0] as obtained last time
- The process of row reduction is equivalent to inverting the B matrix, therefore...
- The optimal w variables are listed in the final tableau, under the initial basic variables!

CPE 778, Lect. 13

96

Solution to the Dual

- The dual is:

$$\begin{aligned} \min \quad & z' = 8w_1 + 15w_2 \\ \text{s.t.} \quad & 2w_1 + 5w_2 \ge 120 \\ & 2w_1 + 3w_2 \ge 100 \\ & w_1 \ge 0 \\ & w_2 \ge 0 \\ & w_1, w_2 \text{ unrestricted} \end{aligned}$$

- The solution from the final tableau was

$$\overline{w}^* = [35,10]$$

- Note $z' = 8*35 + 15*10 = 430 = z$

CPE 778, Lect. 13

97

Conclusions

- Complementary Slackness tells us that if a constraint is inactive at the optimal solution, then the marginal value of that constraint must be zero. Increasing the resource limit won't help.
- We can get the values for the marginal costs (the dual variables) by noting that the process involved with the Simplex method is identical to inverting the *B* matrix. The values of those dual variables are listed under the initial basic variables, in the objective function row.

CPE 778, Lect. 13

98

Nonlinear Programming

- Abbreviated NLP
  - nonlinear constraints and/or a nonlinear objective function
  - no integer/binary variables
- An important relaxation for solving MINLP's
- What are the major challenges here?
  - Locally optimal solutions may exist
  - Solution found can therefore be initial guess-dependent
  - Solutions do not necessarily lie at extreme points

CPE 778, Lect. 15

99

Standard NLP form

$$\begin{aligned} \min \quad & f(\bar{x}) \\ \text{s.t.} \quad & h_i(\bar{x}) = b_i \\ & g_j(\bar{x}) \geq c_j \end{aligned}$$

- All of the functions *f, h,* and *g* may be nonlinear
  - Note we can use slack variables to convert inequalities to equalities
  - All variables (including slacks) may be negative
- Often, *n* (number of variables) as well as *i* and *j* are large

CPE 778, Lect. 15

100

Example: Lagrange Multipliers

- Let's look at a different example:

$$\begin{aligned} \min \quad & f(\bar{x}) = x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 - 1 = 0 \end{aligned}$$

- Note the only nonlinearity is in the constraint – still an NLP
- Let's first solve it graphically, to get a feel for the problem

CPE 778, Lect. 16

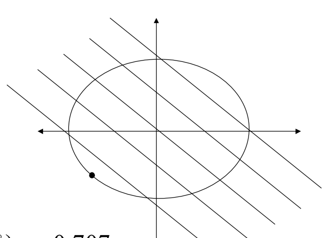
101

Example: Lagrange Multipliers

$$\begin{aligned} \min \quad & f(\bar{x}) = x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 - 1 = 0 \end{aligned}$$

- Where is the feasible region?
- Where is the minimum?

Can find using trig:



$x_1 = x_2 = 1 \cdot \cos(45^\circ) = -0.707$

CPE 778, Lect. 16

102

Example: Lagrange Multipliers

$$\min \quad f(\bar{x}) = x_1 + x_2$$
$$\text{s.t.} \quad h(\bar{x}) = x_1^2 + x_2^2 - 1 = 0$$

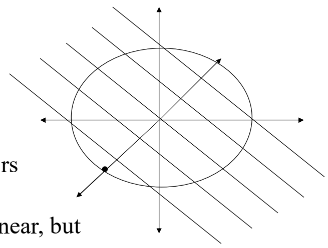
• Now take the gradient of f and h:

$\nabla f(\bar{x}) = [1,1]$

$\nabla h(\bar{x}) = [2x_1, 2x_2]_{x^*}$

$= [-1.414, -1.414]$

• Plot these as vectors from the origin:



• Note these are colinear, but differ by a constant

CPE 778, Lect. 16

103

Example: Lagrange Multipliers

- This colinearity is true for any constraint and objective function of an NLP, at a local optimum.
- The constant difference is called a Lagrange Multiplier: in this case,

$$\lambda = -\left(1/1.414\right)$$

• We can therefore write

$$\nabla f(\bar{x}^*) + \lambda^* \nabla h(\bar{x}^*) = 0$$

CPE 778, Lect. 16

104

The Lagrangian Function

• If we define the Lagrangian Function as:

$$L(\bar{x}, \lambda) = f(\bar{x}) + \lambda\, h(\bar{x})$$

• We can then write

$$\nabla L(\bar{x}, \lambda)\,|_{\bar{x}^*, \lambda^*} = 0$$

which is a necessary condition for optimality, along with

$h(\bar{x}) = 0$ for feasibility

CPE 778, Lect. 16

105

Lagrange Multipliers: Meaning

- The derivative of *f* with respect to *b* tells us how the objective changes with small changes in the constraint:

$$\frac{df}{db} = -\frac{1}{2}(2b)^{-\frac{1}{2}} \cdot 2 = -(2b)^{-\frac{1}{2}}$$

- But this is equal to λ^*!

So the Lagrange Multiplier tells us how sensitive the objective function is to changes in the constraint *h*, much like a marginal cost.

CPE 778, Lect. 16

106

Lagrange Multipliers for Inequalities

- A vector in N-dimensional space is "between" a set of other vectors if we can write it as nonnegative linear combination of the other vectors
- In our case, the other vectors are the active constraints
- We can write:

$$\nabla f(\bar{x}^*) = \sum_j u_j^* (-\nabla g_j(\bar{x}^*))$$

$$u_j^* \geq 0,\ \ j : \text{all active constraints}$$

CPE 778, Lect. 16

107

Logic of an Optimum Point

- Think of an optimum in the following way:

At any local optimum, no small feasible change in the values of the variables will improve the value of the objective function

- This logic allows us to write conditions required for a point to be locally optimal. Writing such conditions converts a constrained NLP problem into a nonlinear equation solving problem

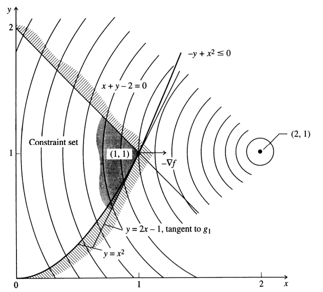
CPE 778, Lect. 17

108

Feasible Region

$$
\begin{aligned}
\text{min} \quad & f(x,y) = (x-2)^2 + (y-1)^2 \\
\text{s.t.} \quad & g_1(x,y) = -y + x^2 \le 0 \\
& g_2(x,y) = y + x \le 2 \\
& g_3(y) = y \ge 0
\end{aligned}
$$

- Note the optimum of the constrained problem is [1,1]
- First two constraints are active, third is inactive at [1,1]

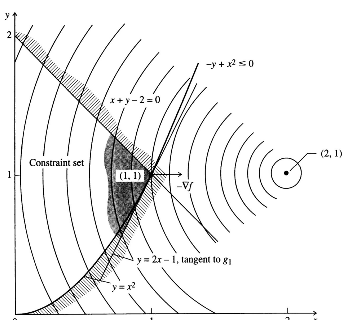


CPE 778, Lect. 17

109

Feasible Search Directions

- Move us in a direction that will remain in the feasible region
- Based on local information – tangents at current point
- Also note the negative gradient is plotted (steepest descent direction)



CPE 778, Lect. 17

110

Descent Directions

- Remember that any direction within 90° of grad *f* is a descent direction
- For this problem

$$
\begin{aligned}
& f(x,y) = (x-2)^2 + (y-1)^2 \\
& -\nabla f = [-2(x-2), -2(y-1)] = [2,0]
\end{aligned}
$$

- Note all feasible search directions lie between the tangent lines to the two active constraints:

$$
\begin{aligned}
\nabla g_1 &= [2x, -y] = [2, -1] \\
\nabla g_2 &= [1,1]
\end{aligned}
$$

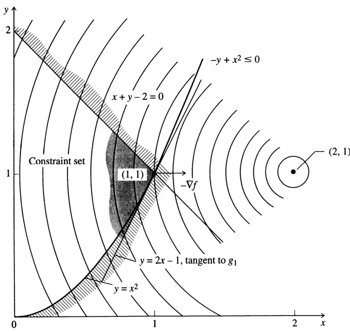
CPE 778, Lect. 17

111

Feasible Descent Directions

$$\begin{aligned} \text{min} \quad & f(x,y) = (x-2)^2 + (y-1)^2 \\ \text{s.t.} \quad & g_1(x,y) = -y + x^2 \le 0 \\ & g_2(x,y) = y + x \le 2 \\ & g_3(y) = y \ge 0 \end{aligned}$$

- Since none of the feasible search directions are within 90° of – grad f, we must be at an optimal point



CPE 778, Lect. 17

112

The KKT Conditions, Algebraically

- If we convert this geometric logic to a set of algebraic equations, we can solve them to find our optimum point
- This idea was worked on independently by Karush and Kuhn & Tucker in the 1960's
- What does it mean for a vector to be "between" two other vectors, algebraically?

CPE 778, Lect. 17

113

The KKT Conditions, Algebraically

- So the full set of equations becomes:

$$0 = \nabla f(\bar{x}^*) + \sum_j u_j^* \nabla g_j(\bar{x}^*)$$

$$u_j^* \ge 0, \; u_j^*[g_j(\bar{x}^*) - c_j] = 0$$

$$g_j(\bar{x}^*) - c_j \le 0$$

CPE 778, Lect. 17

114

Example: Writing the KKT Conditions

- Write the KKT conditions for:

$$\begin{aligned} &\min && f(\bar{x}) = x_1 x_2 \\ &\text{s.t.} && g_1(\bar{x}) = x_1^2 + x_2^2 \le 25 \end{aligned}$$

- Remember we need

$$0 = \nabla f(\bar{x}^*) + \sum_j u_j^* \nabla g_j(\bar{x}^*)$$

$$u_j^* \ge 0,\ u_j^*[g_j(\bar{x}^*) - c_j] = 0$$

$$g_j(\bar{x}^*) - c_j \le 0$$

CPE 778, Lect. 17

115

Example: Writing the KKT Conditions

- Note this is the same as grad *L*:

$$L = x_1 x_2 + u(x_1^2 + x_2^2 - 25)$$

$$\frac{\partial L}{\partial x_1} = x_2 + 2ux_1 = 0 \qquad \frac{\partial L}{\partial x_2} = x_1 + 2ux_2 = 0$$

$$\frac{\partial L}{\partial u} = g(x) = x_1^2 + x_2^2 - 25 \le 0 \qquad u[x_1^2 + x_2^2 - 25] = 0$$

- Remember we also need $u \ge 0$
- 3 equations, three unknowns, plus an inequality – solve numerically
- Note that if we had a maximization, the u's would need to be negative

CPE 778, Lect. 17

116