PRIMAL DUAL INTERIOR POINT METHOD

N. Arvind Kumar

NASA JSC / CACI
n.a.kumar@nasa.gov

TABLE OF CONTENTS

1. OPTIMIZATION PROBLEMS

An optimization problem is the problem of finding the best solution among all feasible solutions. It is typically stated as either a problem to minimize or maximize a function $f(\boldsymbol{x})$.

When $\boldsymbol{x}$ is written in bold font, it represents a vector of multiple variables.

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \dots \\ \dots \\ x_m \end{bmatrix}$$

The problem can either be an unconstrained optimization problem or a constrained optimization problem. For an unconstrained optimization problem, the solution is the global minimum of the function. For a constrained optimization problem, the constraints can be either equality constraints or inequality constraints.

Standard form of an optimization problem with several equality and inequality constraints:

> Minimize $f(\boldsymbol{x})$
> subject to
> $h_i(\boldsymbol{x}) \leq 0.$ $\quad (i = 1, \dots, m)$
> $\ell_j(\boldsymbol{x}) = 0$ $\quad (j = 1, \dots, n)$

The problem of maximizing $f(\boldsymbol{x})$ can be written in the above form by treating it as a problem of minimizing $-f(\boldsymbol{x})$. An inequality constraint of the form $h_i(\boldsymbol{x}) \geq 0$ can be written as $-h_i(\boldsymbol{x}) \leq 0$.

A feasible set is the set of points that satisfies the constraints. An optimal point is a solution to the problem, and the value at which it occurs is written as $\boldsymbol{x}^*$.

The notation $c^T\boldsymbol{x}$ is used for linear functions while $\frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x} + c^T\boldsymbol{x}$ is used for quadratic functions. $f(\boldsymbol{x})$ is used for the general case.

Not all problems have closed form solutions that can be calculated easily. Iterative methods are commonly used to arrive at the solution. Iterative algorithms such as the Simplex method search for the solution by traversing from one vertex to another with the worst-case complexity being exponential. In contrast, interior point methods traverse through the interior of the feasible region and are more efficient.

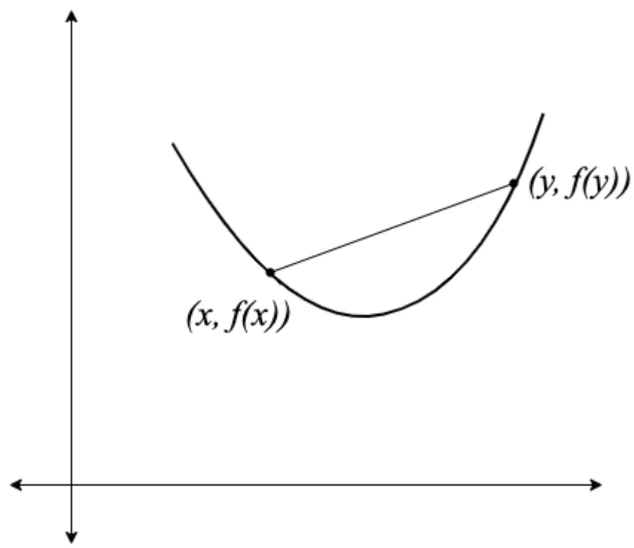
2. CONVEX OPTIMIZATION PROBLEMS

We will deal with a class of problems known as convex optimization problems as the local minimum of a convex function is also its global minimum.

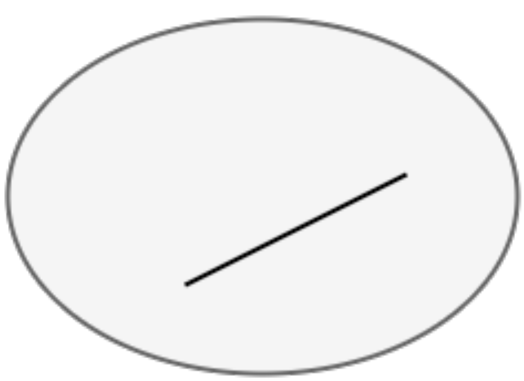
A function acting over domain $\mathcal{D}$ is a convex function if it satisfies the following condition.

$$f(\theta x + (1-\theta) y) \quad \leq \quad \theta f(x) + (1-\theta) f(y) \quad \forall \; x, y \in \mathcal{D} \text{ and } 0 \leq \theta \leq 1 \tag{2.1}$$

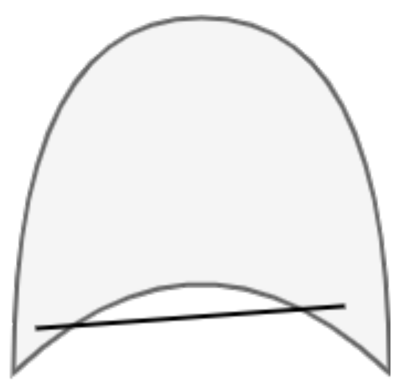
Geometrically, the inequality means that the line segment between $(x, f(x))$ and $(y, f(y))$ lies above the graph of f.



A line between two points in a convex set lies wholly within the set.



Convex Set



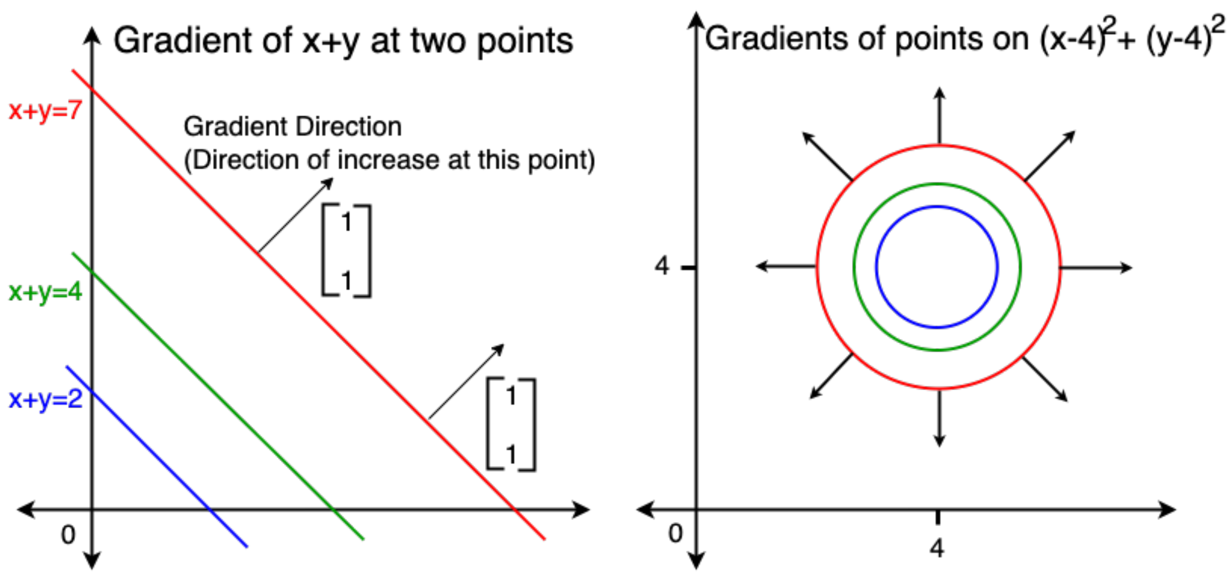
Not a Convex Set

A convex optimization problem is a minimization problem in which the objective function and all the constraints are convex functions, and the feasible region is a convex set. Note that the equality constraints for convex functions will be linear.

3. GRADIENT

The gradient of a scalar valued function is the vector field whose value at a point p gives the direction and rate of fastest increase at that point.

$$\nabla f(p) \quad = \begin{bmatrix} \frac{\partial f(p)}{\partial x_1} \\ \cdots \\ \cdots \\ \frac{\partial f(p)}{\partial x_n} \end{bmatrix} \tag{3.1}$$



The gradient of each point on x+y is the same since $\nabla_{(x+y)} f$ is independent of both x and y.

$$\nabla_{(x+y)} f(p) \quad = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} \quad = \begin{bmatrix} 1 \\ \\ 1 \end{bmatrix} \tag{3.2}$$

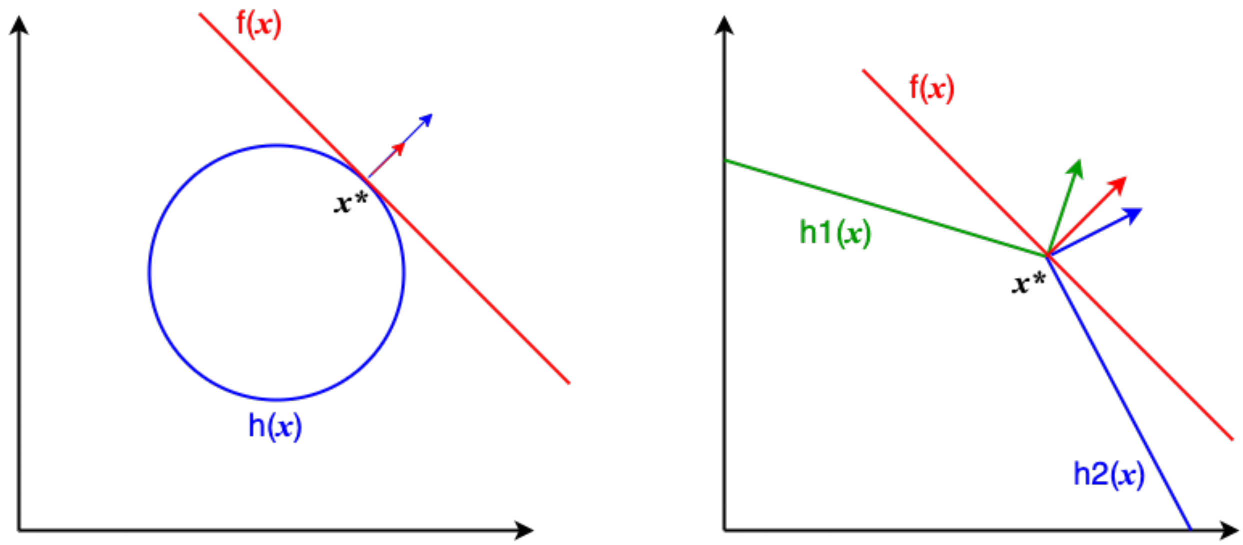
The gradient of the function $(x-4)^2 + (y-4)^2$ is different at each point and can be obtained by taking the partial derivative of the function with respect to x and y and plugging in the coordinate of each point. The directions of the gradients of various points on the function are shown in the diagram above.

4. BINDING CONSTRAINTS

A constraint is called a binding constraint if the optimal solution occurs on the constraint, that is, the constraint is satisfied as an equality. Otherwise, the constraint is a redundant constraint.

If the optimal solution exists but no constraint is binding, then the optimal solution occurs in the interior of the feasible region.

The gradient of the objective function at the optimal point x^* is a linear combination of the gradients of the binding constraints at that point.



$\nabla f(x^*) + \lambda \nabla h(x^*) = 0$ $\qquad$ $\nabla f(x^*) + \lambda_1 \nabla h_1(x^*) + \lambda_2 \nabla h_2(x^*) = 0$

In the general case with several constraints,

$$\frac{\partial f(x*)}{\partial x} + \sum_{i=1}^{m} \lambda_i \frac{\partial h_i(x*)}{\partial x} + \sum_{j=1}^{n} \nu_j \frac{\partial \ell_j(x*)}{\partial x} = 0 \qquad (4.1)$$

5. LAGRANGE MULTIPLIER

The purpose of using the Lagrange multiplier is to convert a constrained optimization problem with equality constraints into a form in which the derivative test can be applied.

Consider the problem:

Minimize $f(x)$
subject to
$\ell(x) = 0$

As the gradient of the objective function is a linear combination of the gradients of the constraints at the optimal point x^*, the following two equations should be satisfied:

$$\nabla f(x^*) + \lambda \nabla \ell(x^*) = 0 \tag{5.1}$$

and the given condition

$$\ell(x^*) = 0 \tag{5.2}$$

We can rewrite the above two equations by defining the Lagrange function $\mathcal{L}(x, \lambda)$ where λ is the Lagrange multiplier, and solve the optimization problem by solving the new equations.

Define

$$\mathcal{L}(x, \lambda) = f(x) + \lambda \ell(x) \tag{5.3}$$

Then

$$\frac{\partial \mathcal{L}}{\partial x} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \lambda} = 0 \tag{5.4}$$

Problem with one variable

In this case, we solve the following equations which correspond to (5.1) and (5.2)

$$\frac{\partial f(x)}{\partial x} + \lambda \frac{\partial \ell(x)}{\partial x} = 0 \tag{5.5}$$

and

$$\ell(x) = 0 \tag{5.6}$$

Problem with two variables

In case of a problem with two variables,

Minimize $f(x, y)$
subject to
$\ell(x, y) = 0$

we solve the following equations which correspond to (5.1) and (5.2)

$$\nabla f(x^*, y^*) + \lambda \nabla \ell(x^*, y^*) = 0 \tag{5.7}$$

and

$$\ell(x^*, y^*) = 0 \tag{5.8}$$

That is,

$$\begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} + \lambda \begin{bmatrix} \frac{\partial \ell(x,y)}{\partial x} \\ \frac{\partial \ell(x,y)}{\partial y} \end{bmatrix} = 0 \tag{5.9}$$

and

$$\ell(x^*, \mathrm{y}^*) = 0 \tag{5.10}$$

6. DUAL FUNCTION

An equivalent function that can be solved instead of solving the given objective function is known as the dual function.

6.1 Linear Case

Consider the following problem.

$$\text{Maximize } f(x) = 4x_1 + 3x_2 \tag{6.1}$$
subject to
$$2x_1 + x_2 \le 8 \tag{6.2}$$
$$x_1 + x_2 \le 4 \tag{6.3}$$
$$x_1, \quad x_2 \ge 0 \tag{6.4}$$

Multiplying (6.2) by 3 yields

$$6x_1 + 3x_2 \le 24 \tag{6.5}$$

Since x_1, $x_2 \ge 0$, the left-hand side of (6.5) $\ge f(x)$

$$6x_1 + 3x_2 \ge 4x_1 + 3x_2$$

Therefore, 24 is an upper bound for the objective function $f(x)$.

Similarly, multiplying (6.3) by 4 yields

$$4x_1 + 4x_2 \le 16 \tag{6.6}$$

Again, since x_1, $x_2 \ge 0$, the left-hand side of (6.6) $\ge f(x)$

$$4x_1 + 4x_2 \ge 4x_1 + 3x_2$$

Therefore, 16 is a better upper bound for the objective function $f(x)$.

Generalizing this idea of finding upper bounds from the constraints, we multiply the first inequality constraint by λ_1 and the second inequality constraint by λ_2 to obtain

$$2\lambda_1 x_1 + \lambda_1 x_2 \le 8\lambda_1 \tag{6.7}$$
$$\lambda_2 x_1 + \lambda_2 x_2 \le 4\lambda_2 \tag{6.8}$$

Adding (6.6) and (6.7)

$$(2\lambda_1 + \lambda_2)x_1 + (\lambda_1 + \lambda_2)x_2 \le 8\lambda_1 + 4\lambda_2 \tag{6.9}$$

Our goal is to find the least possible upper bound of $f(x)$, i.e., the minimum of the right-hand side of (6.9). Additionally, to guarantee the left-hand side of (6.9) being an upper bound, the coefficients in (6.9) must be greater than or equal to the corresponding coefficients in $f(x)$.

We can now rewrite our goal as the dual function of the original problem.

Minimize $8\lambda_1 + 4\lambda_2$	(Minimize the right-hand side of (6.9))	(6.10)
subject to		
$2\lambda_1 + \lambda_2 \geq 4$	(coefficient of x_1 in (6.9) $\geq$ coefficient of x_1 in $f(x)$)	(6.11)
$\lambda_1 + \lambda_2 \geq 3$	(coefficient of x_2 in (6.9) $\geq$ coefficient of x_2 in $f(x)$)	(6.12)
$\lambda_1, \ \lambda_2 \geq 0$	(this ensures (6.11) and (6.12) hold)	(6.13)

General notation in the Linear Case

In general, in the linear case, the maximization [minimization] of the primal function

$$\left.\begin{array}{l} max \;\; c^T x \\ s.t. \\ Ax \leq b, x \geq 0 \end{array}\right\} \quad (6.14)$$

corresponds to the minimization [maximization] of the dual function

$$\left.\begin{array}{l} min \;\; b^T \lambda \\ s.t. \\ A\lambda \geq c, \lambda \geq 0 \end{array}\right\} \quad (6.15)$$

and the right-hand side values of the constraints in the primal function are the coefficients of the dual function and vice versa.

The duality gap is the difference between the two solutions and is given by

$$c^T x - b^T \lambda \quad (6.16)$$

6.2 Non-Linear Case

Consider the problem:

Minimize $f(\boldsymbol{x})$
subject to
$h_i(\boldsymbol{x}) \leq 0 \quad (i = 1, \ldots, m)$

An equivalent problem can be defined as the problem of minimizing

$$J(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_i \mathrm{I}(h_i(\boldsymbol{x})) \tag{6.17}$$

where
$$\mathrm{I}(u) = \begin{cases} 0 \text{ if } u \leq 0 \\ \infty \text{ otherwise} \end{cases}$$

However, this problem is difficult to solve as the indicator function $\mathrm{I}(u)$ is not a continuous function. So let us replace $\mathrm{I}(u)$ by λu where $\lambda \geq 0$ is a constant to get the following Lagrangian function:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x}) \tag{6.18}$$

Then,

- $\max\limits_{\boldsymbol{\lambda} \geq 0} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = J(\boldsymbol{x})$
    - Proof: If x satisfies all constraints, each $h_i(x) \leq 0$, and $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ attains its maximum value of 0 when each $\lambda_i = 0$. Instead, if x violates some constraint, some $h_i(x) > 0$, and again, the maximum depends on λ_i as $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \to \infty$ when $\lambda_i \to \infty$. Thus, maximizing $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ over $\boldsymbol{\lambda}$ results in $\sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x})$ equaling $\sum_i \mathrm{I}(h_i(\boldsymbol{x}))$.
- By construction, minimizing $J(\boldsymbol{x})$ over $\boldsymbol{x}$ results in the solution of the original problem.

Combining the above two points, we note that maximizing $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ over $\boldsymbol{\lambda}$ yields $J(\boldsymbol{x})$, and then minimizing $J(\boldsymbol{x})$ over $\boldsymbol{x}$ solves the original problem.

Therefore, the original problem is equivalent to

$$\min_{\boldsymbol{x} \in \mathrm{X}} \max_{\boldsymbol{\lambda} \geq 0} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{6.19}$$

Swapping the order of min and max, we get

$$\max_{\boldsymbol{\lambda} \geq 0} \min_{\boldsymbol{x} \in \mathrm{X}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{6.20}$$

$g(\boldsymbol{x}, \boldsymbol{\lambda})$ is the dual function where

$$g(\boldsymbol{x}, \boldsymbol{\lambda}) = \min_{\boldsymbol{x} \in \mathrm{X}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{6.21}$$

and the Lagrangian dual program is

$$\max_{\boldsymbol{\lambda} \geq 0} g(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{6.22}$$

6.3 Properties of the Dual Function

- A solution for one of the primal or dual problems guarantees a solution for the other.
- If the primal problem is a maximization problem, then the dual problem is a minimization problem and vice versa.
- The number of dual variables in the dual function is equal to the number of constraints in the primal function.
- The number of constraints in the dual function is equal to the number of variables in the primal function.

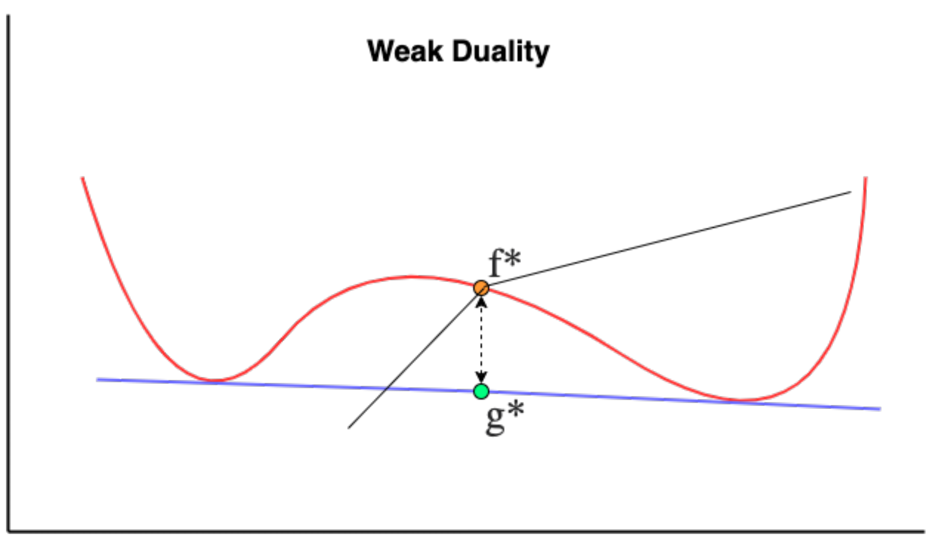
6.4 Duality Gap

The difference between the primal and dual solutions is known as the duality gap.

Note that $f^* \geq g^*$ always holds. This is called weak duality.

If $f^* - g^* = 0$, it is called strong duality.

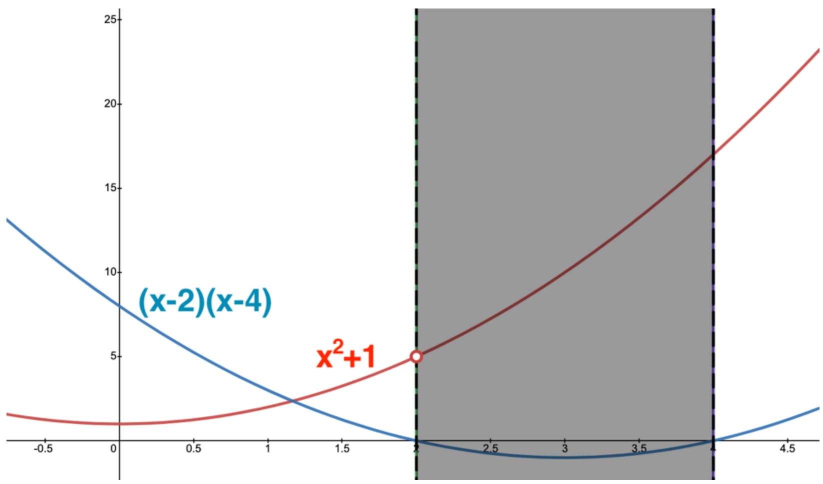
Strong duality holds for most of the convex problems.



6.5 Examples

Example 1

$min\ x^2+1$
such that
$(x-2)(x-4) \le 0$



$$\mathcal{L}(x, \lambda) = x^2 + 1 + \lambda(x-2)(x-4) \qquad (6.23)$$

$$= (1+\lambda)x^2 - 6\lambda x + (1+8\lambda) \qquad (6.24)$$

To find the minimum of (6.24) over x, differentiate $\mathcal{L}(x, \lambda)$ with respect to x and equate it to 0.

$$2(1+\lambda)x - 6\lambda = 0 \qquad (6.25)$$

$$x = 3\lambda\, /\, (1+\lambda) \quad \text{for } \lambda > -1 \qquad (6.26)$$

(Since the second derivative is positive for $\lambda > -1$, this value of x yields a minimum.)

Substituting for x, we obtain

$$\min_{x \in X} \mathcal{L}(x, \lambda) = -9\lambda^2 / (1+\lambda) + 1 + 8\lambda \qquad (6.27)$$

We now need to maximize (6.27) over λ to obtain

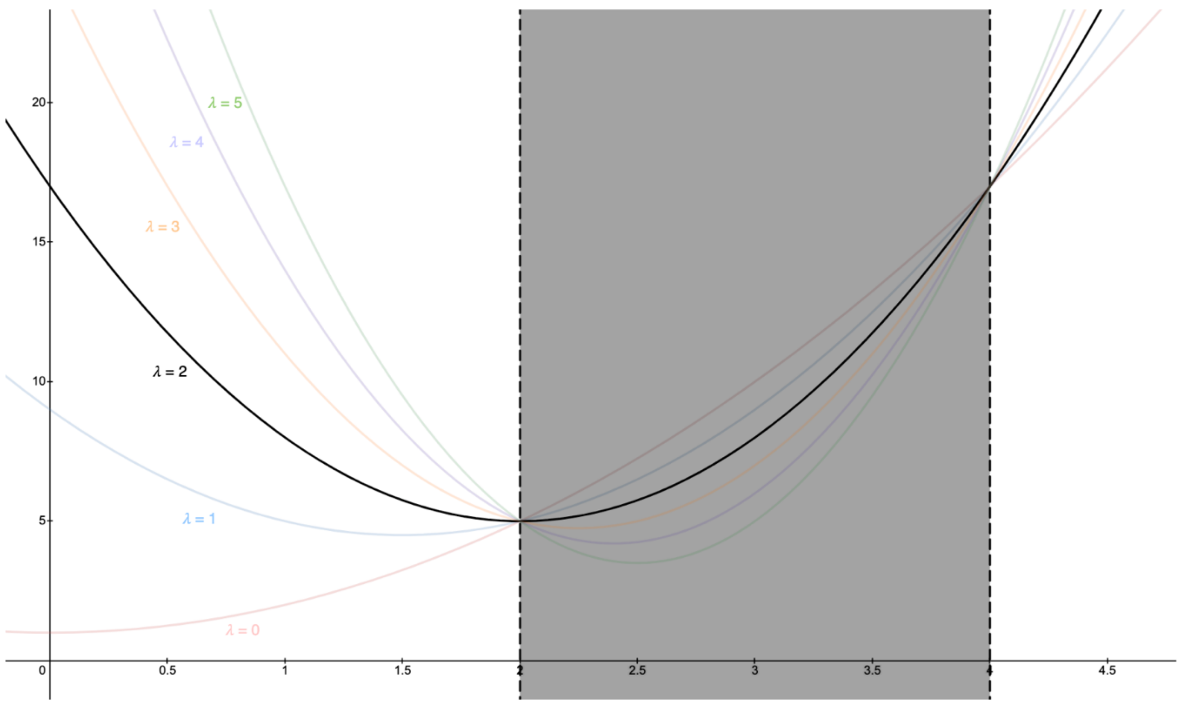
$$\max_{\lambda \ge 0} \min_{x \in X} \mathcal{L}(x, \lambda) \qquad (6.28)$$

Differentiating (6.27) with respect to λ, equating it to 0, we get

$$\lambda^2 + 2\lambda - 8 = 0 \qquad (6.29)$$

$\lambda = 2$ is the only solution which satisfies $\lambda \ge 0$. The value of (6.27), which is the dual function, is 5 when $\lambda = 2$. The second derivative is negative and hence it is a maximum.

$x^2+1 + \lambda(x-2)(x-4)$ for various values of λ
Maximum of the set of minima of these curves occurs at $\lambda = 2$



<u>Observations</u>

- Each curve passes through the optimal point x^* because $h(x) = 0$ at x^* and adding $\lambda h(x)$ to $f(x)$ at x^* has the effect of adding 0 to the original objective function.

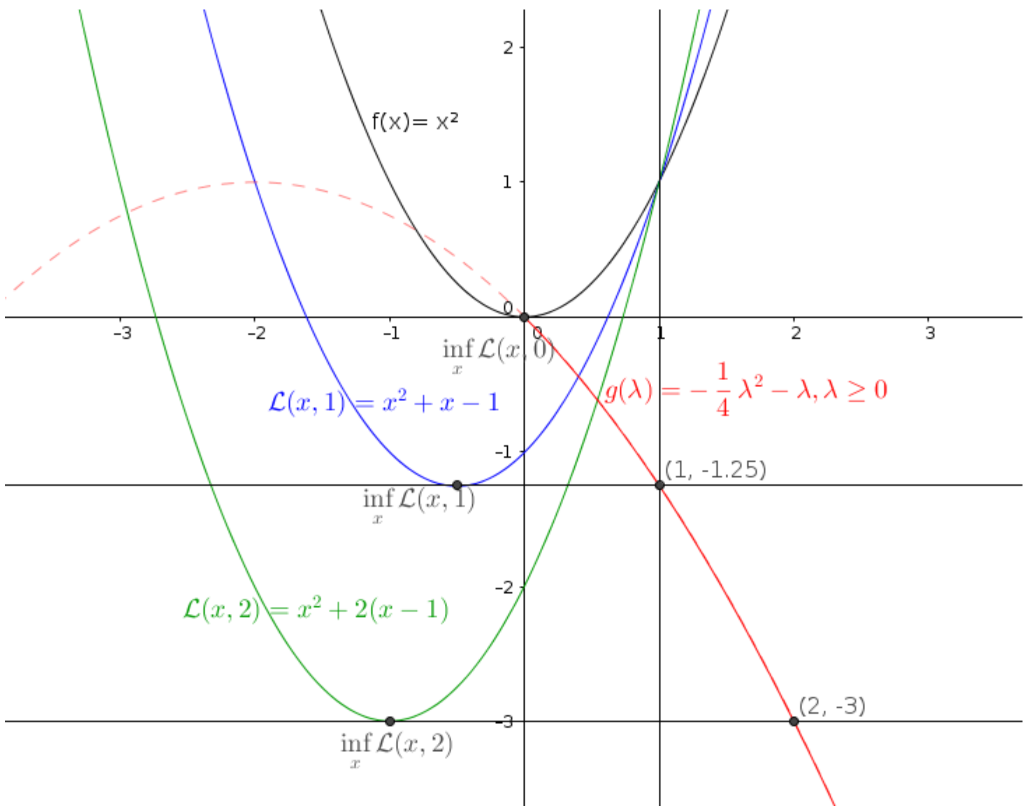
- Since we add 0 at x^*, the minimum of each of these curves is a lower bound for our problem, and the maximum of these lower bounds gets us to the point $(x^*, f(x^*))$.

- To add more detail, since each curve passes through x^*, the minimum of each curve taken individually must be $\leq f(x^*)$. Hence, the curve with its minimum value exactly at x^* is the maximum among the set of minima of all such curves (as the minimum of every other curve $\leq f(x^*)$). So, to find the minimum value of the curve which has its minimum occurring exactly at x^*, we find the maximum among the set of minima of all such curves. The value of the minimum of this curve yields $f(x^*)$ which is also the minimum of the objective function since the value of this curve at x^* is the sum of the value of the objective function at x^* and 0.

- In other words, minimizing each curve over x and then maximizing the resulting set over λ is a way of arriving at $(x^*, f(x^*))$ and hence solving the optimization problem.

Example 2

$min\ x^2$
such that
$x - 1 \leq 0$



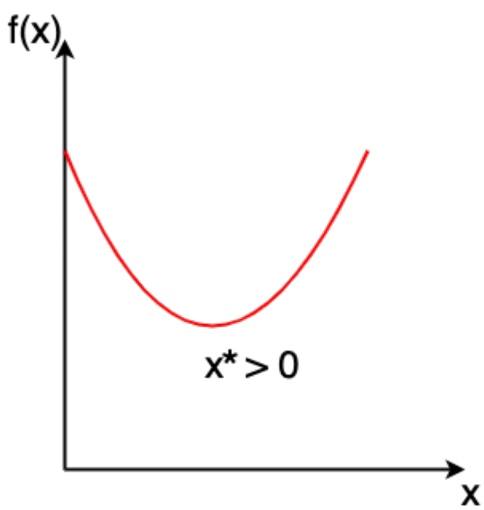
Differentiating $g(x, \lambda)$ with respect to λ and setting it to 0 does not yield a value of λ that is greater than 0, but -2 corresponding to the peak of the red curve.

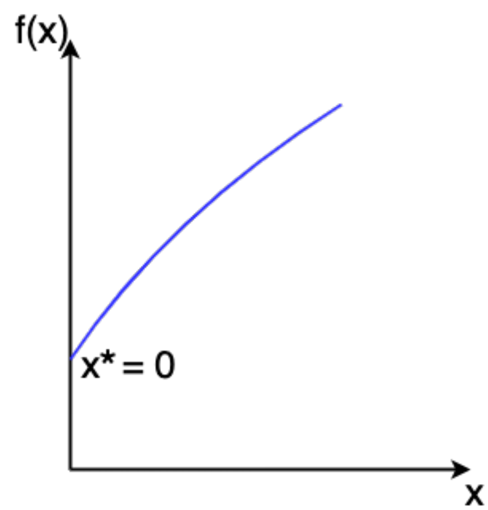
7. NON-NEGATIVITY CONSTRAINTS

For a univariate function f with non-negativity constraint $x > 0$, let the minimum occur at x^*.

- Either x^* is an interior point of the feasible region,
  $x^* > 0$ and $f'(x^*) = 0$

- Or x^* is a boundary point of the feasible region.
  $x^* = 0$ and $f'(x^*) \geq 0$





Combining the above two conditions, we get

$$f'(x^*) \geq 0, \qquad x^* \geq 0, \qquad \text{and} \qquad x^* f'(x^*) = 0 \tag{7.1}$$

For a multivariate function $f(\boldsymbol{x})$ with non-negativity constraints $x_i > 0$, we get such a condition for each of the variables:

$$f_{x_i}(\boldsymbol{x}^*) \geq 0, \qquad x_i^* \geq 0, \qquad \text{and} \qquad x_i^* f_{x_i}(\boldsymbol{x}^*) \geq 0 \tag{7.2}$$

8. SLACK VARIABLES

Slack variables convert an inequality constraint into an equality constraint. For example,

$$x_1^2 + x_2^2 \leq 20 \tag{8.1}$$

becomes

$$x_1^2 + x_2^2 - 20 + s = 0, \quad s \geq 0 \tag{8.2}$$

Consider the following optimization problem.

Minimize $f(\boldsymbol{x})$
subject to
$h_i(\boldsymbol{x}) \leq 0 \qquad (i = 1, \ldots, m)$

By introducing the slack variables $s_i \geq 0$, the inequalities $h_i(\boldsymbol{x}) \leq 0$ are converted to the equality constraints $h_i(\boldsymbol{x}) + s_i = 0$, and the problem is transformed into the following.

Minimize $f(\boldsymbol{x})$
subject to
$h_i(\boldsymbol{x}) + s_i = 0 \qquad (i = 1, \ldots, m)$
$s_i \geq 0$

The corresponding Lagrange function is

$$\tilde{\mathcal{L}}(\boldsymbol{x}, s, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i (h_i(\boldsymbol{x}) + s_i) \tag{8.3}$$

Applying the non-negativity conditions in (7.2), we get

$$\frac{\partial \tilde{\mathcal{L}}}{\partial s_i} \geq 0, \qquad s_i \geq 0, \qquad \text{and} \qquad s_i \frac{\partial \tilde{\mathcal{L}}}{\partial s_i} \geq 0 \tag{8.4}$$

Since $\frac{\partial \tilde{\mathcal{L}}}{\partial s_i} = \lambda_i$, this can be written as

$$\lambda_i \geq 0, \qquad s_i \geq 0, \qquad \text{and} \qquad \lambda_i s_i = 0 \tag{8.5}$$

Substituting $s_i = -h_i(\boldsymbol{x})$,

$$\boxed{\lambda_i \geq 0, \qquad h_i(\boldsymbol{x}) \leq 0, \quad \text{and} \qquad \lambda_i h_i(\boldsymbol{x}) = 0 \tag{8.6}}$$

The slack variables have been eliminated, and we can obtain the same conditions by replacing $\tilde{\mathcal{L}}(x, s, \lambda)$ by $\mathcal{L}(x, \lambda)$, since

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x}) \tag{8.7}$$

and

(8.6) can be written as $\lambda_i \geq 0, \quad \frac{\partial \mathcal{L}}{\partial \lambda_i} \leq 0 \quad \text{and} \quad \lambda_i \frac{\partial \mathcal{L}}{\partial \lambda_i} = 0$ (8.8)

9. KARUSH KUHN TUCKER CONDITIONS (KKT CONDITIONS)

Karush Kuhn Tucker Conditions or KKT Conditions generalize the method of lagrangian multipliers for inequality constraints and capture the concepts described in the previous sections.

Equations (7.2) and (8.8) form the KKT conditions for the problem,

Minimize $f(\boldsymbol{x})$
subject to
$h_i(\boldsymbol{x}) \leq 0$ ($i = 1, \ldots, m$)
$x_j \geq 0$

Rewriting (7.2) and restating (8.8), the KKT conditions are as follows:

$$x_j \geq 0, \quad \frac{\partial \mathcal{L}}{\partial x_j} \geq 0 \quad \text{and} \quad x_j \frac{\partial \mathcal{L}}{\partial x_j} = 0 \tag{9.1}$$

and

$$\lambda_i \geq 0, \quad \frac{\partial \mathcal{L}}{\partial \lambda_i} \leq 0 \quad \text{and} \quad \lambda_i \frac{\partial \mathcal{L}}{\partial \lambda_i} = 0 \tag{9.2}$$

For the problem:

Minimize $f(\boldsymbol{x})$
subject to
$h_i(\boldsymbol{x}) \leq 0$ ($i = 1, \ldots, m$)
$\ell_j(\boldsymbol{x}) = 0$ ($j = 1, \ldots, n$),
$x_j \geq 0$

if we ignore the case of $x_j = 0$ and consider only $x_j > 0 \ \forall j$, the KKT conditions become:

Stationarity

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\boldsymbol{x}^*) + \sum_{j=1}^{n} \nu_i \nabla \ell_i(\boldsymbol{x}^*) = 0 \tag{9.3}$$

Complementary Slackness

$$\sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x}^*) = 0 \tag{9.4}$$

Primal feasibility

$$h_i(\boldsymbol{x}^*) \leq 0 \quad (i = 1, \ldots, m) \qquad \text{and} \qquad \ell_j(\boldsymbol{x}^*) = 0 \quad (j = 1, \ldots, n) \tag{9.5}$$

Dual feasibility

$$\lambda_j \geq 0 \tag{9.6}$$

KKT conditions are sufficient conditions to solve the optimization problem when the primal problem is convex

If (x^*, λ^*, ν^*) satisfies the KKT conditions, then x^* and (λ^*, ν^*) are primal and dual optimal with zero duality gap.

Proof

Since KKT conditions are satisfied, x^* satisfies (9.5). Therefore, x^* is primal feasible.

(x^*, λ^*, ν^*) also satisfies (9.6). Therefore, $\lambda_j \geq 0$. Since $f(\boldsymbol{x})$ is convex and $\lambda_j \geq 0$, $\mathscr{L}(x, \lambda^*, \nu^*)$ is also convex. Since the stationarity condition in (9.3) is satisfied, the gradient of $\mathscr{L}(x, \lambda^*, \nu^*)$ is 0 at x^*, and as $\mathscr{L}(x, \lambda^*, \nu^*)$ is convex, it follows that x^* minimizes $\mathscr{L}(x, \lambda^*, \nu^*)$.

That is,

$$g(\lambda^*, \nu^*) = min\ \mathscr{L}(x, \lambda^*, \nu^*)$$

$$= \mathscr{L}(x^*, \lambda^*, \nu^*) \quad \text{(since } x^* \text{ minimizes } \mathscr{L}(x, \lambda^*, \nu^*) \text{ as shown above)}$$

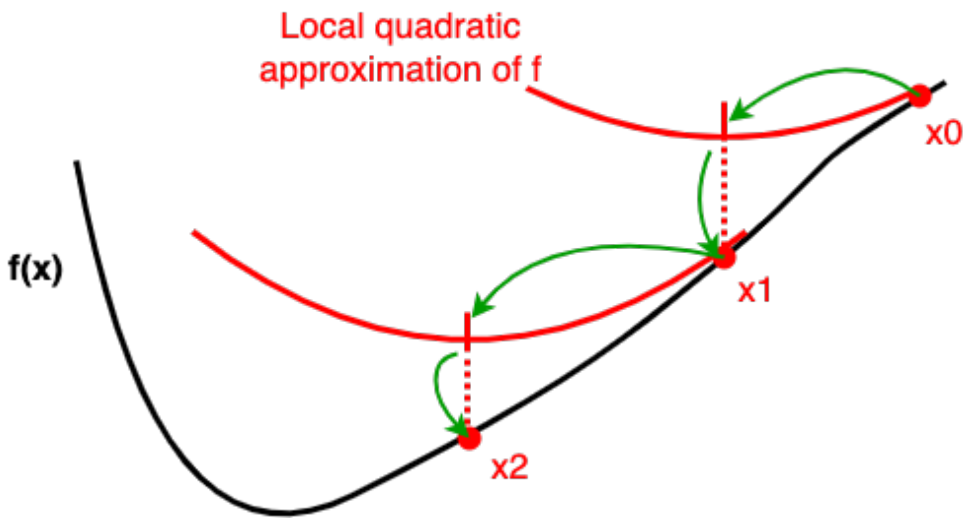
$$= f(\boldsymbol{x^*}) + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x^*}) + \sum_{j=1}^{m} \nu_i \ell_i(\boldsymbol{x^*})$$

$$= f(\boldsymbol{x^*}) + 0 + 0 \quad \text{(due to complementary slackness and primal feasibility)}$$

$$= f(\boldsymbol{x^*})$$

This means x^* and (λ^*, ν^*) have zero duality gap.

10. NEWTON'S METHOD



From the Taylor's series expansion of $f(x+\ \Delta x)$,

$$f(x+\ \Delta x) = f(x) + \Delta x\, f'(x) + \frac{(\Delta \mathrm{x})^2}{2!} f''(x) + \ldots \qquad (10.1)$$

Ignoring terms with degree greater than 2,

$$f(x+\ \Delta x) \approx f(x) + \Delta x\, f'(x) + \frac{(\Delta \mathrm{x})^2}{2!} f''(x) \qquad (10.2)$$

The minimum of $f(x+\ \Delta x)$ as Δx varies occurs when

$$\frac{df(x+\ \Delta \mathrm{x})}{d\Delta \mathrm{x}} = 0$$

$$0 + f'(x) + f''(x)\,\Delta x = 0$$

$$f'(x) + f''(x)\,\Delta x = 0$$

$$\Delta x = -\,f'(x)\,/\,f''(x)$$

$$x_{\mathrm{n+1}} = x_{\mathrm{n}} - \ f'(x_{\mathrm{n}})\,/\,f''(x_{\mathrm{n}}) \qquad (10.3)$$

This can be generalized to the multivariate case by replacing $f'(x)$ by the Jacobian and the reciprocal of $f''(x)$ by the inverse of the Hessian.

$$\boldsymbol{x}_{\mathrm{n+1}} = \boldsymbol{x}_{\mathrm{n}} - [\nabla^2 f(\boldsymbol{x}_{\mathrm{n}})]^{-1} \nabla f(\boldsymbol{x}_{\mathrm{n}}) \qquad (10.4)$$

Jacobian and Hessian

The Jacobian of a function in several variables is

$$\nabla \boldsymbol{f}(x_1, x_2, \ldots x_n) = [\,\partial \boldsymbol{f}/\partial x_1 \quad \partial \boldsymbol{f}/\partial x_2 \quad \ldots \quad \partial \boldsymbol{f}/\partial x_n\,]$$

$$= \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \ldots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \ldots & \partial f_2/\partial x_n \\ \vdots & \vdots & & \vdots \\ \partial f_n/\partial x_1 & \partial f_n/\partial x_2 & \ldots & \partial f_n/\partial x_n \end{bmatrix}$$

The Jacobian of the gradient of a scalar function in several variables is the Hessian.

$$\nabla^2 f(x, y, z) = \begin{bmatrix} \partial^2 f/\partial x^2 & \partial^2 f/\partial x\partial y & \partial^2 f/\partial x\partial z \\ \partial^2 f/\partial y\partial x & \partial^2 f/\partial y^2 & \partial^2 f/\partial y\partial z \\ \partial^2 f/\partial z\partial x & \partial^2 f/\partial z\partial y & \partial^2 f/\partial z^2 \end{bmatrix}$$

11. INTERIOR POINT METHODS

OUTLINE

(1) Define the Barrier Method that keeps $\boldsymbol{x}$ in the interior of the feasible region

(2) Define the corresponding Log Barrier function and the Central Path

(3) Obtain KKT Conditions for the Central Path

(4) Conditions for the original problem from the Central Path KKT Conditions

(5) Interpretation as Perturbed KKT Conditions

(6) Find the solution of the system of equations given by the Perturbed KKT Conditions

11.1 Barrier Method

The barrier method is a method by which a penalty function is added to the objective function to force it to remain within the interior of the feasible region.

$$\left.\begin{array}{l} \text{Minimize } f(\boldsymbol{x}) \\ \text{subject to} \\ h_i(\boldsymbol{x}) \le 0 \qquad (i = 1, \ldots, m) \\ A\boldsymbol{x} = b \end{array}\right\} \tag{11.1}$$

For $i = 1, \ldots, m$, define

$$I_{\{h_i(x)\,\le\,0\}}(x) \;=\; \begin{cases} 0 & \text{if } h_i(\boldsymbol{x}) \le 0 \\ \infty & \text{if } h_i(\boldsymbol{x}) > 0 \end{cases} \tag{11.2}$$

The original problem is equal to

$$\begin{array}{l} \text{Minimize } f(\boldsymbol{x}) + \sum_{i=1}^{m} I_{\{h_i(x)\,\le\,0\}}(\boldsymbol{x}) \\ \text{subject to } A\boldsymbol{x} = b \end{array} \tag{11.3}$$

There are no inequality constraints as they are now implied the indicator functions.

11.2 Log Barrier Method and the Central Path

The sum of the indicator functions is not continuous. So we approximate it by the penalty function $(1/t)\ \phi(\boldsymbol{x})$ where $\phi(\boldsymbol{x})$ is the following function known as the log barrier function:

$$\phi(\boldsymbol{x}) \;=\; \begin{cases} -\sum_{i=1}^{m} log(-h_i(\boldsymbol{x})) & \text{if } h_i(\boldsymbol{x}) \le 0 \\ \infty & \text{if } h_i(\boldsymbol{x}) > 0 \end{cases} \tag{11.4}$$

Define the central path as $f(\boldsymbol{x}) + (1/t)\ \phi(\boldsymbol{x})$. Then, the original problem is equivalent to

$$\begin{array}{l} \text{Minimize } f(x) + (1/t)\ \phi(\boldsymbol{x}) \\ \text{subject to } A\boldsymbol{x} = b \end{array} \tag{11.5}$$

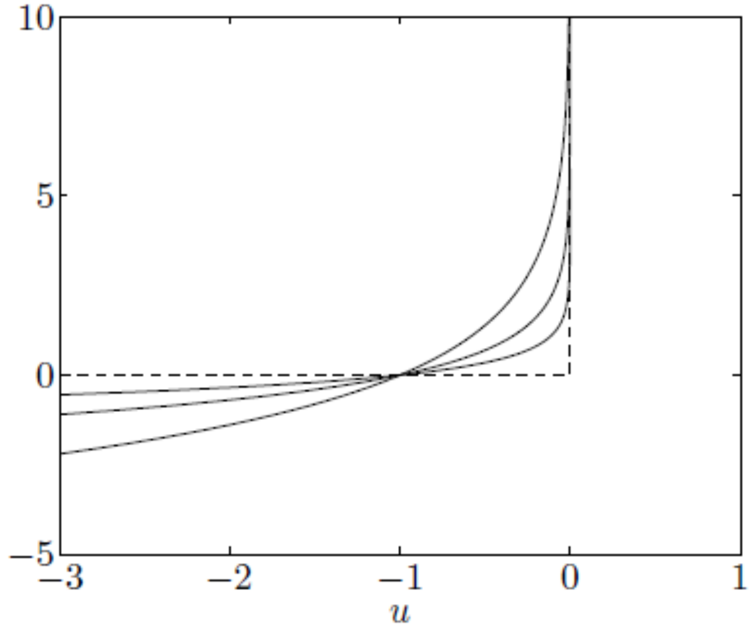
<u>Property 1</u>: Solving the above problem yields $f(\boldsymbol{x}^*)$, the minimum of the original problem.

<u>Property 2</u>: The minimum of $\phi(\boldsymbol{x})$ occurs when x in the interior of the feasible region. Proof: For $h_i(\boldsymbol{x}) > 0$, $\phi(\boldsymbol{x}) = \infty$. For $h_i(\boldsymbol{x}) \le 0$, as $\boldsymbol{x}$ gets closer to the boundary of the feasible region (i.e., $\boldsymbol{x}$ tends to leave the interior), $h_i(x)$ gets closer to 0 for some i, and hence $-\log(-h_i(\boldsymbol{x}))$ approaches ∞.

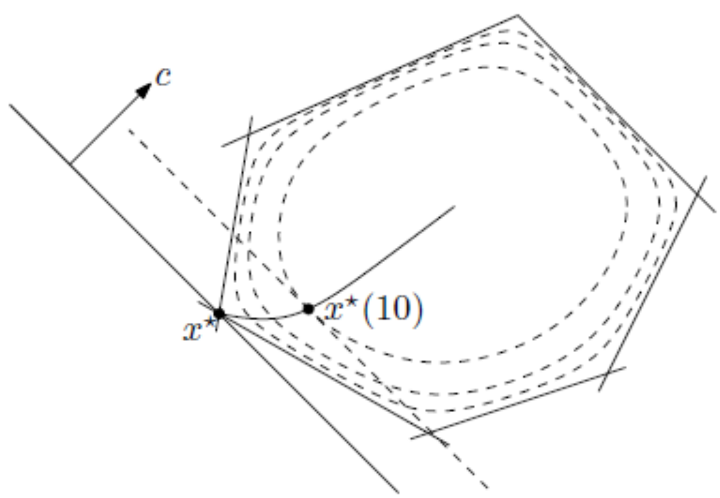
<u>Property 3</u>: The penalty function, $(1/t)\ \phi(\boldsymbol{x})$, converges to zero as $t \to \infty$, and the solution approaches the optimal value of the original objective function $f(\boldsymbol{x}^*)$.

Illustration

[Diagrams taken from *Convex Optimization* by Stephen Boyd and Lieven Vandenberghe, p. 563 and p. 566.]



As $t \to \infty$, the approximation of the second term becomes closer to the indicator function, and the limit of the central path is the solution to our original problem.



The contours of the log barrier function for different values of t over different iterations.

11.3 KKT Conditions for the Central Path

Since the central path is

$$f(\boldsymbol{x}) + (1/t)\,\phi(\boldsymbol{x}) \quad = \quad f(\boldsymbol{x}) - (1/t)\ \sum_{i=1}^{m} log(-h_i(\boldsymbol{x})) \tag{11.6}$$

We can write the problem as

Minimize $tf(\boldsymbol{x}) + \phi(\boldsymbol{x})$ (11.7)
subject to $A\boldsymbol{x} = b$

As $\boldsymbol{x}^*(t)$ is the solution of this problem for each t, and

$$\nabla\phi(\boldsymbol{x}) \quad = \quad -\sum_{i=1}^{m} \frac{1}{h_i(x)} \nabla h_i(\boldsymbol{x}) \tag{11.8}$$

the necessary and sufficient conditions for $\boldsymbol{x}^*(t)$ to be optimal are:

Primal Feasibility

$$\left.\begin{aligned} &A\boldsymbol{x}^*(t) = b, \\ &h_i(\boldsymbol{x}^*(t)) \le 0 \qquad (i = 1, \ldots, m) \end{aligned}\right\} \tag{11.9}$$

Stationarity

$$t\nabla f(\boldsymbol{x}^*(t)) \ -\sum_{i=1}^{m} \frac{1}{h_i(x)} \nabla h_i(\boldsymbol{x}) + A^{\mathrm{T}} w \ = \ 0$$

(11.10)

for some $w \in \mathbb{R}^n$ (w is the dual variable for the equality constraint in the definition of the Lagrangian).

11.4 Conditions for the original problem from the Central Path KKT Conditions

The original problem is

$$\min_{x \in X} f(\boldsymbol{x})$$

subject to

$h_i(\boldsymbol{x}) \leq 0 \quad (i = 1, \ldots, m)$

$A\boldsymbol{x} = b$

Primal Feasibility

The primal feasibility conditions for the original problem are the same as the primal feasibility conditions for the central path:

$$\boxed{A\boldsymbol{x^*}(t) = b, \qquad h_i(\boldsymbol{x^*}(t)) \leq 0 \qquad (i = 1, \ldots, m)} \tag{11.11}$$

Stationarity

Dividing the stationarity condition of the Central Path by t we get

$$\nabla f(\boldsymbol{x^*}(t)) - \sum_{i=1}^{m} \frac{1}{t\, h_i(x)} \nabla h_i(x) + A^{\mathrm{T}} \frac{w}{t} = 0 \tag{11.12}$$

Define

$$\lambda_i^*(t) = -\frac{1}{t\, h_i(x^*(t))} \qquad \text{and} \quad \nu^*(t) = \frac{w}{t} \tag{11.13}$$

Then, we can write (11.12) as

$$\boxed{\nabla f(\boldsymbol{x^*}(t)) + \sum_{i=1}^{m} \lambda_i^*(t) \nabla h_i(x) + A^{\mathrm{T}} \nu^*(t) = 0} \tag{11.14}$$

If we can verify that $[\lambda_i^*(t), \nu_i^*(t)]$ is dual feasible for the original problem, (11.14) would be the stationarity condition for the original problem.

Dual Feasibility

To verify that $[\lambda_i^*(t), \nu_i^*(t)]$ is dual feasible for the original problem, we need to verify that the signs of $\lambda_i^*(t)$ are positive, and that each $[\lambda_i^*(t), \nu_i^*(t)]$ lies in the domain of $g(\lambda, \nu)$.

Since

$$h_i(x^*(t)) < 0,$$

$$\boxed{\lambda_i^*(t) > 0 \qquad (11.15)}$$

The domain of

$$g(\lambda, \nu) = \min_{x \in X} \mathcal{L}(x, \lambda, \nu)$$

consists of all points (λ, ν) such that $\lambda > 0$ and $\min_{x \in X} \mathcal{L}(x, \lambda, \nu) >$ -∞.

This is because a solution for one of the primal or dual problems guarantees a solution for the other and so $g(\lambda, \nu)$ must have a minimum if the primal problem has a solution.

Since (11.14) gives us the equation

$$\nabla f(\boldsymbol{x^*(t)}) + \sum_{i=1}^{m} \lambda_i^*(t) \nabla h_i(x) + A^{\mathrm{T}} \nu^*(t) = 0$$

$\min_{x \in X} \mathcal{L}(x, \lambda^*(t), \nu^*(t))$ occurs at $\boldsymbol{x^*(t)}$ which means the minimum occurs at some value $>$ -∞.

Hence, $\left(\lambda_i^*(t), \nu_i^*(t)\right)$ lies in the domain of $g(\lambda, \nu)$ and is dual feasible for the original problem (not the barrier problem). This also means that (11.14) is the Stationarity condition for the original problem as $\lambda_i^*(t)$ *and* $\nu_i^*(t)$ in (11.14) are dual feasible.

Complementary Slackness

From (11.13), we get

$$\boxed{\lambda_i h_i(\boldsymbol{x}) = -\frac{1}{t} \qquad (i = 1, \ldots, m) \qquad (11.16)}$$

As $t \to \infty$, the right-hand side goes to 0.

11.5 Interpretation as Perturbed KKT

Taken together, (11.11), (11.14), (11.15) and (11.16) in the Section 6.4 are as follows:

$$\nabla f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + A^{\mathrm{T}}\nu = 0$$

$$\lambda_i h_i(\boldsymbol{x}) = -\frac{1}{t} \qquad (i = 1, \ldots, m)$$

$$A\boldsymbol{x} = b \ \text{ and } \ h_i(\boldsymbol{x}) \leq 0 \qquad (i = 1, \ldots, m)$$

$$\lambda_i \geq 0 \qquad (i = 1, \ldots, m)$$

We can think of these as perturbed KKT conditions where the second equation has $-\frac{1}{t}$ instead of 0 on the right-hand side while the other equations correspond to KKT conditions.

11.6 Solution of system of equations given by Perturbed KKT

The log barrier method can be viewed as solving a set of perturbed KKT conditions with $\frac{1}{t}$ converging to 0.

To solve this set of equations, we eliminate λ_i by substituting it with $-\frac{1}{t\, h_i(x)}$ and apply Newton's Method during successive iterations and until the stopping condition is met.

As t becomes larger, $\frac{1}{t}$ approaches 0, and we get closer to solving the actual KKT conditions.

<u>Dual Feasibility provides a stopping condition</u>

Dual feasibility in Section 6.4 allows us to use a stopping condition for the log barrier method.

$$g(\lambda_i^*(t), \nu_i^*(t)) = f(\boldsymbol{x^*(t)}) + \sum_{i=1}^{m} \lambda_i^*(t)\, h_i(\boldsymbol{x}^*(t)) + \nu^*(t)^{\mathrm{T}}(A\, \boldsymbol{x^*(t)} - b) \qquad (11.17)$$

Since $\lambda_i^*(t) = -\frac{1}{t\, h_i(x^*(t))}$, the second term of (11.17) is $-\frac{m}{t}$ and the third term is 0, and so

$$g(\lambda^*(t), \nu^*(t)) = f(\boldsymbol{x^*(t)}) - \frac{m}{t}$$

That is, $\frac{m}{t}$ is the duality gap.

As $t \to \infty$, we get closer to the solution, and we can check if $f(\boldsymbol{x^*(t)}) - f^* \leq \frac{m}{t}$ and use it as the stopping condition.

12. PRIMAL-DUAL INTERIOR POINT METHOD

The Primal-Dual Interior Point Method solves the perturbed KKT conditions, but unlike in the barrier method, we do not eliminate the dual variable λ.

We also take just one Newton step instead of going through the entire Newton Method in each iteration and use backtracking to correct the step size. Our stopping conditions are also different.

We can write the perturbed KKT system as

$$\mathrm{r}(x, \lambda, \nu) \;=\; 0 \tag{12.1}$$

where

$$\mathrm{r}(x, \lambda, \nu) \;=\; \begin{bmatrix} \nabla f(x) + \nabla h(x)^T \lambda + A^T \nu \\ -diag(\lambda) h(x) - 1/t \\ Ax - b \end{bmatrix} \begin{array}{l} \longleftarrow r_{dual} \text{ (dual residual)} \\ \longleftarrow r_{cent} \text{ (centrality residual)} \\ \longleftarrow r_{prim} \text{ (primal residual)} \end{array} \tag{12.2}$$

$$h(\mathrm{x}) \;=\; \begin{bmatrix} h_1(x) \\ \cdots \\ \cdots \\ h_m(x) \end{bmatrix} \qquad \nabla h(x) \;=\; \begin{bmatrix} \nabla h_1(x)^T \\ \cdots \\ \cdots \\ \nabla h_m(x)^T \end{bmatrix}$$

This is a non-linear set of equations. We use a linear approximation using Taylor's series and solve the approximation to get to the next point. This amounts to taking one Newton step.

Let $y = (x, \lambda, \nu)$ be the current iterate, $\Delta y = (\Delta x, \Delta\lambda, \Delta\nu)$ be the update direction.

$$\mathrm{r}(x{+}\Delta x, \lambda{+}\Delta\lambda, \nu{+}\Delta\nu) \;=\; 0$$

$$\mathrm{r}(x{+}\Delta x, \lambda{+}\Delta\lambda, \nu{+}\Delta\nu) \;\approx\; \mathrm{r}(x, \lambda, \nu) \;+\; \nabla \mathrm{r}(x, \lambda, \nu) \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{bmatrix} \;=\; 0 \tag{12.3}$$

$$\nabla \mathrm{r}(x, \lambda, \nu) \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{bmatrix} \;=\; -\,\mathrm{r}(x, \lambda, \nu) \tag{12.4}$$

$$\nabla \left(\begin{bmatrix} \nabla f(x) + \nabla h(x)^T \lambda + A^T \nu \\ -diag(\lambda) h(x) - 1/t \\ Ax - b \end{bmatrix} \right) \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{bmatrix} = -\begin{bmatrix} \nabla f(x) + \nabla h(x)^T \lambda + A^T \nu \\ -diag(\lambda) h(x) - 1/t \\ Ax - b \end{bmatrix} \tag{12.5}$$

Then,

$$\boxed{\begin{bmatrix} \nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x) & \nabla h(x)^T & A^T \\ -diag(\lambda)\nabla h(x) & -diag(h(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + \nabla h(x)^T \boldsymbol{\lambda} + A^T \nu \\ -diag(\lambda)h(x) - 1/t \\ Ax - b \end{bmatrix} \quad (12.6)}$$

Writing $\ell(x)$ instead of $Ax = b$ (equality constraints), the general form is as given below. Optimization problems can be reformulated in this form by adding slack variables to the nonlinear inequality constraints and converting them into equality constraints.

$$\begin{bmatrix} \nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x) + \nabla^2 \ell(x)\nu & \nabla h(x)^T & \nabla(\ell(x))^T \\ -diag(\lambda)\nabla h(x) & -diag(h(x)) & 0 \\ \nabla(\ell(x)) & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + \nabla h(x)^T \lambda + \nabla(\ell(x))^T \nu \\ -diag(\lambda)h(x) - 1/t \\ \ell(x) \end{bmatrix}$$

We obtain the values of Δx, $\Delta \lambda$, and $\Delta \nu$ from the above equation, and use these values to update x, λ, and ν. We go through multiple iterations of this process until we hit the stopping condition.

Stopping condition for the Primal-Dual Interior Point method

Since the intermediate points are not necessarily dual feasible unlike in the case of the barrier method (they may violate either the primal equality condition $Ax - b = 0$ or the dual equality condition $A^T \nu = 0$), we do not have the duality gap as the stopping condition.

Instead, we create a surrogate duality gap and check that the surrogate duality gap is small.

Let $\eta = \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\lambda} = -\sum_{i=1}^{m} \lambda_i h_i(x)$

Then $t = \dfrac{m}{\eta}$

This would be the duality gap if x was primal feasible, and λ and ν were dual feasible (if $r_{prim} = 0$ and $r_{dual} = 0$), but we do not. So, we check that the surrogate duality gap is small, and also check that the primal and dual residuals, r_{prim} and r_{dual} (amount of feasibility violated), are small.

13. ALGORITHM

Initialize strictly feasible points $x^{(0)} \mid h(x^{(0)}) < 0$, $\lambda^{(0)} > 0$, $\nu^{(0)}$

$$\eta^{(0)} \;=\; -h\left(x^{(0)}\right)^T \lambda^{(0)}$$

Iterate

Define $t = \dfrac{\mu m}{\eta^{(k-1)}}$

Compute primal-dual update direction $\Delta y = (\Delta x, \Delta\lambda, \Delta\nu)$.

Determine step size s using backtracking method.

Update $y^{(k)} = y^{(k-1)} + s\Delta y$

Compute $\eta^{(k)} = -h\left(x^{(k)}\right)^T \boldsymbol{\lambda}^{(k)}$

Stop if $\eta^{(k)} \leq \varepsilon$ and $\|r_{prim}\|_2 \leq \varepsilon$ and $\|r_{dual}\|_2 \leq \varepsilon$

<u>Backtracking Line Search</u>

At each step, we arrive at $y^{(k)} = y^{(k-1)} + s\Delta y$, i.e.,

$x^{(k)} = x^{(k-1)} + s\Delta x$ $\qquad$ $\lambda_i^{(k)} = \lambda_i^{(k-1)} + s\Delta\lambda$ $\qquad$ $\nu^{(k)} = \nu^{(k-1)} + s\Delta\nu$

by choosing s which maintains $h_i(\boldsymbol{x}) < 0$ and $\lambda_i > 0$ $(i = 1, \ldots, m)$

Start with largest step size $s_{max} \leq 1$ that makes $\lambda + s\Delta\lambda \geq 0$, i.e.,

$s_{max} = min\,(1,\ min\,[\,\lambda_i \,/\, \Delta\lambda_i : \Delta\lambda_i < 0\,]\,)$

Then, with $\alpha, \beta \in (0, 1)$, set $s = 0.999\, s_{max}$

Set $s = \beta s$ until $h_i(\boldsymbol{x}) < 0$ $(i = 1, \ldots, m)$

Set $s = \beta s$ until $\|r(x^+, \lambda^+, \nu^+)\|_2 \leq (1 - \alpha s)\, \|r(x, \lambda, \nu)\|_2$

Example

Minimize $x_1(x_2+5)$
subject to
$x_1 x_2 \geq 5$
$x_1^2 + x_2^2 \leq 20$
$x_1, x_2 \geq 0$

[Problem taken from https://optimization.cbe.cornell.edu/index.php?title=Interior-point_method_for_NLP#Numerical_Example]

Step 1: Transform into the form below consistent with the form in the section on slack variables.

Minimize $f(x)$
subject to
$\ell(x) = 0$
$x \geq 0$

Do this by writing all inequality constraints except the non-negativity constraints in the form $h(x) \leq 0$ and convert them to equality constraints by adding slack variables.

Minimize $x_1(x_2+5)$
subject to

$$\left.\begin{aligned} x_1 x_2 - 5 - x_3 &= 0 \\ -x_1^2 - x_2^2 + 20 - x_4 &= 0 \end{aligned}\right\} \text{ equality constraints with } x_3 \text{ and } x_4 \text{ as slack variables}$$

$x_1, \; x_2, \; x_3, \; x_4 \; \geq 0$

Step 2: Make initial guesses and calculate each element of the matrices.

$x_1 = 2,\; x_2 = 3,\; x_3 = 1,\; x_4 = 1,\; \lambda_1 = 1,\; \lambda_2 = 1,\; \lambda_3 = 1,\; \lambda_{4,} = 1,\; \nu_1 = 1,\; \nu_2 = 1$

$$h(x) = \begin{bmatrix} -x_1 \\ -x_2 \\ -x_3 \\ -x_4 \end{bmatrix}$$

$$\nabla h(x) = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

$$diag(h(x)) = \begin{bmatrix} -x_1 & 0 & 0 & 0 \\ 0 & -x_2 & 0 & 0 \\ 0 & 0 & -x_3 & 0 \\ 0 & 0 & 0 & -x_4 \end{bmatrix} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

$$\ell(x) = \begin{bmatrix} x_1x_2 - 5 - x_3 \\ -x_1^2 - x_2^2 + 20 - x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

$$\nabla\ell(x)^{\mathrm{T}} = \begin{bmatrix} x_2 & -2x_1 \\ x_1 & -2x_2 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 3 & -4 \\ 2 & -6 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\nabla^2(\ell_1(x)) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \nabla^2(\ell_2(x)) = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\nabla(f(x)) = \begin{bmatrix} 5 + x_2 \\ x_1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \\ 0 \\ 0 \end{bmatrix} \qquad \nabla^2(f(x)) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Step 3: Plug in these values into the equation and solve for the update direction $[\Delta x, \Delta\lambda, \Delta\nu]$.

$$\begin{bmatrix} \nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x) + \sum_{j=1}^{n} \nu_j \nabla^2 \ell_j(x) & \nabla h(x)^T & \nabla(\ell(x))^T \\ -diag(\lambda)\nabla h(x) & -diag(h(x)) & 0 \\ \nabla(\ell(x)) & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + \nabla h(x)^T \boldsymbol{\lambda} + \nabla(\ell(x))^T \boldsymbol{\nu} \\ -diag(\lambda)h(x) - 1/t \\ \ell(x) \end{bmatrix}$$

$$\left[\begin{array}{cccc|cccc|cc} 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 3 & -4 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 2 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 \\ \hline 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 3 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -4 & -6 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \\ \Delta x_4 \\ \Delta\lambda_1 \\ \Delta\lambda_2 \\ \Delta\lambda_3 \\ \Delta\lambda_4 \\ \Delta\nu_1 \\ \Delta\nu_2 \end{bmatrix} = - \begin{bmatrix} 6 \\ -3 \\ -2 \\ -2 \\ 2 - 1/t \\ 3 - 1/t \\ -1/t \\ -1/t \\ 0 \\ 6 \end{bmatrix}$$

Step 4: Modify (x, λ, ν) by adding the calculated values of $(\Delta x, \Delta\lambda, \Delta\nu)$ multiplied by a positive number less than 1 that is determined using the backtracking line search method to limit the magnitude in the new update direction, check for stopping condition, and iterate through all steps until the stopping condition is met.

The progression of the values of (x, λ, v) after each iteration until the stopping condition is met is shown below. These values are the updated values after going through Step 4 of each iteration.

Iteration	x_1	x_2	x_3	x_4	λ_1	λ_2	λ_3	λ_4	ν_1	ν_2
Initial	2.0000	3.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	1.5058	3.6460	0.8094	1.2562	0.8459	0.3975	0.7475	0.3007	0.7475	0.3007
2	1.3179	3.8290	0.3405	1.0811	0.7907	0.3021	1.0355	0.2829	1.0355	0.2829
3	1.2259	3.9876	0.1088	0.8662	0.6076	0.1986	1.4256	0.2531	1.4256	0.2531
4	1.1866	4.1318	0.0290	0.5959	0.3516	0.1014	1.8240	0.2172	1.8240	0.2172
5	1.1673	4.2501	0.0048	0.2842	0.1215	0.0305	2.1211	0.1850	2.1211	0.1850
6	1.1609	4.2962	0.0013	0.1072	0.0385	0.0092	2.2097	0.1729	2.2097	0.1729
7	1.1586	4.3123	0.0004	0.0351	0.0118	0.0028	2.2358	0.1689	2.2358	0.1689
8	1.1578	4.3175	0.0001	0.0109	0.0036	0.0008	2.2437	0.1676	2.2437	0.1676
9	1.1576	4.3191	0.0000	0.0033	0.0011	0.0002	2.2460	0.1672	2.2460	0.1672

The solution for the problem is

$x_1 = 1.1576$
$x_2 = 4.3191$

14. REFERENCES

[1] Stephen Boyd and Lieven Vandenberghe, Convex Optimization: https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

[2] Ryan Tibshirani, Primal-Dual Interior Point Methods, Lecture Notes: https://www.stat.cmu.edu/~ryantibs/convexopt/lectures/primal-dual.pdf

[3] Andreas Wächter and Lorenz T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming: http://cepac.cheme.cmu.edu/pasilectures/biegler/ipopt.pdf

[4] Example 1 of section 6.5: https://see.stanford.edu/materials/lsocoee364a/hw4sol.pdf

[5] Example 2 of section 6.5: Mass Zhou (https://math.stackexchange.com/users/129681/mass-zhou), Please explain the intuition behind the dual problem in optimization., URL (version: 2022-02-01): https://math.stackexchange.com/q/1938329

[6] Diagram in section 10: Nick Alger (https://math.stackexchange.com/users/3060/nick-alger), Newton's method intuition, URL (version: 2013-12-17): https://math.stackexchange.com/q/609755

[7] Example in Section 13 taken from Cornell University Computational Optimization Open Textbook: https://optimization.cbe.cornell.edu/index.php?title=Interior-point_method_for_NLP