(Convex) optimization under constraints:

Example

minimize $x+y$ s.t. $x^2+y^2=1$

Recall: method of Lagrange multipliers

$$L(x,y,\lambda) = x+y+\lambda(x^2+y^2-1)$$

$L(x,y,\lambda)$ has stationary pts given by $\nabla L(x,y,\lambda) = 0$ where

$$\nabla L(x,y,\lambda) = \begin{pmatrix} \partial L/\partial x \\ \partial L/\partial y \\ \partial L/\partial \lambda \end{pmatrix} = \begin{pmatrix} 1+2\lambda x \\ 1+2\lambda y \\ x^2+y^2-1 \end{pmatrix}$$

so the stationary pts satisfy

$y = x = -\frac{1}{2\lambda}$ so that $\left(-\frac{1}{2\lambda}\right)^2 + \left(-\frac{1}{2\lambda}\right)^2 - 1 = 0$

i.e $\lambda = \pm\sqrt{2}/2$

leading to $(x,y) = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ ← maximizer of $x+y$ s.t. x^2+y^2

& $(x,y) = (-\sqrt{2}/2, -\sqrt{2}/2)$ ← minimizer

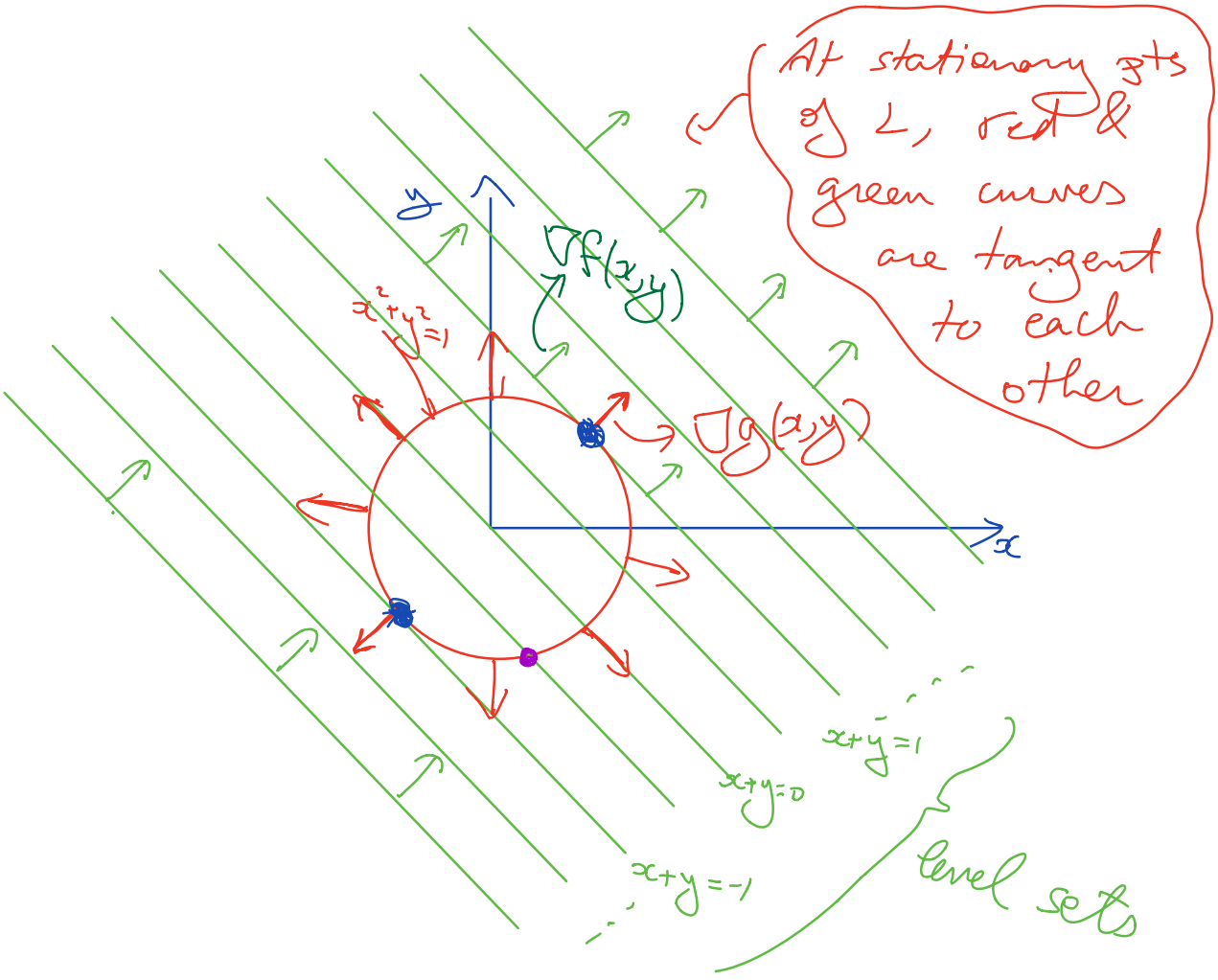
Interpretation ?

Consider

$$\min f(x,y) \quad \text{s.t.} \quad g(x,y) = 0$$

$$\Rightarrow L(x,y,\lambda) = f(x,y) + \lambda g(x,y)$$

$$\& \ \nabla L = 0 \Rightarrow \begin{cases} \nabla f(x,y) = -\lambda \nabla g(x,y) \\ g(x,y) = 0 \end{cases}$$



If the level curves of f & g are not tangent (e.g. at the purple pt.) then we can still move along the red curve and increase/decrease the function $f(x,y)$.

To summarize:

At the stationary pts of $\mathcal{L}$:

$$\begin{cases} \nabla f \text{ \& } \nabla g \text{ are in the same direction } (\Leftrightarrow \nabla f = -\lambda \nabla g \\ g(\vec{x}) = 0 \end{cases}$$

both are satisfied by finding a stationary pt of $\mathcal{L}(x,y,\lambda)$

Caveat: Stationary pts of $\mathcal{L}$ may be a min, max, or neither.

Multiple Constraints:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad \begin{aligned} h_1(x) &= 0 \\ h_2(x) &= 0 \\ &\vdots \\ h_p(x) &= 0 \end{aligned}$$

Then $L(x, \nu) = f(x) + \sum_{i=1}^{p} \nu_i h_i(x)$

$x \in \mathbb{R}^n$, $\nu \in \mathbb{R}^p$

$$= f(x) + \nu^T h(x)$$

$$h(x) = \begin{pmatrix} h_1(x) \\ h_2(x) \\ \vdots \\ h_p(x) \end{pmatrix}$$

$$\nabla L(x, \nu) = 0 \iff \begin{cases} \nabla f(x) = -\sum_{i=1}^{p} \nu_i \nabla h_i(x) \\ h_i(x) = 0 \quad \forall i \in \{1, \dots, p\} \end{cases}$$

Graphical interpretation: The intersection of all the level surfaces $h_i(x) = 0$ is tangent to the level surface of f.

More generally:

Consider the constrained opt. problem

$$\min f(x) \quad \text{s.t.} \quad \begin{cases} g_i(x) \le 0 & i = 1, \dots, m \\ h_i(x) = 0 & i = 1, \dots, p \end{cases}$$

The Lagrangian associated with it is:

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

The variables λ_i, ν_i are called Lagrange multipliers.

Some important examples:

Least squares: $\min x^T x$ s.t. $Ax = b$ (A is $p \times n$)

$$\Rightarrow L(x, \nu) = x^T x + \nu^T (Ax - b)$$

Linear optimization:

$$\min c^T x \quad \text{s.t.} \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

$$\Rightarrow L(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (Ax - b)$$

Lagrangian Duality:

Definition: Given a Lagrangian $L(x, \lambda, \nu)$, the Lagrangian dual is

$$F(\lambda, \nu) = \min_x L(x, \lambda, \nu)$$

→ In reality this should be the infimum

Example: Dual of least squares:

Here $L(x, v) = x^T x + v^T(Ax - b)$

$$\Rightarrow F(v) = \min_x \{x^T x + v^T(Ax - b)\}$$

convex in x

$\Rightarrow$ we can find the min by finding a stationary pt.

$$\nabla_x L(x, v) = 0 \Rightarrow 2x + A^T v = 0$$

$$\Rightarrow x^* = -\frac{A^T v}{2}$$

Plugging back into $L(x, v)$ we get

$$F(v) = \frac{1}{4} v^T A A^T v - \frac{1}{2} v^T A A^T v - v^T b$$

$$= -\frac{1}{4} v^T A A^T v - v^T b$$

Example: Dual of linear opt.

Here $L(x,\lambda,\nu) = (c^T + \nu^T A - \lambda^T)x - \nu^T b$

$\Rightarrow F(\lambda,\nu) = \min_x \{ \underbrace{(c^T + \nu^T A - \lambda^T)x}_{\text{linear fct of } x} - \underbrace{\nu^T b}_{\text{const. in } x} \}$

Two cases: either (*) $c^T + \nu^T A - \lambda^T = 0$

$\Rightarrow F(\lambda,\nu) = -\nu^T b$

or (**) $c^T + \nu^T A - \lambda^T \neq 0$

$\Rightarrow F(\lambda,\nu) = -\infty$

Thus $F(\lambda,\nu) = \begin{cases} -b^T\nu & c + A^T\nu - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$

Let $y := -\nu$, then when $\lambda \geq 0$

$F(\lambda,y) = \begin{cases} b^T y, & A^T y = c - \lambda \leq c \\ -\infty & \text{otherwise} \end{cases}$

↑ dual of Linear program.

Dual optimization problem:

Defn: We refer to the constrained opt. problem

$$\min_x f(x) \quad \text{s.t.} \begin{cases} g_i(x) \leq 0 & i = 1, \dots, m \\ h_j(x) = 0 & j = 1, \dots, p \end{cases}$$

as the primal optimization problem.

Defn: Its associated dual opt. problem is then

$$\max_{\lambda, \nu} F(\lambda, \nu) \quad \text{s.t. } \lambda \geq 0$$

where $F(\lambda, \nu) = \min_x L(x, \lambda, \nu)$

Duality Theory:

Weak duality: Theorem

If $\alpha^* = f(x^*)$ is the optimal value for the primal opt. problem & if β^* is the optimal (max) value for the dual opt. prob. then

$$\beta^* \leq \alpha^*$$

Proof: Recall that

$$F(\lambda, \nu) = \min_{x \in D} \left(f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

- $x \in D$: Feasible region of primal
- $\lambda_i g_i(x) \leq 0$ for all feasible x
- $\nu_i h_i(x) = 0$ for all feasible x

So, for all $\lambda_i \geq 0$, $\nu_i \in \mathbb{R}$ we have

$$f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \leq f(x)$$

In particular for λ_i^*, ν_i^* which maximize $F(\lambda, \nu)$, i.e. $F(\lambda^*, \nu^*) = \beta^*$, we have

$$\beta^* = \inf_{x \in I} L(x, \lambda^*, \nu^*) \leq \inf_{x \in \mathcal{D}} f(x) = \alpha^*$$

In words: Maximizing the dual problem gives a lower bound for the minimum of the primal problem.

↳ This is always true

Strong Duality:

Here, we want $\alpha^* = \beta^*$ so that we can find the min of the primal by finding the max of the dual.

Example (will return to this later)

When the function $f(x)$ is linear & when the constraints are linear, we have $\alpha^* = \beta^*$

Def'n: A primal opt. problem satisfies Slater's condition if

- f is convex
- g_i, $i = 1, \dots, m$ are convex
- h_j, $j = 1, \dots, p$ are linear
- $\exists \bar{x}$: $g_i(\bar{x}) < 0 \quad \forall i = 1, \dots, m$
  $h_j(\bar{x}) = 0 \quad \forall j = 1, \dots, p$

Theorem: Slater's condition $\Rightarrow \alpha^* = \beta^*$

To prove the theorem, we first need a lemma:

Lemma: Let

$$\mathcal{A} = \{(u, v, t) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \text{ such that}$$

$$\exists x \in D,\ g_i(x) \le u_i \quad i = 1, \dots, m$$

$$h_j(x) = v_j \quad i = 1, \dots, p$$

$$\& \ f(x) \le t\}$$

$$\mathcal{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \text{ s.t.}$$

$$s < \alpha^*\}$$

Then $\mathcal{A}$ & $\mathcal{B}$ are convex and $\mathcal{A} \cap \mathcal{B} = \emptyset$

Proof sketch: $\mathcal{A}$ & $\mathcal{B}$ convex follows from their def'n.

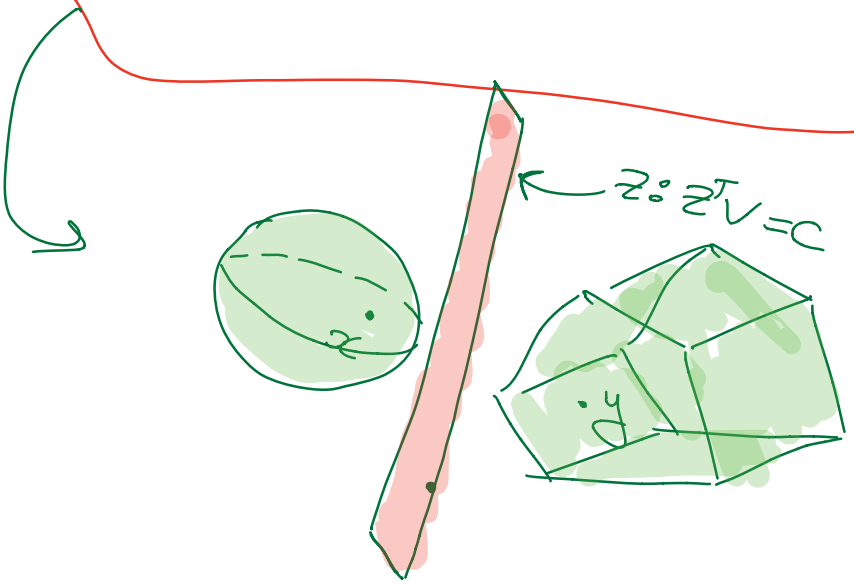
$\mathcal{A} \cap \mathcal{B} = \phi$, by contradiction.

We will also need the hyperplane separation theorem:

If: $A, B \subseteq \mathbb{R}^n$, $B \cap A = \emptyset$, $B \neq \emptyset$, $A \neq \emptyset$ and A, B are convex, then

$\exists\, v \in \mathbb{R}^n$, $c \in \mathbb{R}$ s.t.

$x^T v \geq c$ & $y^T v \leq c$ $\forall x \in A$, $y \in B$



Proof of Theorem (Slater's Cond $\Rightarrow \alpha^* = \beta^*$)
(sketch)

Write the linear constraints as $Ax = b$, and assume A is full rank

- $\mathcal{A}, \mathcal{B}$ convex, $\mathcal{A} \cap \mathcal{B} = \emptyset$ $\Rightarrow$ by hyperplane sepr. theorem

$\Rightarrow \exists (\tilde{\lambda}, \tilde{v}, \tilde{\mu}) \neq 0$ & $\alpha \in \mathbb{R}$:

(1) $\tilde{\lambda}^T u + \tilde{v}^T v + \tilde{\mu} t \geq \alpha$

(2) $\tilde{\mu} s \leq \alpha$

↑ lower bound

$\forall \begin{cases} (u, v, t) \in \mathcal{A} \\ (0, 0, s) \in \mathcal{B} \end{cases}$

(1) $\Rightarrow \tilde{\lambda} \geq 0$ & $\tilde{v} \geq 0$

(otherwise we have unboundedness)

(2) $\Rightarrow \tilde{\mu} s \leq \alpha \quad \forall s < \alpha^*$

so

$$\sum_{i=1}^{m} \tilde{\lambda}_i g_i(x) + \tilde{v}^T (Ax - b) + \tilde{\mu} f(x) \geq \alpha \geq \tilde{\mu} \alpha^*$$

divide by $\tilde{\mu}$ (if $\tilde{\mu} > 0$)

$$\Rightarrow L(x, \tilde{\lambda}/\tilde{\mu}, \tilde{\nu}/\tilde{\mu}) = \sum_{i=1}^{m} \frac{\tilde{\lambda}_i}{\tilde{\mu}_i} g_i(x) + \left(\frac{\tilde{\nu}}{\tilde{\mu}}\right)^T (Ax-b) + f(x) \geq \alpha^*$$

$\forall x \in D$,

$\Rightarrow$ True for the "minimizer" of $L(x, \lambda, \mu)$

so that

$$\alpha^* \leq F(\lambda, \nu) \leq \beta^*$$

($\leq$ by the above; $\leq$ by def'n)

But $F(\lambda, \nu) \leq \alpha^*$ by weak duality ($\beta^* \leq \alpha^*$)

$\Rightarrow F(\lambda, \nu) = \alpha^*$

i.e. strong duality holds

The case $\tilde{\mu} = 0$ leads to the same conclusion ... ∎