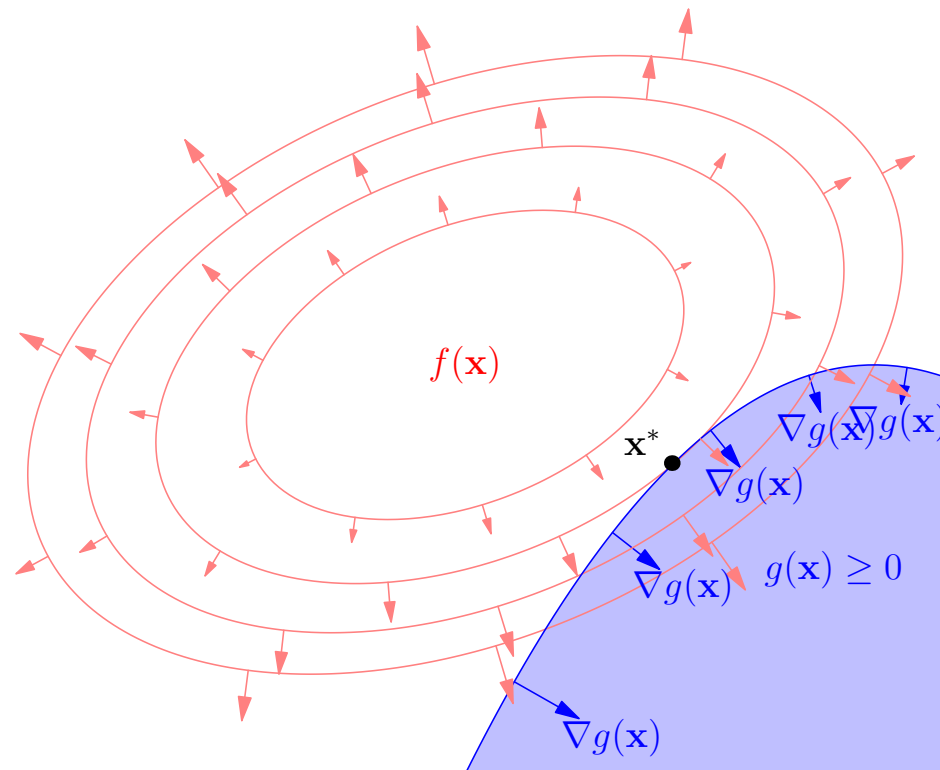


Advanced Machine Learning

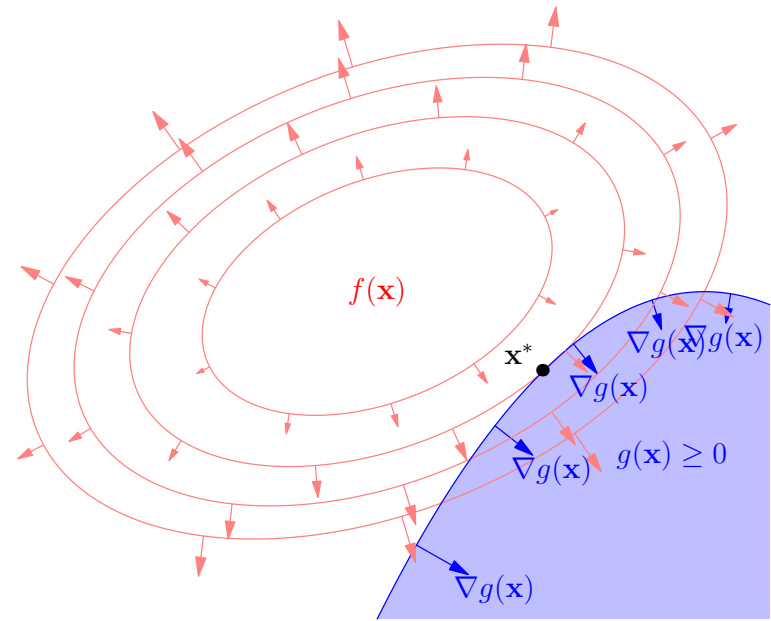
Constrained Optimisation



Lagrangians, Inequalities, KKT, Linear Programming, Quadratic Programming, Duality

Outline

1. **Constrained Optimisation**
2. Inequalities
3. Duality



Optimisation with Constraints

- There are a number of important applications where we wish to minimise an objective function subject to inequality constraints■
- A prominent example of this is support vector machines■
- More generally there are a large number of kernel models that involve constraints■
- However, constraints are ubiquitous in machine learning (e.g. in Wasserstein GANs)■

Solving Constrained Optimisation Problems

- Suppose we have a problem

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to } g(\boldsymbol{x}) = 0$$

- A standard procedure is to define the Lagrangian

$$\mathcal{L}(\boldsymbol{x}, \alpha) = f(\boldsymbol{x}) - \alpha g(\boldsymbol{x})$$

where α is known as a Lagrange multiplier

- In the extended space (\boldsymbol{x}, α) we have to solve

$$\max_{\alpha} \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \alpha)$$

Conditions on Optimum

- The optimisation problem is

$$\max_{\alpha} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \alpha) \quad \text{where} \quad \mathcal{L}(\mathbf{x}, \alpha) = f(\mathbf{x}) - \alpha g(\mathbf{x}) \blacksquare$$

- Assuming differentiability

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \alpha) = \nabla_{\mathbf{x}} f(\mathbf{x}) - \alpha \nabla_{\mathbf{x}} g(\mathbf{x}) = 0$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = -g(\mathbf{x}) = 0 \blacksquare$$

- The second condition is just the constraint \blacksquare
- But what about first condition: $\nabla_{\mathbf{x}} f(\mathbf{x}) = \alpha \nabla_{\mathbf{x}} g(\mathbf{x})?$ \blacksquare

Note on Gradients

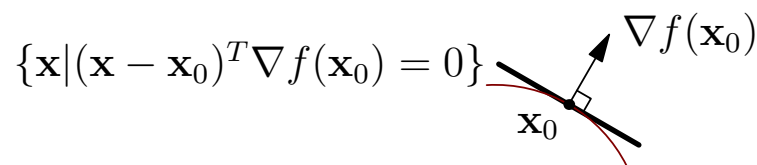
- Note that for any function $f(\mathbf{x})$ we can Taylor expand around \mathbf{x}_0

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top \nabla_{\mathbf{x}} f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top \mathbf{H}(\mathbf{x} - \mathbf{x}_0) + \dots$$

where \mathbf{H} is a matrix of second derivative known as the Hessian■

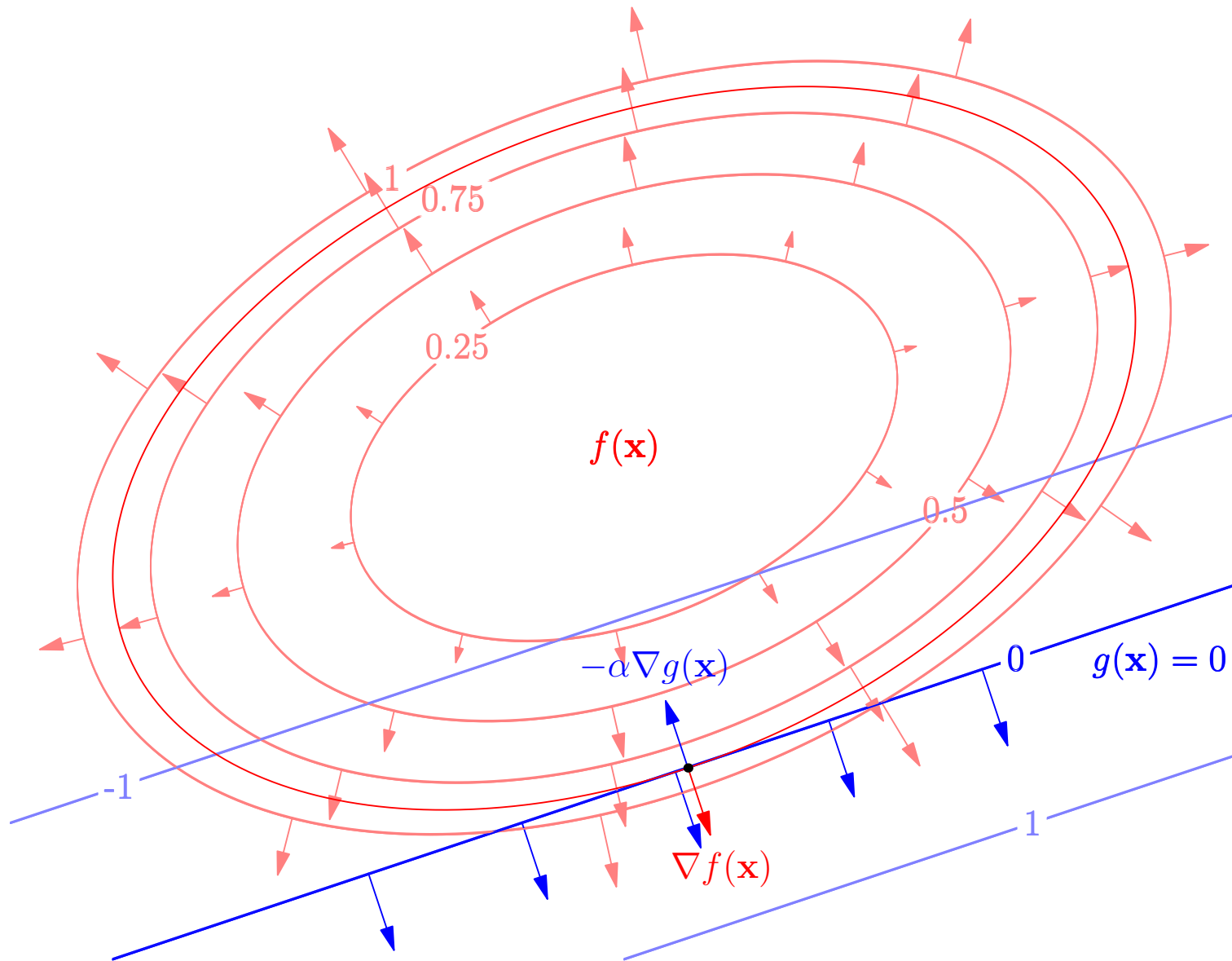
- If we consider the set of points perpendicular to $\nabla_{\mathbf{x}} f(\mathbf{x}_0)$ which go through \mathbf{x}_0 (the tangent plane), these will have values

$$f(\mathbf{x}) = f(\mathbf{x}_0) + O(\|\mathbf{x} - \mathbf{x}_0\|^2)$$



thus $\nabla_{\mathbf{x}} f(\mathbf{x})$ is always orthogonal to the contour lines■

Constrained Optima



Example

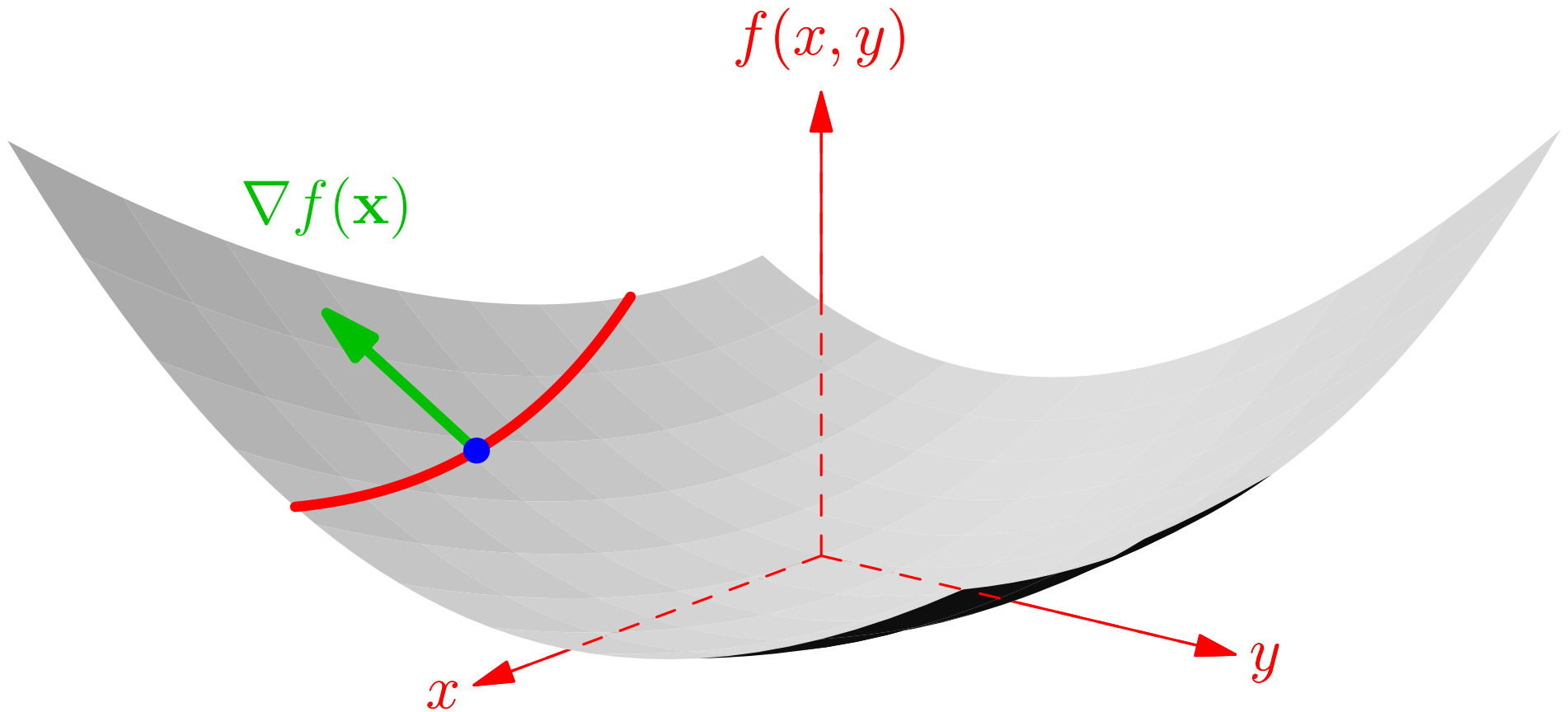
- Minimise $f(\mathbf{x}) = x^2 + 2y^2 - xy$
- Subject to $g(\mathbf{x}) = x - 2y - 3 = 0$ ■
- Writing $\mathcal{L} = f(\mathbf{x}) - \alpha g(\mathbf{x})$ ■
- Condition for minima is $\nabla_{\mathbf{x}}\mathcal{L} = 0$

$$\nabla_{\mathbf{x}}f(\mathbf{x}) = \begin{pmatrix} 2x - y \\ -x + 4y \end{pmatrix} = \alpha \nabla_{\mathbf{x}}g(\mathbf{x}) = \alpha \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

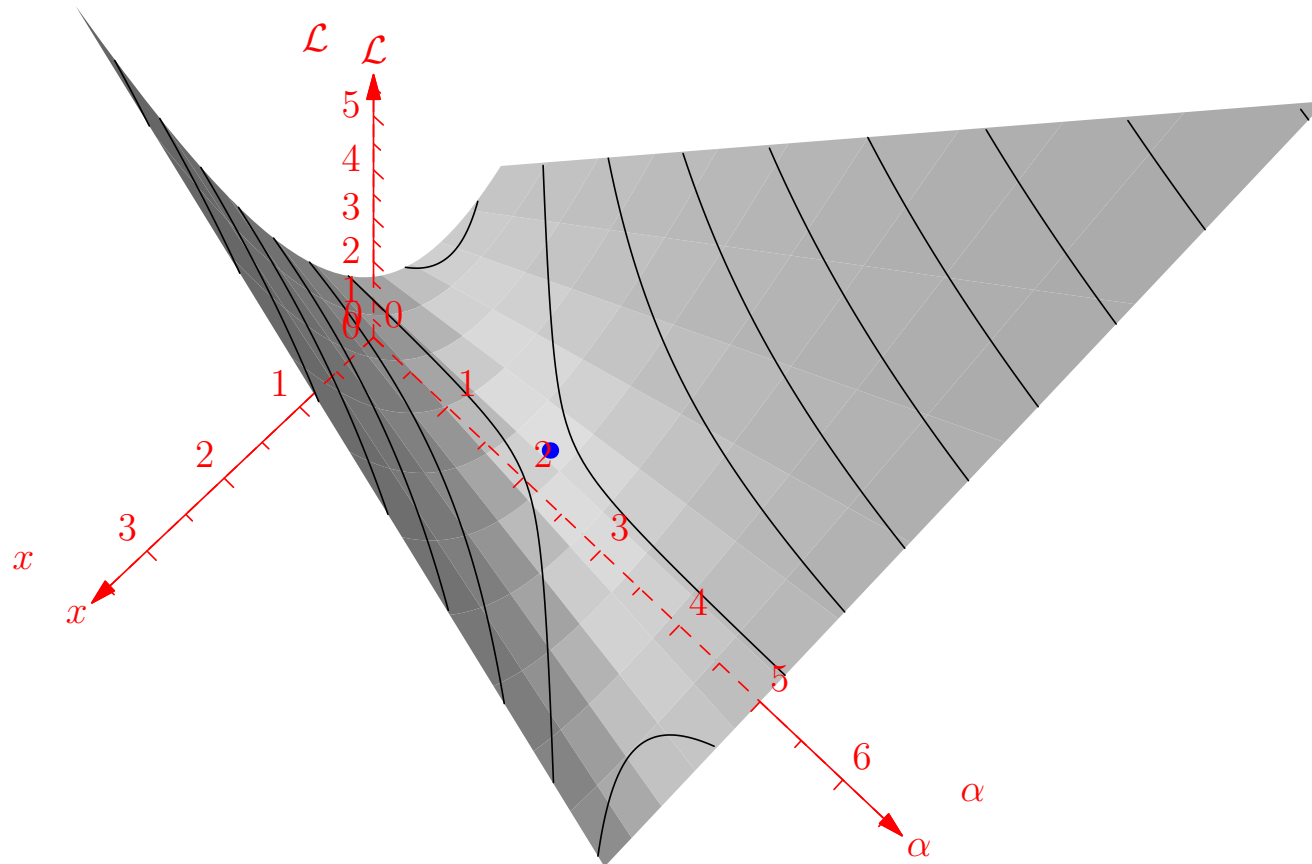
$$\text{and } \frac{\partial \mathcal{L}}{\partial \alpha} = -g(\mathbf{x}) = -x + 2y + 3 = 0$$
■

- Solving simultaneous equations gives minima at $(x, y) = (\frac{3}{4}, -\frac{9}{8})$ with $\alpha = \frac{21}{8}$ ■

Surface



Saddle-Point $y = -9/8$



Multiple Constraints

- Given an optimisation problem with multiple constraints

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to } g_k(\mathbf{x}) = 0 \text{ for } k = 1, 2, \dots, m$$

- We introduce multiple Lagrange multipliers

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}) = f(\mathbf{x}) - \sum_{k=1}^m \alpha_k g_k(\mathbf{x})$$

- The condition for an optima is $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}) = 0$ which implies

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \sum_{k=1}^m \alpha_k \nabla_{\mathbf{x}} g_k(\mathbf{x})$$

plus the original constraints $\frac{\partial \mathcal{L}(\mathbf{x}, \boldsymbol{\alpha})}{\partial \alpha_k} = -g_k(\mathbf{x}) = 0$

Example

- Minimise $f(\mathbf{x}) = x^2 + 2y^2 + 5z^2 - xy - xz$ subject to $g_1(\mathbf{x}) = x - 2y - z - 3 = 0$ and $g_2(\mathbf{x}) = 2x + 3y + z - 2 = 0$ ■
- Writing $\mathcal{L}(\mathbf{x}, \alpha) = f(\mathbf{x}) - \alpha_1 g_1(\mathbf{x}) - \alpha_2 g_2(\mathbf{x})$ ■
- Condition for minima is $\nabla_{\mathbf{x}} \mathcal{L} = 0$ or $\nabla_{\mathbf{x}} f(\mathbf{x}) = \sum_{k=1}^2 \alpha_k \nabla_{\mathbf{x}} g_k(\mathbf{x})$

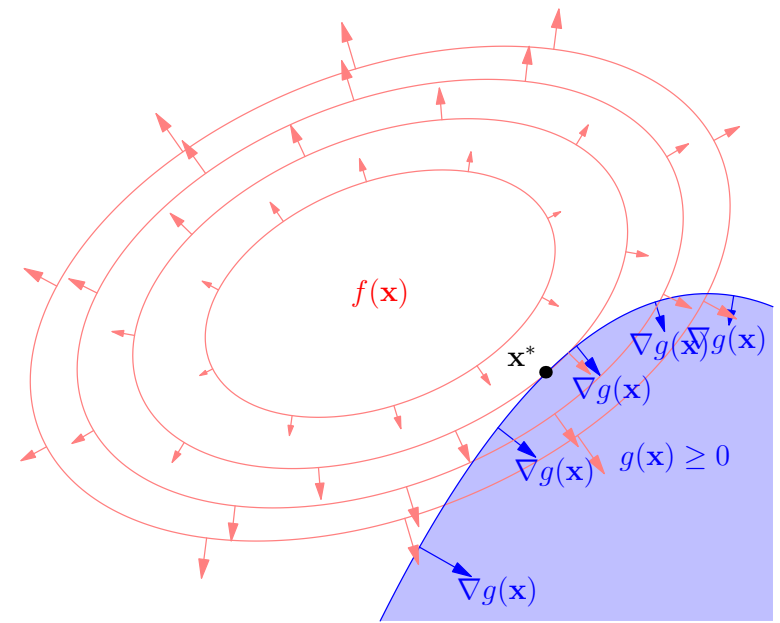
$$\begin{pmatrix} 2x - y - z \\ -x + 4y \\ 10z - x \end{pmatrix} = \alpha_1 \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$$

$$\text{and } \frac{\partial \mathcal{L}}{\partial \alpha_i} = -g_i(\mathbf{x}) = 0 \text{ ■}$$

- Solving simultaneous equations gives minima at $(\frac{37}{20}, -\frac{11}{20}, -\frac{1}{20})$ with $\alpha_1 = 3$ and $\alpha_2 = \frac{13}{20}$ ■

Outline

1. Constrained Optimisation
2. **Inequalities**
3. Duality



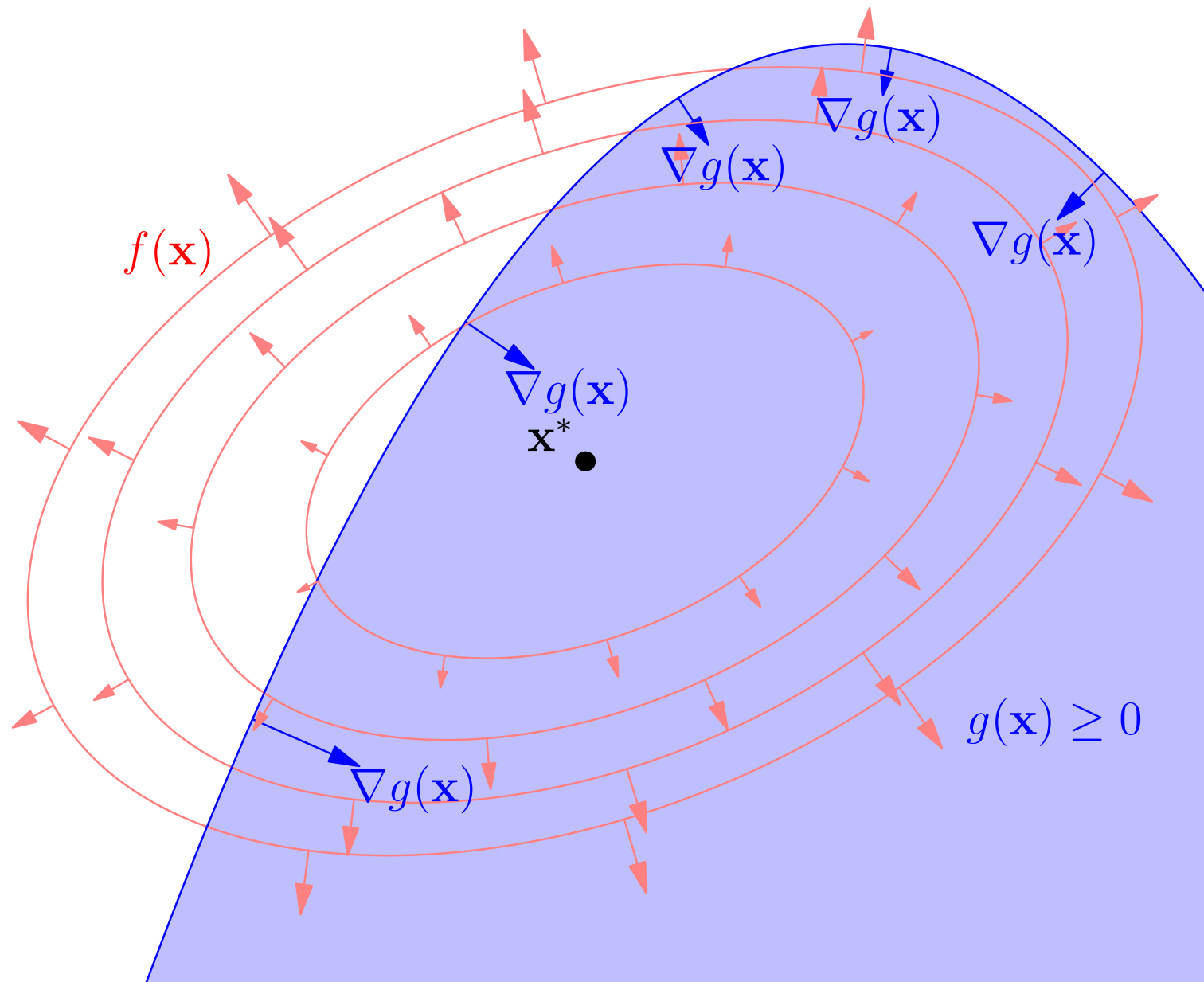
Inequality Constraints

- Suppose we have the problem

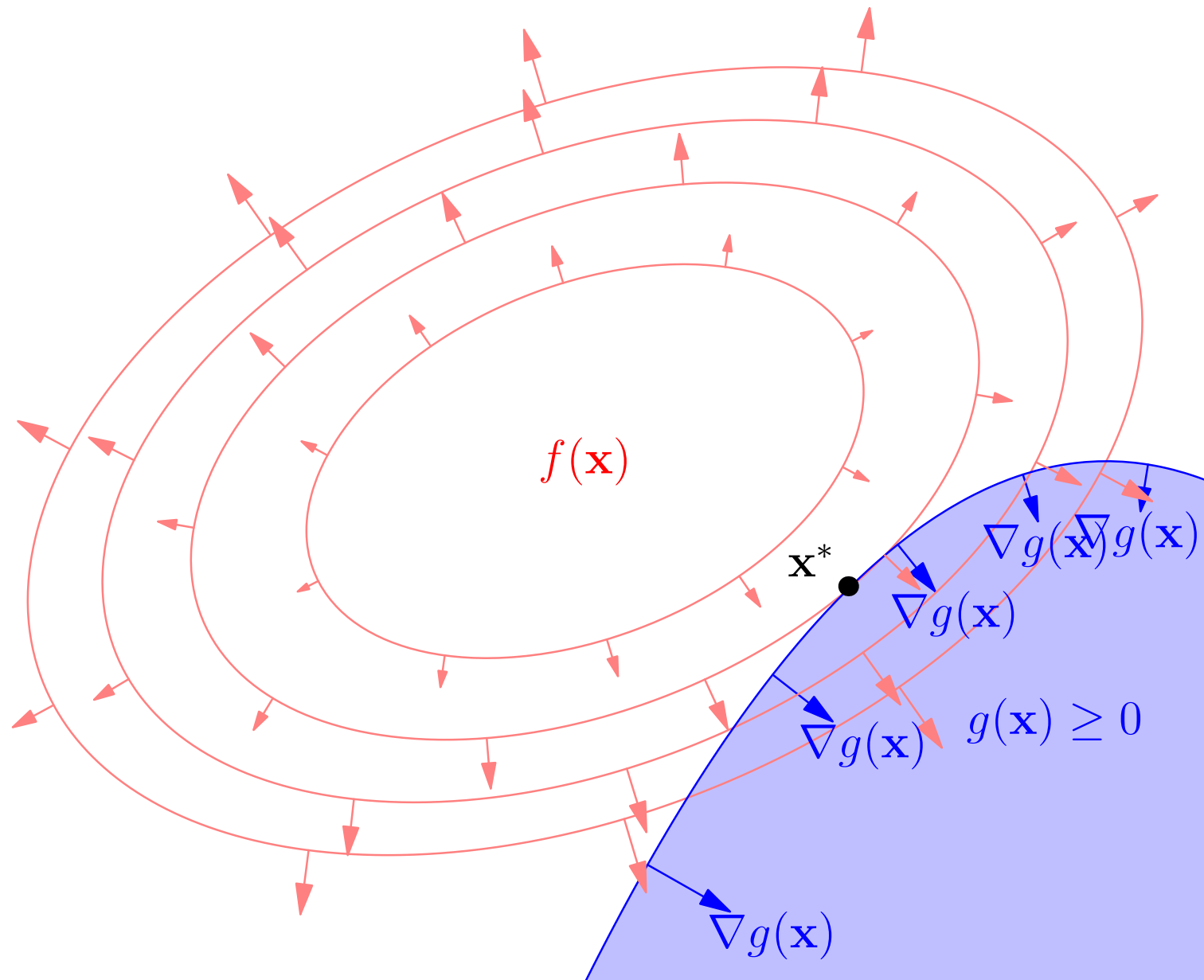
$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to } g(\boldsymbol{x}) \geq 0$$

- Looks much more complicated, but
- Only two things can happen
 - ★ Either a minimum, \boldsymbol{x}^* , of $f(\boldsymbol{x})$ satisfies $g(\boldsymbol{x}^*) > 0$
 - * We then have an unconstrained optimisation problem
 - ★ Otherwise, it satisfies $g(\boldsymbol{x}^*) = 0$
 - * We have a constrained optimisation problem

Inside Region



On the Boundary



KKT Conditions

- To minimise $f(\mathbf{x})$ subject to $g(\mathbf{x}) \geq 0$

$$\mathcal{L}(\mathbf{x}, \alpha) = f(\mathbf{x}) - \alpha g(\mathbf{x})$$

- Then $\nabla_{\mathbf{x}} \mathcal{L} = 0$ or

$$\nabla_{\mathbf{x}} \mathcal{L} = \nabla_{\mathbf{x}} f(\mathbf{x}) - \alpha \nabla_{\mathbf{x}} g(\mathbf{x}) = 0$$

- where either
 - ★ $\alpha = 0$ and the solutions in the interior
 - ★ $\alpha > 0$ and $g(\mathbf{x}) = 0$, i.e. the solution is on the boundary
- These conditions are known as the Karush-Kuhn-Tucker conditions

Many Inequalities

- Given the problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to } g_k(\mathbf{x}) \geq 0 \text{ for } k = 1, 2, \dots, m$$

- We introduce multiple Lagrange multipliers

$$\mathcal{L}(\mathbf{x}, \alpha) = f(\mathbf{x}) - \sum_{k=1}^m \alpha_k g_k(\mathbf{x})$$

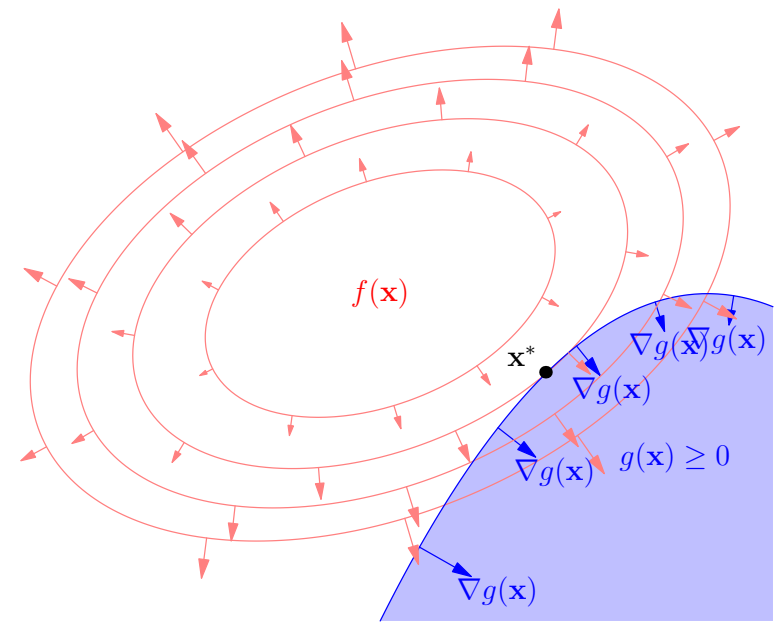
- The condition for an optima is

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \sum_{k=1}^m \alpha_k \nabla_{\mathbf{x}} g_k(\mathbf{x})$$

- Plus the constraints that either $\alpha_k = 0$ or $\alpha_k > 0$ and $g_k(\mathbf{x}) = 0$

Outline

1. Constrained Optimisation
2. Inequalities
3. **Duality**



Solving the Lagrangian for \mathbf{x}

- Consider minimising a function $f(\mathbf{x})$ subject to a set of constraints $g_i(\mathbf{x}) = 0$ or $g_i(\mathbf{x}) \leq 0$ ■
- We can consider this a double optimisation problem

$$\max_{\alpha} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \alpha) = \max_{\alpha} \min_{\mathbf{x}} \left(f(\mathbf{x}) + \sum_i \alpha_i g_i(\mathbf{x}) \right)$$

where there would be constraints on α_i if we had an inequality constraint■

Dual Problem

- If $f(\mathbf{x})$ and $g_i(\mathbf{x})$ are simple we can sometimes find a set of variables $\mathbf{x}^*(\boldsymbol{\alpha})$ that minimises the Lagrangian

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}) = 0 \blacksquare$$

- This leaves us with the **dual problem**

$$\max_{\boldsymbol{\alpha}} \mathcal{L}(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}) \blacksquare$$

- If we had an inequality constraint $g_i(\mathbf{x}) \geq 0$ then we would have the additional constraint in the dual problem $\alpha_i \geq 0 \blacksquare$

Linear Programming

- In linear programming we minimise a linear objective function $\mathbf{c}^\top \mathbf{x}$ subject to linear constraints $\mathbf{g}(\mathbf{x}) = \mathbf{M}\mathbf{x} - \mathbf{b} = 0$ (or $\mathbf{g}(\mathbf{x}) \geq 0$)■
- The Lagrangian becomes

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{c}^\top \mathbf{x} - \boldsymbol{\alpha}^\top (\mathbf{M}\mathbf{x} - \mathbf{b}) \blacksquare$$

- An equivalent way of writing the Lagrangian is

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{b}^\top \boldsymbol{\alpha} - \mathbf{x}^\top (\mathbf{M}^\top \boldsymbol{\alpha} - \mathbf{c}) \blacksquare$$

- An entirely equivalent interpretation is that we maximise an objective function $\mathbf{b}^\top \boldsymbol{\alpha}$ subject to constraints $\mathbf{M}^\top \boldsymbol{\alpha} - \mathbf{c} = 0$ (or $\mathbf{M}^\top \boldsymbol{\alpha} - \mathbf{c} \leq 0$)■

Linear Programming Example

- Suppose we eat potatoes and rice and we want to ensure that we get enough vitamin A and C

	Potatoes	Rice	Daily Requirement
Vitamin A	3	5	20
Vitamin C	5	2	24
Price	5	4	

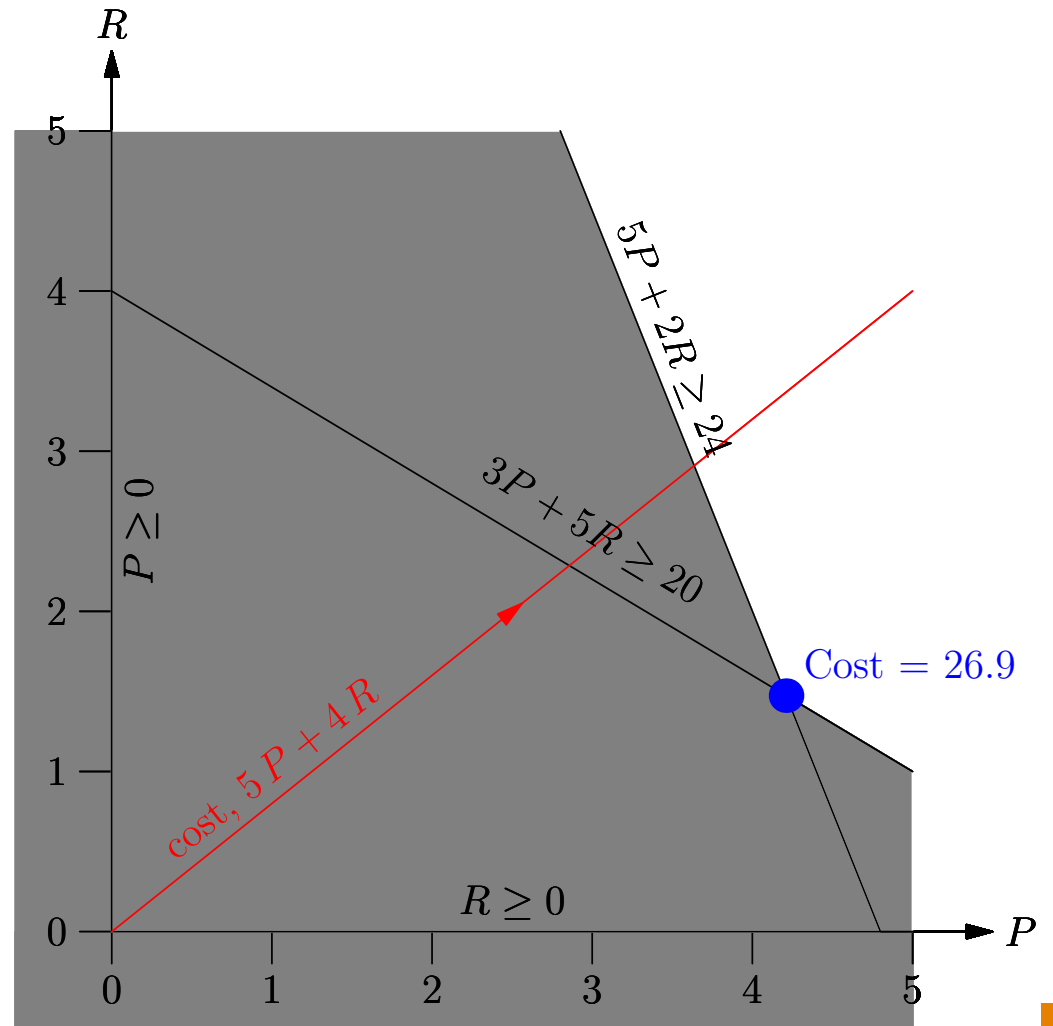
- We want to buy P kg potatoes and R kg of rice as cheaply as possible subject to fulfilling our vitamin requirement

$$\min_{P,R} 5P + 4R$$

$$\text{subject to } P, R \geq 0, \quad 3P + 5R \geq 20 \quad \text{and} \quad 5P + 2R \geq 24$$

Linear Programming

- Minimise $5P + 4R$
- Subject to
 - ★ $3P + 5R \geq 20$
 - ★ $5P + 2R \geq 24$
 - ★ $P, R \geq 0$



Lagrangian

- We can write the problem as a Lagrange problem

$$\min_{P,R} \max_{A,C} 5P + 4R - A(3P + 5R - 20) - C(5P + 2R - 24) \blacksquare$$

- subject to $P, R, A, B \geq 0$ \blacksquare
- A and C are Lagrange multipliers for vitamin A and C \blacksquare
- We can rearrange the Lagrangian to obtain

$$\max_{A,C} \min_{P,R} 20A + 24C - P(3A + 5C - 5) - R(5A + 2C - 4) \blacksquare$$

Dual Problem

- The Lagrangian

$$\max_{A,C} \min_{P,R} 20A + 24C - P(3A + 5C - 5) - R(5A + 2C - 4) \blacksquare$$

leads to the dual problem

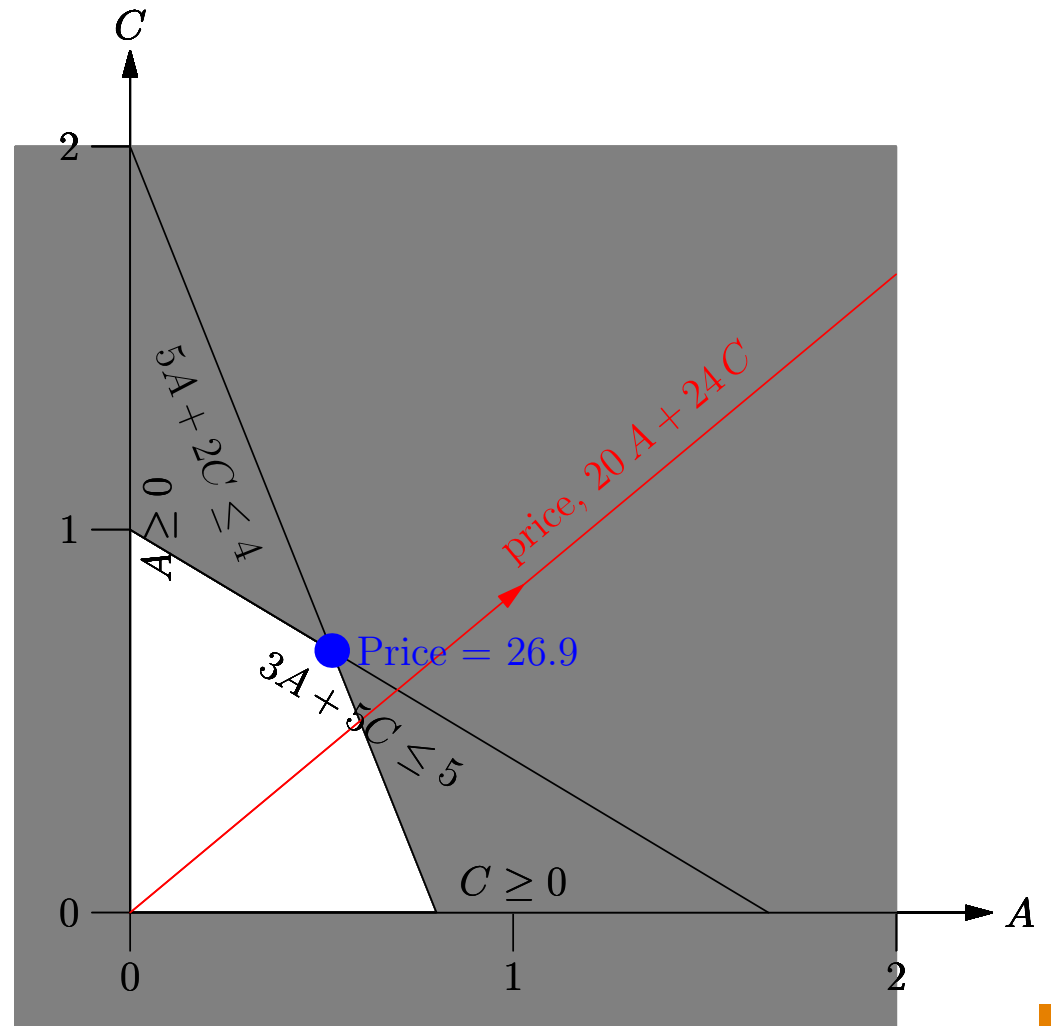
$$\max_{A,C} 20A + 24C$$

$$\text{subject to } 3A + 5C \leq 5 \quad 5A + 2C \leq 4 \quad A, C \geq 0 \blacksquare$$

- Consider someone selling vitamins A and C. They want to maximise the price of vitamins A and C, but their prices cannot exceed the price of the vitamins in potatoes or rice \blacksquare

Dual Linear Programme

- Maximise $20A + 24C$
- Subject to
 - ★ $3A + 5C \leq 5$
 - ★ $5A + 2C \leq 4$
 - ★ $A, C \geq 0$



Why?

- Why are we bothered about translating one linear programme into another?■
- Sometime one form is massively easier to solve than the other■
- This is because the first linear programme depends on the dimensionality of x while the second linear programme depends on the number of constraints (or dimensionality of α)■
- This is important, for example, in Wasserstein GANs■

Quadratic Programming

- A quadratic programme involves minimising a quadratic function $x^T Q x$ (with $Q \succ 0$) subject to linear constraints $Mx = b$ (or $Mx \leq b$)■
- We can define the Lagrangian

$$\mathcal{L}(x, \alpha) = x^T Q x - \alpha^T (Mx - b) \blacksquare$$

- Where the solution is given by $\max_{\alpha} \min_x \mathcal{L}(x, \alpha)$ ■
- If the constraints are inequality constraints then $\alpha_i \geq 0$ ■

Solution to Quadratic Programming Problem

- Using

$$\mathcal{L}(x, \alpha) = x^\top Q x - \alpha^\top (M x - b) \blacksquare$$

- Then

$$\nabla_x \mathcal{L}(x, \alpha) = 2Qx - M^\top \alpha \blacksquare$$

- So $\nabla_x \mathcal{L}(x, \alpha) = 0$ implies

$$x^* = \frac{1}{2} Q^{-1} M^\top \alpha \blacksquare$$

Dual Quadratic Programming Problem

- Substituting $x^* = \frac{1}{2}\mathbf{Q}^{-1}\mathbf{M}^\top\alpha$ into

$$\mathcal{L}(x, \alpha) = x^\top \mathbf{Q} x - \alpha^\top (\mathbf{M} x - b) \blacksquare$$

- We get the dual problem

$$\max_{\alpha} -\frac{1}{4}\alpha^\top \mathbf{M} \mathbf{Q}^{-1} \mathbf{M}^\top \alpha + \alpha^\top b \blacksquare$$

- If the constraints were inequality constraints then we have $\alpha_i \geq 0 \blacksquare$
- We have exchanged one quadratic programme for another, but sometimes that very useful (e.g. SVMs) \blacksquare

Lessons

- A useful tool for performing constrained optimisation is the introduction of Lagrange multipliers■
- This is particularly useful for problems with unique solutions (it will work when there are multiple solutions, but finding many saddle points is a pain)■
- For inequality constraints we need to satisfy KKT conditions■
- For simple situations (linear and quadratic programming) we can eliminate the original variables to obtain the dual problem■