Continuous Optimization 2 of 2

Constrained Optimization

We can extend our previous optimization discussion to one in which we now have constraints. Specifically

$$\min_{\mathbf{x}} f(\mathbf{x}) \tag{1}$$

$$\text{subject to } \; g_i(\mathbf{x}) \leq 0 \;\; \forall i = 1, \ldots m \tag{2}$$

where $g_i : \mathbb{R}^D \to \mathbb{R}$, for $i = 1, \ldots, m$, are our constraints.

- The question is how do we solve this type of problem?
- There are actually a number of way to tackle this, in optimization as a whole
  - We are however going to look at a specific classic approach, often referred to as a penalty method.

Constrained Optimization

The most harsh penalty would be to do the following

$$J(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{1}(g_i(\mathbf{x})) \tag{3}$$

where $\mathbf{1}(z)$ is an infinite step function

$$\mathbf{1}(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ \infty & \text{otherwise} \end{cases} \tag{4}$$

So basically, $J(\mathbf{x})$ is pushed infinity high if we violate a constraint.

- Why would J be hard to optimize?

Constrained Optimization

We can deal with the constraints in more nuanced approach, namely

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \tag{5}$$

$$= f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}) \tag{6}$$

we have concatenated all constraints $g_i(\mathbf{x})$ into a vector $\mathbf{g}(\mathbf{x})$, and all the Lagrange multipliers into a vector $\boldsymbol{\lambda} \in \mathbb{R}^m$.

- Is this easier to solve though?
- The challenge is now that we need to find both $\mathbf{x}$ and $\boldsymbol{\lambda}$.
  - Luckily, for a class of objective functions/constraints we can convert this problem and solve the converted problem.

Primal and Lagrangian Dual Problem

Primal and Lagrangian Dual

The problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \tag{7}$$

$$\text{subject to} \;\; g_i(\mathbf{x}) \leq 0 \;\; \forall i = 1, \ldots m \tag{8}$$

is the *primal problem*, corresponding to the *primal variables* x_i.
The associated *Lagrangian dual problem* is given by

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \mathcal{D}(\boldsymbol{\lambda}) \tag{9}$$

$$\text{subject to} \;\; \boldsymbol{\lambda} \geq \mathbf{0} \tag{10}$$

where $\boldsymbol{\lambda}$ are the dual variables and $\mathcal{D}(\boldsymbol{\lambda}) = \min_{\mathbf{x} \in \mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$

Primal and Lagrangian Dual Problem

In order to understand why it is worth trying to solve the dual problem we need to make couple of observations:

- The minimax inequality

$$\max_{\mathbf{y}} \min_{\mathbf{x}} \rho(\mathbf{x}, \mathbf{y}) \leq \min_{\mathbf{x}} \max_{\mathbf{y}} \rho(\mathbf{x}, \mathbf{y}) \tag{11}$$

we can build to this fact as follows, note that

$$\min_{\mathbf{x}} \rho(\mathbf{x}, \mathbf{y}) \leq \rho(\mathbf{x}, \mathbf{y}) \qquad \text{for all } \mathbf{x}, \mathbf{y} \tag{12}$$

$$\implies \min_{\mathbf{x}} \rho(\mathbf{x}, \mathbf{y}) \leq \max_{\mathbf{y}} \rho(\mathbf{x}, \mathbf{y}) \qquad \text{for all } \mathbf{x}, \mathbf{y} \tag{13}$$

$$\implies \max_{\mathbf{y}} \min_{\mathbf{x}} \rho(\mathbf{x}, \mathbf{y}) \leq \max_{\mathbf{y}} \rho(\mathbf{x}, \mathbf{y}) \qquad \text{for all } \mathbf{x}, \mathbf{y} \tag{14}$$

$$\implies \max_{\mathbf{y}} \min_{\mathbf{x}} \rho(\mathbf{x}, \mathbf{y}) \leq \min_{\mathbf{x}} \max_{\mathbf{y}} \rho(\mathbf{x}, \mathbf{y}) \qquad \text{for all } \mathbf{x}, \mathbf{y} \tag{15}$$

Primal and Lagrangian Dual Problem

The second concept is *weak duality*:

- Namely: the primal values are always greater than or equal to the dual values. $f(x)$

  $L(x)$
- Note that

$$f(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{1}(g_i(\mathbf{x})) = J(\mathbf{x}) = max_{\boldsymbol{\lambda} \geq \mathbf{0}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \tag{16}$$

$$= max_{\boldsymbol{\lambda} \geq \mathbf{0}} \left[f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \right] \tag{17}$$

If we are trying to solve the original problem using $J(x)$ we where looking for

$$min_{\mathbf{x} \in \mathbb{R}^d} J(x) = \overbrace{min_{\mathbf{x} \in \mathbb{R}^d} max_{\boldsymbol{\lambda} \geq \mathbf{0}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})}^{\text{Primal Problem}} \tag{18}$$

Primal and Lagrangian Dual Problem

$$min_{\mathbf{x}\in\mathbb{R}^d} J(x) = min_{\mathbf{x}\in\mathbb{R}^d} max_{\boldsymbol{\lambda}\geq\mathbf{0}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \tag{19}$$

now by applying the minimax inequality we see that

P   Dual Problem

$$min_{\mathbf{x}\in\mathbb{R}^d} max_{\boldsymbol{\lambda}\geq\mathbf{0}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \geq max_{\boldsymbol{\lambda}\geq\mathbf{0}} min_{\mathbf{x}\in\mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \tag{20}$$

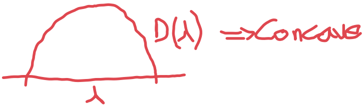
The right hand side is what we are solving in the dual problem, which is a lower bound of our primal problem

- Equation (20) represents *weak duality*
- If we had strict equality we would actually have *strong duality*
- The difference between the LHS and the RHS is called the *duality gap*

Primal and Lagrangian Dual Problem

The dual objective function, $\mathcal{D}(\boldsymbol{\lambda}) = \min_{\mathbf{x}\in\mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$, is an unconstrained optimization problem for a given value of $\boldsymbol{\lambda}$.

- If solving $min_{\mathbf{x}\in\mathbb{R}^d}\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$, for fixed $\boldsymbol{\lambda}$, is easy, then the overall problem is easy to solve.
  - Observe that $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ is affine with respect to $\boldsymbol{\lambda}$.
  - Therefore $\min_{\mathbf{x}\in\mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ is a pointwise minimum of affine functions of $\boldsymbol{\lambda}$,
    - ★ and hence $\mathcal{D}(\boldsymbol{\lambda})$ is concave even though $f(\cdot)$ and $g_i(\cdot)$ may be non-convex.
- Assuming $f(\cdot)$ and $g_i(\cdot)$ are differentiable and convex, we find the Lagrange dual problem by differentiating the Lagrangian with respect to $\mathbf{x}$, setting the differential to zero, and solving for the optimal value.



Dealing with Equality Constraints

In our original formulation we only had inequality constraints but we can extend our discussion to include them

$$\min_{\mathbf{x}} f(\mathbf{x}) \tag{21}$$

$$\text{subject to} \;\; g_i(\mathbf{x}) \leq 0 \;\; \forall i = 1, \ldots m \tag{22}$$

$$h_j(\mathbf{x}) = 0 \;\; \forall j = 1, \ldots n \tag{23}$$

The previous argument can be replicated with the inclusion of the h_js. You can also think of and equality constraint as $h_j(\mathbf{x}) \geq 0$ AND $h_j(\mathbf{x}) \leq 0$

- For many practical problems equality constraints are modeled as ϵ-inequalities.

Convex Optimization

Convex optimization is likely the best understood area of optimization

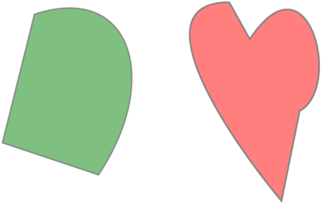
- It requires meaningful assumptions on both the objective function as well as the constraints.
- Provides us with global optimality as well efficient methods.

Convex set

A set $\mathcal{C}$ is a convex set if for any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ and any scalar $\theta \in [0, 1]$, we have

$$\theta\mathbf{x} + (1 - \theta)\mathbf{y} \in \mathcal{C} \tag{24}$$

Which of the below shaded regions are convex?



Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

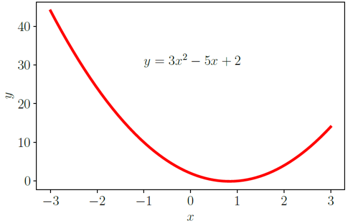
Convex Optimization

Convex function

Let function $f : \mathbb{R}^D \to \mathbb{R}$ be a function whose domain is a convex set. The function f is a *convex function* if for all $\mathbf{x}$, $\mathbf{y}$ in the domain of f, and for any scalar $\theta \in [0, 1]$, we have

$$f(\theta\mathbf{x} + (1 - \theta)\mathbf{y}) \le \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}) \tag{25}$$

- A concave function is the negative of a convex function.



Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

Convex Optimization: Differentiable Objective

If a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, we can specify convexity in terms of its gradient, $\nabla_{\mathbf{x}} f(\mathbf{x})$

Differentiable function:1st order criterion

A function $f(\mathbf{x})$ is convex if and only if for any two points $\mathbf{x}$, $\mathbf{y}$ it holds that:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla_{\mathbf{x}} f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \tag{26}$$

Twice differentiable function:2nd order criterion

A function $f(\mathbf{x})$ is convex if and only if $\nabla^2_{\mathbf{x}} f(\mathbf{x})$ is positive semi-definite

Convex function example

Consider the function $f(x) = xlog_2x$.

- This function is convex for all $x > 0$.
- See example 7.3 for a intuitive sense of why this function is convex, using the two possible tests (but example 7.3 is not a prove, as is stated)
- We will show it in general using the last mentioned test.

Convex function example

The easiest way is to use the fact that f is twice differentiable. First note that

$$f'(x) = 1 \cdot log_2 x + x \cdot \frac{1}{xln(2)} = log_2 x + \frac{1}{ln(2)} \tag{27}$$

$$f''(x) = \frac{1}{xln(2)} \tag{28}$$

in this case $\nabla^2_{\mathbf{x}} f(\mathbf{x})$ being positive semi-definite, is just $f''(x) > 0$. Which if we consider $x > 0$ follow readily

$$x > 0 \implies x \ln 2 > 0 \implies \frac{1}{xln2} > 0 \tag{29}$$

therefore f is convex.

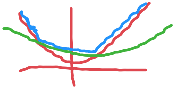
Convex functions

While we can use any of the three stated test (there are others equivalent ones) it can become rather tricky in practice.

- Luckily convex function are closed under certain operations.
    - Specifically if f_1 and f_2 are convex on A, then so is

$$\alpha f_1(\mathbf{x}) + \beta f_2(\mathbf{x}) \tag{30}$$

      for $\alpha, \beta \geq 0$. (This is called a conic combination)
- The application of an affine map preserves convexity
- Pointwise supreme preserves convexity $f(\mathbf{x}) = sup_{i \in \mathcal{I}} f_i(\mathbf{x})$ (over a family of convex functions)



Convex Optimization Problem

In summary, a constrained optimization problem is called a convex opti-
convex optimization problem if

$$\min_{\mathbf{x}} f(\mathbf{x}) \tag{31}$$

$$\text{subject to} \quad g_i(\mathbf{x}) \leq 0 \quad \forall i = 1, \ldots m \tag{32}$$

$$h_j(\mathbf{x}) = 0 \quad \forall j = 1, \ldots n \tag{33}$$

where all functions $f(\mathbf{x})$ and $g_i(\mathbf{x})$ are convex functions, and all $h_j(\mathbf{x}) = 0$ are convex sets.

Linear Programming

Linear Program

The following is a linear program of d linear variables and m linear constraints

$$\min_{\mathbf{x}\in\mathbb{R}^d} \mathbf{c}^T\mathbf{x} \tag{34}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{35}$$

where $\mathbf{A} \in \mathbb{R}^{m\times d}$ and $\mathbf{b} \in \mathbb{R}^m$

Linear Programming

The Lagrangian is given by

$f(x)$ $\quad Ax \leq b$ $\Downarrow$

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}^T\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) \tag{36}$$

$$= (\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{x} - \boldsymbol{\lambda}^T\mathbf{b} \tag{37}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$ is the vector of non-negative Lagrange multipliers. From here we can find the dual Lagrangian

$$\mathcal{D}(\boldsymbol{\lambda}) = min_{\mathbf{x}\in\mathbb{R}^d}\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \tag{38}$$

The min is found by using the gradient (remember we are dealing with convex functions):

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = (\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T = \mathbf{0}^T \tag{39}$$

It follows that $\mathcal{D}(\boldsymbol{\lambda}) = -\boldsymbol{\lambda}^T\mathbf{b}$

The dual optimization problem is therefore

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \; -\boldsymbol{\lambda}^T \mathbf{b} \tag{40}$$

$$\text{subject to} \quad \mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda} = \mathbf{0} \tag{41}$$

$$\boldsymbol{\lambda} \geq \mathbf{0} \tag{42}$$

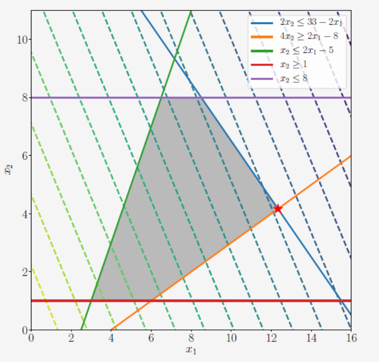
This is also a linear program, but with m variables. We have the choice of solving the primal or the dual program depending on whether m or d is larger

Linear Programming: Example

Consider the linear program

$$\min_{\mathbf{x}\in\mathbb{R}^2} \; -\begin{bmatrix}5\\3\end{bmatrix}^{\top}\begin{bmatrix}x_1\\x_2\end{bmatrix} \tag{43}$$

$$\text{subject to} \quad \begin{bmatrix}2 & 2\\ 2 & -4\\ -2 & 1\\ 0 & -1\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} \leq \begin{bmatrix}33\\8\\5\\-1\\8\end{bmatrix} \tag{44}$$



Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

Quadratic Programming

Quadratic program

The following is a *quadratic program* of d variables and m linear constraints.

$$\min_{\mathbf{x} \in \mathbb{R}^d} \frac{1}{2}\mathbf{x}^T \mathbf{Q}\mathbf{x} + \mathbf{c}^T \mathbf{x} \tag{45}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{46}$$

where $\mathbf{A} \in \mathbb{R}^{m \times d}$, $\mathbf{b} \in \mathbb{R}^m$, and $\mathbf{c} \in \mathbb{R}^d$ and $\mathbf{Q} \in \mathbb{R}^{d \times d}$ is positive definite[a] (so the objective function is convex)

[a] We consider the positive definite case, but it is not necessary to be that strict for convexity, but positive definite case means $\mathbf{Q}^{-1}$ exists

Quadratic Programming

We can construct the dual if Q is invertible. The Lagrangian is given by

$$\begin{aligned}
\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) &= \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{c}^T\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) && (47)\\
&= \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + (\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{x} - \boldsymbol{\lambda}^T\mathbf{b} && (48)
\end{aligned}$$

From here we can find the dual Lagrangian

$$\mathcal{D}(\boldsymbol{\lambda}) = min_{\mathbf{x}\in\mathbb{R}^d}\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \tag{49}$$

The min is found by using the gradient (remember we are dealing with convex functions):

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = (\mathbf{Q}\mathbf{x} + (\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}))^T = \mathbf{0}^T \tag{50}$$

Using the positive definiteness of $\mathbf{Q}$ we can solve for $\mathbf{x}$ as

$$\hat{\mathbf{x}} = -\mathbf{Q}^{-1}(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}) \tag{51}$$

It follows that

$$\mathcal{D}(\boldsymbol{\lambda}) = \frac{1}{2}\hat{\mathbf{x}}^T\mathbf{Q}\hat{\mathbf{x}} + (\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T\hat{\mathbf{x}} - \boldsymbol{\lambda}^T\mathbf{b} \tag{52}$$

$$= \frac{1}{2}\hat{\mathbf{x}}^T\mathbf{Q}\hat{\mathbf{x}} - (\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{Q}^{-1}(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}) - \boldsymbol{\lambda}^T\mathbf{b} \tag{53}$$

where

$$\frac{1}{2}\hat{\mathbf{x}}^T\mathbf{Q}\hat{\mathbf{x}} = \frac{1}{2}(\mathbf{Q}^{-1}(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}))^T\mathbf{Q}\mathbf{Q}^{-1}(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}) \tag{54}$$

$$= \frac{1}{2}(\mathbf{Q}^{-1}(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}))^T(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}) \tag{55}$$

$$= \frac{1}{2}(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T(\mathbf{Q}^{-1})^T(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}) \tag{56}$$

$$= \frac{1}{2}(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{Q}^{-1}(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}) \tag{57}$$

so

$$\mathcal{D}(\boldsymbol{\lambda}) = -\frac{1}{2}(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{Q}^{-1}(\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda}) - \boldsymbol{\lambda}^T\mathbf{b} \tag{58}$$

Quadratic Programming

Therefore, the dual optimization problem is given by

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \; -\frac{1}{2}(\mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda})^T \mathbf{Q}^{-1} (\mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda}) - \boldsymbol{\lambda}^T \mathbf{b} \tag{59}$$

$$\text{subject to} \;\; \boldsymbol{\lambda} \geq \mathbf{0} \tag{60}$$

Quadratic programming is the backbone of Support Vector Machines

Convex conjugate

Convex Conjugate

The *convex conjugate* of a function $f : \mathbb{R}^d \to \mathbb{R}$ is the function f^* defined by

$$f^*(\mathbf{s}) = \sup_{\mathbf{x}\in\mathbb{R}^D} (\langle \mathbf{s}, \mathbf{x} \rangle - f(\mathbf{x})) \tag{61}$$

$$= - \inf_{\mathbf{x}\in\mathbb{R}^D} (f(\mathbf{x}) + \langle \mathbf{s}, \mathbf{x} \rangle) \tag{62}$$

We will just use standard dot product between finite-dimensional vectors ($\langle \mathbf{s}, \mathbf{x} \rangle = \mathbf{s}^T \mathbf{x}$)

- f^* is a convex function, since it is the pointwise supremum of a family of convex (in this case, affine) functions of $\mathbf{s}$.
- If f is convex then the convex conjugate of f^* is once again f

Convex conjugate

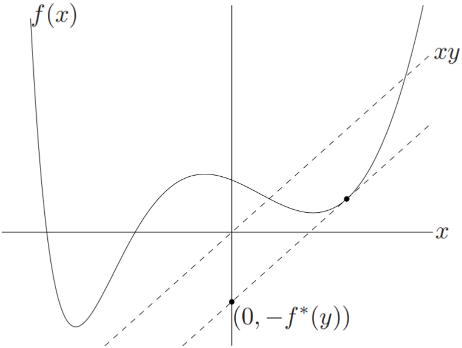


Figure 3.8 A function $f : \mathbf{R} \to \mathbf{R}$, and a value $y \in \mathbf{R}$. The conjugate function $f^*(y)$ is the maximum gap between the linear function yx and $f(x)$, as shown by the dashed line in the figure. If f is differentiable, this occurs at a point x where $f'(x) = y$.

Source: S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.

Convex conjugate: Example

Consider the quadratic function

$$f(\mathbf{y}) = \frac{\lambda}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} \tag{63}$$

based on a positive definite matrix $\mathbf{K}^{-1} \in \mathbb{R}^{n\times n}$ and the primal variable $\mathbf{y} \in \mathbb{R}^n$.

- The conjugate is then

$$f^*(\boldsymbol{\alpha}) = \sup_{\mathbf{y}\in\mathbb{R}^n}\left[\mathbf{y}^T\boldsymbol{\alpha} - \frac{\lambda}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y}\right] \tag{64}$$

  Since the function is differentiable and concave, we can find the maximum by taking the derivative and with respect to **y** setting it to zero.

Convex conjugate: Example

Specifically,

$$\frac{\partial\left[\mathbf{y}^T\boldsymbol{\alpha}-\frac{\lambda}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y}\right]}{\partial\mathbf{y}}=(\boldsymbol{\alpha}-\lambda\mathbf{K}^{-1}\mathbf{y})^T \tag{65}$$

and hence when the gradient is zero we have $\mathbf{y}=\frac{1}{\lambda}\mathbf{K}\boldsymbol{\alpha}$ into (64) yields

$$f^*(\boldsymbol{\alpha})=(\frac{1}{\lambda}\mathbf{K}\boldsymbol{\alpha})^T\boldsymbol{\alpha}-\frac{\lambda}{2}(\frac{1}{\lambda}\mathbf{K}\boldsymbol{\alpha})^T\mathbf{K}^{-1}\frac{1}{\lambda}\mathbf{K}\boldsymbol{\alpha} \tag{66}$$

$$=\frac{1}{\lambda}\boldsymbol{\alpha}^T\mathbf{K}\boldsymbol{\alpha}-\frac{1}{2\lambda}(\mathbf{K}\boldsymbol{\alpha})^T\boldsymbol{\alpha} \tag{67}$$

$$=\frac{1}{2\lambda}\boldsymbol{\alpha}^T\mathbf{K}\boldsymbol{\alpha} \tag{68}$$