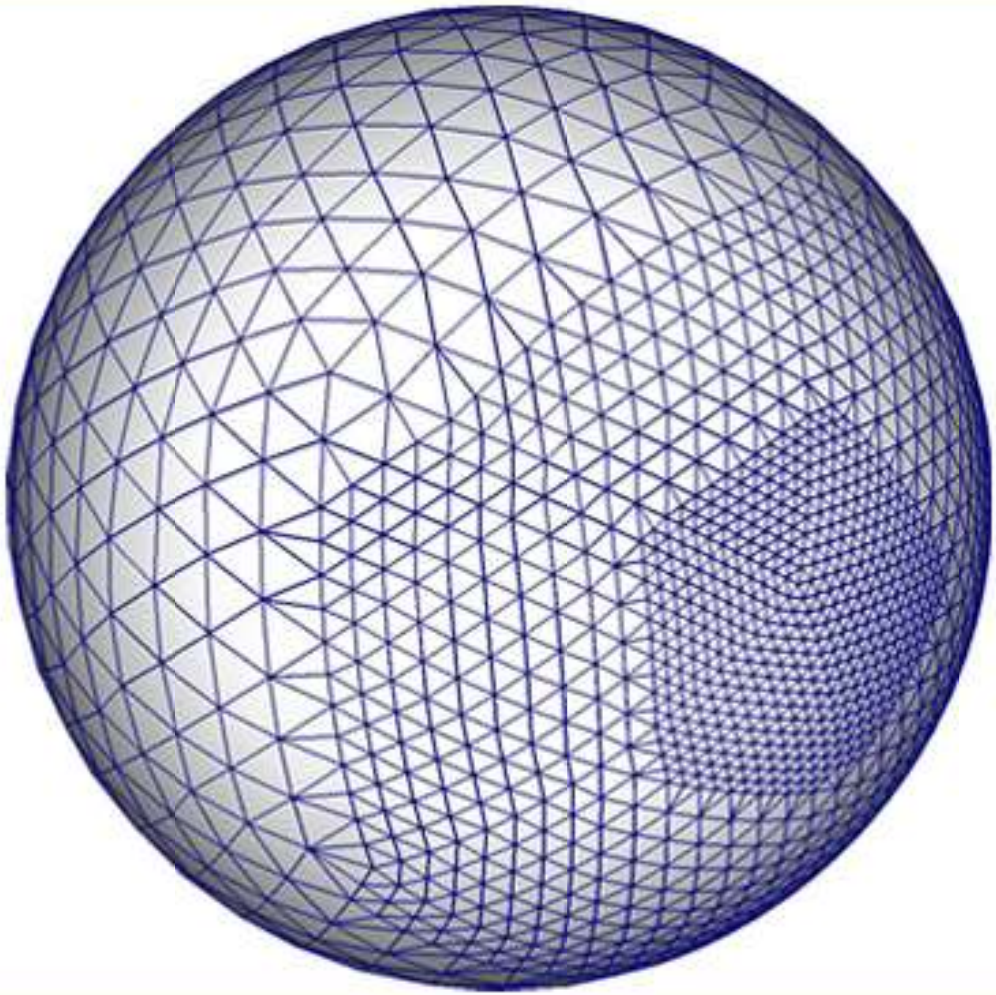
Engineering Ratios & Mathematical Tensors

(Concepts & Applications)



Mindi Harms

Anjali Burbank

First Edition, 2012

ISBN 978-81-323-0878-2

Published by:
Academic Studio
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1

Air-Fuel Ratio

Air-fuel ratio (**AFR**) is the mass ratio of air to fuel present during combustion. If exactly enough air is provided to completely burn all of the fuel, the ratio is known as the stoichiometric mixture (often abbreviated to **stoich**). AFR is an important measure for anti-pollution and performance tuning reasons. Lambda (λ) is an alternative way to represent AFR.

stoichiometric mixture fraction The relative amounts of oxygen enrichment and fuel dilution can be quantified by the stoichiometric mixture fraction, Zst, defined as Zst=(1+YF,0WOvO/YO,0WFvF)−1, where YF,0 and YO,0 represent the fuel and oxidizer mass fractions at the inlet, WF and WO are the species molecular weights, and vF and vO are the fuel and oxygen stoichiometric coefficients, respectively.

In industrial fired heaters, power plant steam generators, and large gas-fired turbines, the more common term is **percent excess combustion air**. For example, excess combustion air of 15 percent means that 15 percent more than the required stoichiometric air is being used.

A stoichiometric mixture is the working point that modern engine management systems employing fuel injection attempt to achieve in light load cruise situations. For gasoline fuel, the stoichiometric air/fuel mixture is approximately 14.7; i.e. the approximate mass of air is 14.7 mass of fuel. Any mixture less than 14.7 to 1 is considered to be a rich mixture, any more than 14.7 to 1 is a lean mixture - given perfect (ideal) "test" fuel (gasoline consisting of solely n-heptane and iso-octane). In reality, most fuels consist of a combination of heptane, octane, a handful of other alkanes, plus additives including detergents, and possibly oxygenators such as MTBE (methyl tert-butyl ether) or ethanol/methanol. These compounds all alter the stoichiometric ratio, with most of the additives pushing the ratio downward (oxygenators bring extra oxygen to the combustion event in liquid form that is released at time of combustions; for MTBE-laden fuel, a stoichiometric ratio can be as low as 14.1:1). Vehicles using an oxygen sensor(s) or other feedback-loop to control fuel to air ratios (usually by controlling fuel volume) will usually compensate automatically for this change in the fuel's stoichiometric rate by

measuring the exhaust gas composition, while vehicles without such controls (such as most motorcycles until recently, and cars predating the mid-1980s) may have difficulties running certain boutique blends of fuels (esp. winter fuels used in some areas) and may need to be rejetted (or otherwise have the fueling ratios altered) to compensate for special boutique fuel mixes. Vehicles using oxygen sensors enable the air-fuel ratio to be monitored by means of an air fuel ratio meter.

Synopsis

In theory a stoichiometric mixture has just enough air to completely burn the available fuel. In practice this is never quite achieved, due primarily to the very short time available in an internal combustion engine for each combustion cycle. Most of the combustion process completes in approximately 4-5 milliseconds at an engine speed of 6000 rpm. This is the time that elapses from when the spark is fired until the burning of the fuel air mix is essentially complete after some 80 degrees of crankshaft rotation.

Catalytic converters are designed to work best when the exhaust gases passing through them show nearly perfect combustion has taken place.

A stoichiometric mixture unfortunately burns very hot and can damage engine components if the engine is placed under high load at this fuel air mixture. Due to the high temperatures at this mixture, detonation of the fuel air mix shortly after maximum cylinder pressure is possible under high load (referred to as knocking or pinging). Detonation can cause serious engine damage as the uncontrolled burning of the fuel air mix can create very high pressures in the cylinder. As a consequence stoichiometric mixtures are only used under light load conditions. For acceleration and high load conditions, a richer mixture (lower air-fuel ratio) is used to produce cooler combustion products and thereby prevent detonation and overheating of the cylinder head.

In the typical air to natural gas combustion burner, a double cross limit strategy is employed to ensure ratio control. (This method was used in World War 2). The strategy involves adding the opposite flow feedback into the limiting control of the respective gas (air or fuel).This assures ratio control within an acceptable margin.

Other terms used

There are other terms commonly used when discussing the mixture of air and fuel in internal combustion engines.

Mixture

Mixture is the predominant word that appears in training texts, operation manuals and maintenance manuals in the aviation world.

AFR

The **Air fuel ratio** is the most common reference term used for mixtures in internal combustion engines.

$$AFR = \frac{m_{air}}{m_{fuel}}$$

It is the ratio between the *mass* of air and the mass of fuel in the fuel-air mix at any given moment.

For pure octane the stoichiometric mixture is approximately 14.7:1 or λ of 1.00 exactly.

In naturally aspirated engines powered by octane, maximum power is frequently reached at AFRs ranging from 12.5 - 13.3:1 or λ of 0.850 - 0.901.

FAR

Fuel Air ratio is commonly used in the gas turbine industry as well as in government studies of internal combustion engine and refers to the ratio of fuel to the air, it is 1/AFR.

Lambda

Most practical AFR devices actually measure the amount of residual oxygen (for lean mixes) or unburnt hydrocarbons (for rich mixtures) in the exhaust gas. Lambda (λ) is the ratio of actual AFR to stoichiometry for a given mixture. Lambda of 1.0 is at stoichiometry, rich mixtures are less than 1.0, and lean mixtures are greater than 1.0.

There is a direct relationship between lambda and AFR. To calculate AFR from a given lambda, multiply the measured lambda by the stoichiometric AFR for that fuel. Alternatively, to recover lambda from an AFR, divide AFR by the stoichiometric AFR for that fuel. This last equation is often used as the definition of lambda:

$$\lambda = \frac{AFR}{AFR_{stoich}}$$

Because the composition of common fuels varies seasonally, and because many modern vehicles can handle different fuels, when tuning, it makes more sense to talk about lambda values rather than AFR.

Equivalence ratio

The **equivalence ratio** of a system is defined as the ratio of the fuel-to-oxidizer ratio to the stoichiometric fuel-to-oxidizer ratio. Mathematically,

$$\phi = \frac{\text{fuel-to-oxidizer ratio}}{(\text{fuel-to-oxidizer ratio})_{st}} = \frac{m_{fuel}/m_{ox}}{(m_{fuel}/m_{ox})_{st}} = \frac{n_{fuel}/n_{ox}}{(n_{fuel}/n_{ox})_{st}}$$

where, *m* represents the mass, *n* represents number of moles, suffix *st* stands for stoichiometric conditions.

stoichiometric mixture fraction

$Zst = \lambda / (1 + \lambda)$

The advantage of using equivalence ratio over fuel-to-oxidizer ratio is that it does not have the same dependence as fuel-to-oxidizer ratio on the units being used. For example fuel-to-oxidizer ratio based on mass of fuel and oxidizer is not same as one define based on number of moles. This is not the case for equivalence ratio. The following example can help clarify the point. Consider a mixture of one mole of ethane (C_2H_6) and one mole of oxygen (O_2).

fuel-to-oxidizer ratio of this mixture based on the mass of fuel and air is

$$\frac{m_{C_2H_6}}{m_{O_2}} = \frac{1 \cdot (2 \cdot 12 + 6 \cdot 1)}{1 \cdot (2 \cdot 16)} = \frac{30}{32} = 0.938$$

fuel-to-oxidizer ratio of this mixture based on the number of moles of fuel and air is $\frac{n_{C_2H_6}}{n_{O_2}} = \frac{1}{1} = 1.$

Clearly the two values are not equal. To compare it to the equivalence ratio, we need to determine the fuel-to-oxidizer ratio of ethane and oxygen mixture. For this we need to consider the stoichiometric reaction of ethane and oxygen,

$$C_2H_6 + \tfrac{7}{2}O_2 \rightarrow 2CO_2 + 3H_2O$$

This gives,

$$(\text{fuel-to-oxidizer ratio based on mass})_{st} = (\frac{m_{C_2H_6}}{m_{O_2}})_{st} = \frac{1 \cdot (2 \cdot 12 + 6 \cdot 1)}{3.5 \cdot (2 \cdot 16)} = \frac{30}{112} = 0.268$$

$$(\text{fuel-to-oxidizer ratio based on number of moles})_{st} = (\frac{n_{C_2H_6}}{n_{O_2}})_{st} = \frac{1}{3.5} = 0.286$$

Thus we can determine the equivalence ratio of the give mixture as,

$$\phi = \frac{m_{C_2H_6}/m_{O_2}}{(m_{C_2H_6}/m_{O_2})_{st}} = \frac{0.938}{0.268} = 3.5$$

or equivalently as,

$$\phi = \frac{n_{C_2H_6}/n_{O_2}}{(n_{C_2H_6}/n_{O_2})_{st}} = \frac{1}{0.286} = 3.5$$

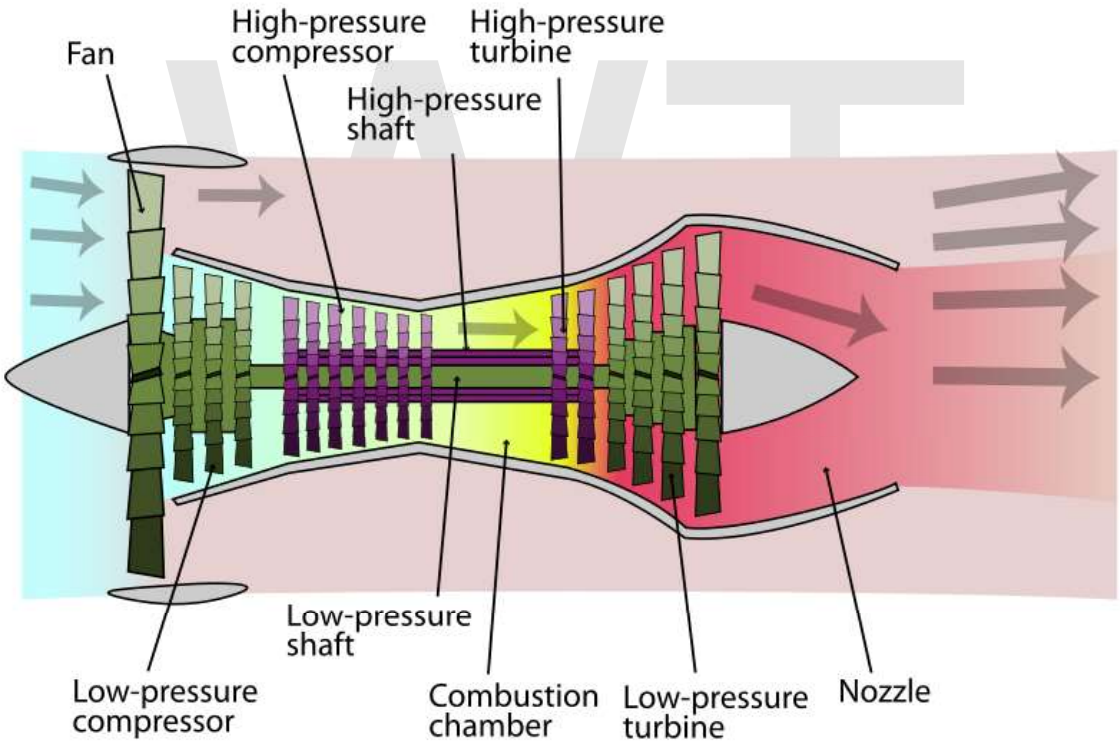
Another advantage of using the equivalence ratio is that ratios greater than one always represent excess fuel in the fuel-oxidizer mixture than would be required for complete combustion (stoichiometric reaction) irrespective of the fuel and oxidizer being used, while ratios less than one represent a deficiency of fuel or equivalently excess oxidizer in the mixture. This is not the case if one uses fuel-to-oxidizer ratio, which will take different values for different mixtures.

Equivalence ratio is related to λ (defined previously) as follows,

$$\phi = \frac{1}{\lambda}$$

Chapter 2

Bypass Ratio



Schematic turbofan engines; the high-bypass engine (lower) has a large fan that routes much air around the turbine, the low-bypass engine (upper) has a smaller fan routing more air into the turbine.
The bypass air is shown in pink, whilst the core gases are shown in red.

The term **bypass ratio (BPR)** relates to the design of turbofan engines, commonly used in aviation. It is defined as the ratio between the mass flow rate of air drawn in by the fan bypassing the engine core to the mass flow rate passing through the engine core.

A high bypass ratio gives a lower (actual) exhaust speed. This reduces the thrust specific fuel consumption, but reduces the top speed and gives a heavier engine.

A lower bypass ratio gives a higher exhaust speed, which is needed to sustain higher, usually supersonic, airspeeds. This increases the thrust specific fuel consumption.

In spite of this, it turns out that for jet engines in general, at optimum bypass ratios, the fuel burnt to travel any particular *distance* is largely independent of airspeed, but with supersonic jet engines being slightly more efficient in practice, at their design point.

Description

Jet engines are generally able to create considerably more energy than they can use in moving air through the engine core. This is because the limiting factor is the temperature at the turbine face, and that is a function of the total amount of fuel burned. Increasing airflow, and thus thrust, would imply burning more fuel and generating higher temperatures. It is possible to increase the airflow by burning "too much" fuel or adding water in front of the turbine to cool it, but both methods lead to incomplete combustion and very poor fuel efficiency. This was nevertheless commonly practiced in early jet engines because of a need to produce added thrust on takeoff. This is also why the exhaust plumes of older aircraft produce so much visible smoke (which is nothing more than unburned carbon from wasted jet fuel).

Rolls–Royce came up with a better use of the extra energy in their Conway turbofan engine, developed in the early 1950s. In the Conway, an otherwise normal axial-flow turbojet was equipped with an oversized first compressor stage (the one closest to the front of the engine), and centered inside a tubular nacelle (in effect, a ducted fan arrangement). While the inner portions of the compressor worked "as normal" and provided air into the core of the engine, the outer portion blew air around the engine to provide extra thrust. The Conway had a very small bypass ratio of only 0.3, but the improvement in fuel economy was notable; as a result, it and its derivatives like the Spey became some of the most popular jet engines in the world.

If the fan of a turbofan engine drives two kilograms of air around the engine for every kilogram that passes through the engine's core, the engine is said to have a bypass ratio of 2 (or 2 to 1). Higher bypass ratios generally give better Thrust specific fuel consumption as an increasing amount of thrust is being generated without burning more fuel. This is achieved since the engine propels a larger amount of air rearwards at slower speed, rather than a smaller amount of air at higher speed- because thrust is the momentum given to the air per second the thrust is the same. However energy is a square law on speed, and so it takes less energy to generate the same thrust; and hence less fuel is needed, the specific fuel consumption reduces.

Thus, with the example, for engines with the same thrust, the fuel efficiency would be improved by something less than 50%.

High bypass ratios are also correlated with lower noise, since the large flow of air surrounding the jet exhaust from the engine core helps to buffer the noise produced by the latter.

Through the 1960s the bypass ratios grew, making jetliners competitive in fuel terms with piston-powered planes for the first time. Most of the very-large engines in this class were pioneered in the United States by both Pratt & Whitney and General Electric, which for the first time was out-competing the United Kingdom in engine design. Rolls-Royce also started the development of the high-bypass turbofan, and although it caused considerable trouble at the time, the RB.211 would go on to become one of their most successful products.

Turbofans are typically broken into one of two categories: low–bypass and high–bypass ratio. In a low–bypass turbofan, only a small amount of air passes through the fan ducts and the fan is of very small diameter. The fan in a high–bypass turbofan is much larger to force a large volume of air through the ducts. The low–bypass turbofan is more compact, but the high-bypass turbofan can produce much greater thrust, is more fuel efficient, and is much quieter.

Today, almost all jet engines include some amount of bypass. Lower bypass ratios are appropriate at high speeds because the exhaust velocity must exceed the airspeed to give forward net thrust. For lower speed operations, such as airliners, modern engines use bypass ratios up to 17, while for higher speed operations such as fighter aircraft the ratios are much lower, around 1.5; and around 0.5 for sustained speeds around Mach 2 and somewhat above.

At transonic and supersonic speeds, very high bypass ratios still present engineering challenges.

Engine bypass ratios

Engine	Aircraft	Bypass ratio
Rolls-Royce/Snecma Olympus 593	Concorde (turbojet)	0:1
Rolls-Royce Tay	Gulfstream IV, Fokker 70, Fokker 100	3.1:1
SNECMA M88	Dassault Rafale	0.30:1
Pratt & Whitney JT8D	DC-9, MD-80, Boeing 727, Boeing 737 100 and 200 series	0.96:1
Pratt & Whitney F100	F-16, F-15	0.34:1
General Electric F404	F/A-18, T-50, F-117, X-29, X-31	0.34:1
Eurojet EJ200	Eurofighter Typhoon	0.4:1
Klimov RD-33	MiG-29, Il-102	0.49:1
Saturn AL-31F	Su-27, Su-30, Chengdu J-10	0.59:1

Kuznetsov NK-321	Tu-160	1.4:1
PowerJet SaM146	Sukhoi Superjet 100	4.43:1
Pratt & Whitney PW2000	Boeing 757, C-17 Globemaster III	5.9:1
Progress D-436	Yak-42M, Beriev Be-200, An-148	6.2:1
General Electric GEnx	Boeing 787	8.5:1
Rolls-Royce Trent 900	Airbus A380	8.7:1
General Electric GE90	Boeing 777	9:1
Rolls-Royce Trent 1000	Boeing 787	11:1

Chapter 3

Compression Ratio

The **compression ratio** of an internal-combustion engine or external combustion engine is a value that represents the ratio of the volume of its combustion chamber from its largest capacity to its smallest capacity. It is a fundamental specification for many common combustion engines.

In a piston engine it is the ratio between the volume of the cylinder and combustion chamber when the piston is at the bottom of its stroke, and the volume of the combustion chamber when the piston is at the top of its stroke.

Picture a cylinder and its combustion chamber with the piston at the bottom of its stroke containing 1000 cc of air (900 cc in the cylinder plus 100 cc in the combustion chamber). When the piston has moved up to the top of its stroke inside the cylinder, and the remaining volume inside the head or combustion chamber has been reduced to 100 cc, then the compression ratio would be proportionally described as 1000:100, or with fractional reduction, a 10:1 compression ratio.

A high compression ratio is desirable because it allows an engine to extract more mechanical energy from a given mass of air-fuel mixture due to its higher thermal efficiency. High ratios place the available oxygen and fuel molecules into a reduced space along with the adiabatic heat of compression–causing better mixing and evaporation of the fuel droplets. Thus they allow increased power at the moment of ignition and the extraction of more useful work from that power by expanding the hot gas to a greater degree.

Higher compression ratios will however make gasoline engines subject to engine knocking if lower octane-rated fuel is used, also known as detonation. This can reduce efficiency or damage the engine if knock sensors are not present to retard the timing. However, knock sensors have been a requirement of the OBD-II specification used in 1996 model year vehicles and newer.

Diesel engines on the other hand operate on the principle of compression ignition, so that a fuel which resists autoignition will cause late ignition which will also lead to engine knock.

Formula

The ratio is calculated by the following formula:

$$\mathrm{CR} = \frac{\frac{\pi}{4}b^2 s + V_c}{V_c}$$, where

b= cylinder bore (diameter)
s= piston stroke length
V_c= clearance volume. It is the volume of the combustion chamber (including head gasket). This is the minimum volume of the space at the end of the compression stroke, i.e. when the piston reaches top dead center (TDC). Because of the complex shape of this space, it is usually measured directly rather than calculated.

Typical compression ratios

Petrol (gasoline) engine

Due to pinging (detonation), the compression ratio in a gasoline or petrol-powered engine will usually not be much higher than 10:1, although some production automotive engines built for high-performance from 1955–1972 had compression ratios as high as 13.0:1, which could run safely on the high-octane leaded gasoline then available.

A technique used to prevent the onset of knock is the high "swirl" engine that forces the intake charge to adopt a very fast circular rotation in the cylinder during compression that provides quicker and more complete combustion. Recently, with the addition of variable valve timing and knock sensors to delay ignition timing, it is possible to manufacture gasoline engines with compression ratios of over 11:1 that can use 87 MON (octane rating) fuel.

In engines with a 'ping' or 'knock' sensor and an electronic control unit, the CR can be as high as 13:1 (2005 BMW K1200S). In 1981, Jaguar released a cylinder head that allowed up to 14:1 compression; but settled for 12.5:1 in production cars. The cylinder head design was known as the "May Fireball" head; it was developed by a Swiss engineer Michael May.

Petrol/gasoline engine with pressure-charging

In a turbocharged or supercharged gasoline engine, the CR is customarily built at 9.32:1 or lower.

Petrol/gasoline engine for racing

Motorcycle racing engines can use compression ratios as high as 14:1, and it is not uncommon to find motorcycles with compression ratios above 12.0:1 designed for 86 or 87 octane fuel.

Ethanol and methanol engines

Ethanol and methanol can take significantly higher compression ratios than gasoline. Racing engines burning methanol and ethanol fuel often incorporate a CR of 14.5-16:1, with F1 engines coming closer to 17:1(which is very critical for maximizing volumetric/fuel efficiency at around 18000 rpm)

Gas-fueled engine

In engines running exclusively on LPG or CNG, the CR may be higher, due to the higher octane rating of these fuels.

Diesel engine

In an auto-ignition diesel engine, (no electrical sparking plug—the hot air of compression lights the injected fuel) the CR will customarily exceed 14:1. Ratios over 22:1 are common. The appropriate compression ratio depends on the design of the cylinder head. The figure is usually between 14:1 and 16:1 for direct injection engines and between 18:1 and 23:1 for indirect injection engines.

Fault finding and diagnosis

Measuring the compression pressure of an engine, with a pressure gauge connected to the spark plug opening, gives an indication of the engine's state and quality. There is, however, no formula to calculate compression ratio based on cylinder pressure.

If the nominal compression ratio of an engine is given, the pre-ignition cylinder pressure can be estimated using the following relationship:

$$p = p_0 \times \mathrm{CR}^{\gamma}$$

where p_0 is the cylinder pressure at bottom dead center which is usually at 1 atm, CR is the compression ratio, and γ is the specific heat ratio for the working fluid, which is about 1.4 for air, and 1.3 for methane-air mixture.

For example, if an engine running on gasoline has a compression ratio of 10:1, the cylinder pressure at top dead center is

$$p_{\mathrm{TDC}} = 1 \text{ bar} \times 10^{1.4} = 25.1 \text{ bar}$$

This figure, however, will also depend on cam (i.e. valve) timing. Generally, cylinder pressure for common automotive designs should at least equal 10 bar, or, roughly estimated in pounds per square inch (psi) as between 15 and 20 times the compression ratio, or in this case between 150 psi and 200 psi, depending on cam timing. Purpose-built racing engines, stationary engines etc. will return figures outside this range.

Factors including late intake valve closure (relatively speaking for camshaft profiles outside of typical production car range, but not necessarily into the realm of competition engines) can produce a misleadingly low figure from this test. Excessive connecting rod clearance, combined with extremely high oil pump output (rare but not impossible) can sling enough oil to coat the cylinder walls with enough oil to facilitate reasonable piston ring seal artificially give a misleadingly high figure, on engines with compromised ring seal.

This can actually be used to some slight advantage. If a compression test does give a low figure, and it has been determined it is not due to intake valve closure/camshaft characteristics, then one can differentiate between the cause being valve/seat seal issues and ring seal by squirting engine oil into the spark plug orifice, in a quantity sufficient to disperse across the piston crown and the circumference of the top ring land, and thereby effect the mentioned seal. If a second compression test is performed shortly thereafter, and the new reading is much higher, it would be the ring seal that is problematic, whereas if the compression test pressure observed remains low, it is a valve sealing (or more rarely head gasket, or breakthrough piston or rarer still cylinder wall damage) issue.

If there is a significant (greater than 10%) difference between cylinders, that may be an indication that valves or cylinder head gaskets are leaking, piston rings are worn or that the block is cracked.

If a problem is suspected then a more comprehensive test using a leak-down tester can locate the leak.

Saab variable-compression engine

Because cylinder bore diameter, piston stroke length and combustion chamber volume are almost always constant, the compression ratio for a given engine is almost always constant, until engine wear takes its toll.

One exception is the experimental Saab Variable Compression engine (SVC). This engine, designed by Saab Automobile, uses a technique that dynamically alters the volume of the combustion chamber (V_c), which, via the above equation, changes the compression ratio (CR).

To alter V_c, the SVC 'lowers' the cylinder head closer to the crankshaft. It does this by replacing the typical one-part engine block with a two-part unit, with the crankshaft in the lower block and the cylinders in the upper portion. The two blocks are hinged together at one side (imagine a book, lying flat on a table, with the front cover held an inch or so

above the title page). By pivoting the upper block around the hinge point, the V_c (imagine the air between the front cover of the book and the title page) can be modified. In practice, the SVC adjusts the upper block through a small range of motion, using a hydraulic actuator.

Variable Compression Ratio (VCR) engines

The SAAB SVC is an advanced and workable addition to the world of VCR engines, the first being built and tested by Harry Ricardo in the 1920s. This work led to him devising the octane rating system that is still in use today. SAAB has recently been involved in working with the 'Office of Advanced Automotive Technologies', to produce a modern petrol VCR engine that showed an efficiency comparable with that of a Diesel. Many companies have been carrying out their own research in to VCR Engines, including Nissan, Volvo, PSA/Peugeot-Citroën and Renault but so far with no publicly demonstrated results.

The Atkinson cycle engine was one of the first attempts at variable compression. Since the compression ratio is the ratio between dynamic and static volumes of the combustion chamber the Atkinson cycle's method of increasing the length of the powerstroke compared to the intake stroke ultimately altered the compression ratio at different stages of the cycle.

Cortina Variable Compression engine

American inventor Paul Cortina is the latest entry into variable compression engine development. The Cortina engine is the only variable compression concept that slides the entire reciprocating assembly toward and away from the cylinder head. The advantage of this approach is that there is no complication to the workings of the reciprocating assembly itself, nor is there any complication outside of the engine. The way this is accomplished is by sliding the reciprocating assembly on a splined output shaft that is positioned perpendicular to a typical engine’s output orientation. The novel use of a barrel cam in place of a tradition crankshaft uses fewer parts than competing designs and adds the option of a true Atkinson cycle. It also has a more favorable leverage curve than a traditional crankshaft. Since the connecting rods stay straight, with a barrel cam cycling the pistons, it also is possible to box in the lower cylinder for use as an internal supercharger.

Dynamic compression ratio

The calculated compression ratio, as given above, presumes that the cylinder is sealed at the bottom of the stroke, and that the volume compressed is the actual volume.

However: intake valve closure (sealing the cylinder) always takes place after BDC, which may cause some of the intake charge to be compressed backwards out of the cylinder by the rising piston at very low speeds; only the percentage of the stroke after intake valve closure is compressed. Intake port tuning and scavenging may allow a greater mass of

charge (at a higher than atmospheric pressure) to be trapped in the cylinder than the static volume would suggest (This "corrected" compression ratio is commonly called the "*dynamic compression ratio*".

This ratio is higher with more conservative (i.e., earlier, soon after BDC) intake cam timing, and lower with more radical (i.e., later, long after BDC) intake cam timing, but always lower than the static or "nominal" compression ratio.

The actual position of the piston can be determined by trigonometry, using the stroke length and the connecting rod length (measured between centers). The absolute cylinder pressure is the result of an exponent of the dynamic compression ratio. This exponent is a polytropic value for the ratio of variable heats for air and similar gases at the temperatures present. This compensates for the temperature rise caused by compression, as well as heat lost to the cylinder. Under ideal (adiabatic) conditions, the exponent would be 1.4, but a lower value, generally between 1.2 and 1.3 is used, since the amount of heat lost will vary among engines based on design, size and materials used, but provides useful results for purposes of comparison. For example, if the static compression ratio is 10:1, and the dynamic compression ratio is 7.5:1, a useful value for cylinder pressure would be (7.5)^1.3 × atmospheric pressure, or 13.7 bar. (× 14.7 psi at sea level = 201.8 psi. The pressure shown on a gauge would be the absolute pressure less atmospheric pressure, or 187.1 psi.)

The two corrections for dynamic compression ratio affect cylinder pressure in opposite directions, but not in equal strength. An engine with high static compression ratio and late intake valve closure will have a DCR similar to an engine with lower compression but earlier intake valve closure.

Additionally, the cylinder pressure developed when an engine is running will be higher than that shown in a compression test for several reasons.

- The much higher velocity of a piston when an engine is running versus cranking allows less time for pressure to bleed past the piston rings into the crankcase.
- a running engine is coating the cylinder walls with much more oil than an engine that is being cranked at low RPM, which helps the seal.
- the higher temperature of the cylinder will create higher pressures when running vs. a static test, even a test performed with the engine near operating temperature.
- A running engine does not stop taking air & fuel into the cylinder when the piston reaches BDC; The mixture that is rushing into the cylinder during the downstroke develops momentum and continues briefly after the vacuum ceases (in the same respect that rapidly opening a door will create a draft that continues after movement of the door ceases). This is called scavenging. Intake tuning, cylinder head design, valve timing and exhaust tuning determine how effectively an engine scavenges.

Compression ratio versus overall pressure ratio

Compression ratio and overall pressure ratio are interrelated as follows:

Compression ratio	2:1	3:1	5:1	10:1	15:1	20:1	25:1	35:1
Pressure ratio	2.64:1	4.66:1	9.52:1	25.12:1	44.31:1	66.29:1	90.60:1	145.11:1

The reason for this difference is that compression ratio is defined via the volume reduction:

$$\mathrm{CR} = \frac{V_1}{V_2},$$

while pressure ratio is defined as the pressure increase:

$$\mathrm{PR} = \frac{P_2}{P_1}.$$

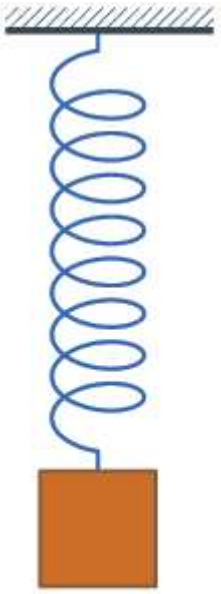
In calculating the pressure ratio, we assume that an adiabatic compression is carried out (i.e. that no heat energy is supplied to the gas being compressed, and that any temperature rise is solely due to the compression). We also assume that air is a perfect gas. With those two assumptions we can define the relationship between change of volume and change of pressure as follows:

$$P_1 V_1^\gamma = P_2 V_2^\gamma \Rightarrow \frac{P_2}{P_1} = \left(\frac{V_1}{V_2}\right)^\gamma$$

where γ is the ratio of specific heats for air (approximately 1.4). The values in the table above are derived using this formula. Note that in reality the ratio of specific heats changes with temperature and that significant deviations from adiabatic behavior will occur.

Chapter 4

Damping Ratio



Underdamped spring-mass system with $\zeta<1$

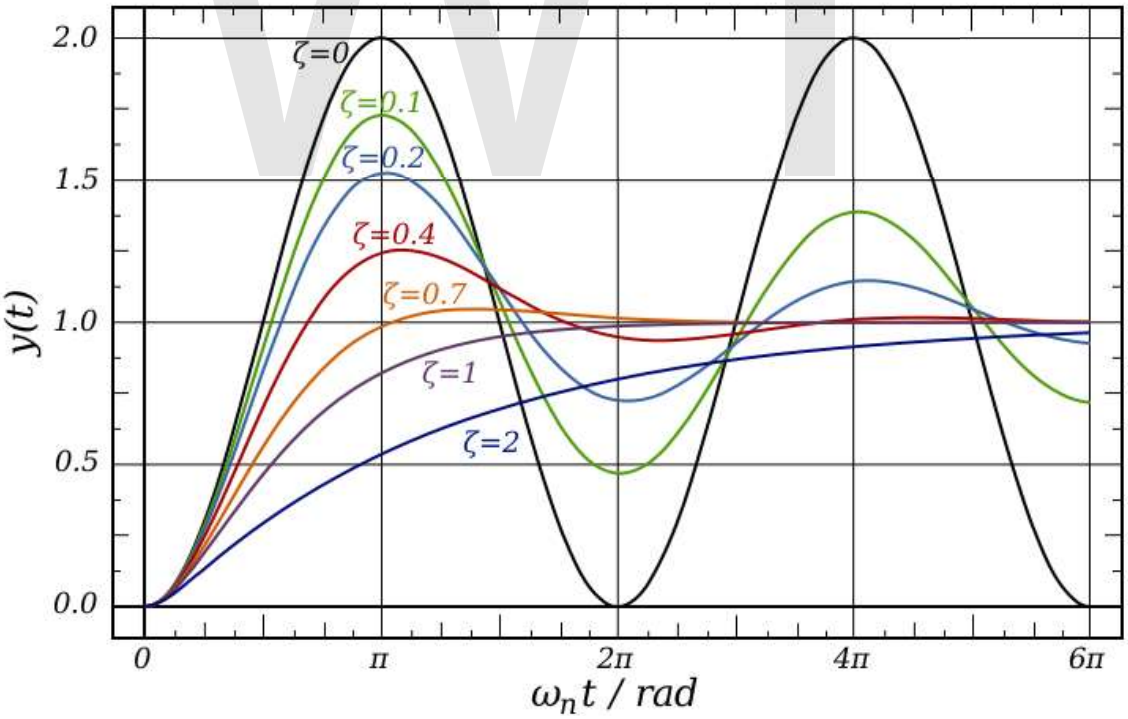
In engineering, the **damping ratio** is a dimensionless measure describing how oscillations in a system decay after a disturbance. Many systems exhibit oscillatory behavior when they are disturbed from their position of static equilibrium. A mass suspended from a spring, for example, will, if pulled and released, bounce up and down. On each bounce, the system is trying to return to its equilibrium position, but overshoots it. Frictional losses damp the system and cause the oscillations to gradually decay in amplitude towards zero. The damping ratio is a measure of describing how rapidly the oscillations decay from one bounce to the next.

The behaviour of oscillating systems is often of interest in a diverse range of disciplines that include control engineering, mechanical engineering and electrical engineering. The physical quantity that is oscillating varies greatly, and could be the swaying of a tall building in the wind, or the speed of an electric motor, but a normalised, or non-dimensionalised approach can be convenient in describing common aspects of behavior.

Oscillation modes

- Were the spring-mass system completely lossless, the mass would oscillate indefinitely, each bounce of equal height to the last. This hypothetical case is called **undamped**.
- If the system contained high losses, for example if the spring-mass experiment were conducted in a viscous fluid, the mass could slowly return to its rest position without ever overshooting. This case is called **overdamped**.
- Commonly, the mass tends to overshoot its starting position, and then return, overshooting again. With each overshoot, some energy in the system is dissipated, and the oscillations die towards zero. This case is called **underdamped**.
- Between the overdamped and underdamped cases, there exists a certain level of damping at which the system will just fail to overshoot and will not make a single oscillation. This case is called **critical damping**.

Definition



The effect of varying damping ratio on a second-order system.

The **damping ratio** is a parameter, usually denoted by ζ (zeta), that characterizes the frequency response of a second order ordinary differential equation. It is particularly important in the study of control theory. It is also important in the harmonic oscillator.

The damping ratio provides a mathematical means of expressing the level of damping in a system relative to critical damping. For a damped harmonic oscillator with mass *m*, damping coefficient *c*, and spring constant *k*, it can be defined as the ratio of the damping coefficient in the system's differential equation to the critical damping coefficient:

$$\zeta = \frac{c}{c_c}.$$

where the system differential equation is

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = 0.$$

and the corresponding critical damping coefficient is

$$c_c = 2\sqrt{km}$$

The damping ratio is dimensionless, being the ratio of two coefficients of identical units.

The damping ratio is also related to the logarithmic decrement δ for underdamped vibrations via the relation

$$\zeta = \frac{\delta}{\sqrt{(2\pi)^2 + \delta^2}} \qquad \text{where} \qquad \delta \triangleq \ln\frac{x_1}{x_2}.$$

This relation is only meaningful for underdamped systems because the logarithmic decrement is defined as the natural log of the ratio of any two successive amplitudes, and only underdamped systems exhibit oscillation.

Derivation of the damping ratio

Using the natural frequency of the simple harmonic oscillator $\omega_0 = \sqrt{k/m}$ and the definition of the damping ratio above, we can rewrite this as:

$$\frac{d^2x}{dt^2} + 2\zeta\omega_0\frac{dx}{dt} + \omega_0^2 x = 0.$$

This equation can be solved with the ansatz

$$x(t) = Ce^{st},$$

where *C* and *s* are both complex constants. That ansatz assumes a solution that is oscillatory and/or decaying exponentially. Using it in the ODE gives a condition on the frequency of the damped oscillations,

$$s = -\omega_0(\zeta \pm \sqrt{\zeta^2 - 1}).$$

- **Overdamped**:If *s* is a real number, then the solution is simply a decaying exponential with no oscillation. This case occurs for $\zeta > 1$, and is referred to as **overdamped**.

- **Underdamped**:If *s* is a complex number, then the solution is a decaying exponential combined with an oscillatory portion that looks like $\exp(i\omega_0\sqrt{1-\zeta^2})$. This case occurs for $\zeta < 1$, and is referred to as **underdamped**. (The case where $\zeta \to 0$ corresponds to the undamped simple harmonic oscillator, and in that case the solution looks like $\exp(i\omega_0)$, as expected.)

- **Critically damped**:The case where $\zeta = 1$ is the border between the overdamped and underdamped cases, and is referred to as **critically damped**. This turns out to be a desirable outcome in many cases where engineering design of a damped oscillator is required (e.g., a door closing mechanism).

Q factor and decay rate

The factors *Q*, damping ratio ζ, and exponential decay rate α are related such that

$$\zeta = \frac{1}{2Q} = \frac{\alpha}{\omega_0}.$$

When a second-order system has $\zeta < 1$ (that is, when the system is underdamped), it has two complex conjugate poles that each have a real part of α; that is, the decay rate parameter α represents the rate of exponential decay of the oscillations. A lower damping ratio implies a lower decay rate, and so very underdamped systems oscillate for long times. For example, a high quality tuning fork, which has a very low damping ratio, has an oscillation that lasts a long time, decaying very slowly after being struck by a hammer.

Chapter 5

Gear Ratio

Gears on a piece of farm equipment, total (3 gears) gear ratio 42/13 = 3.23

The **gear ratio** is the relationship between the numbers of teeth on two gears that are meshed or two sprockets connected with a common roller chain, or the circumferences of two pulleys connected with a drive belt.

General description

The input or driver gear in a gear train is the gear directly connected to the motor or other power source. Thus the driver is the gear that transmits power to the other gears in the gear train. In a simple 2-gear system, the second gear (the gear which is *turned by* the driver) is called the output or driven gear. In a gear train consisting of more than 2 gears, the final gear (the gear connected to a wheel axle or other rotating mechanical component) is the output gear.

gear ratio (gr) = (number of teeth on output or driven gear)/(number of teeth on input or driver gear)

If we assume that in the photo the smallest gear is connected to the motor, then it is the driver gear. The somewhat larger gear on the upper left is called an idler gear -- it is not connected directly to either the motor or the output shaft and serves only to transmit power between the input and output gears. There is a third gear in the upper-right corner of the photo. If we assume that gear is connected to the machine's output shaft, it is the output or driven gear.

The idler gear in this particular gear train has 21 teeth and the input gear has 13. *Considering for the moment only those two gears*, we can regard the idler as the driven gear. Therefore, the gear ratio is driven/driver = 21/13 = ~1.62 or 1.62:1.

The ratio means that the driver gear must make 1.62 revolutions to turn the driven gear 1 revolution. It also means that for every one revolution of the driver, the driven gear has made 1/1.62, or 0.62, revolutions. In practical terms, the larger gear turns more slowly.

Now suppose the third gear in the picture has 42 teeth. The gear ratio between the idler and third gear is thus 42/21, or 2:1, and hence the final gear ratio is 1.62x2=~3.23. For every 3.23 revolutions of the smallest gear, the largest gear turns one revolution, or for every one revolution of the smallest gear, the largest gear turns 0.31 (1/3.23) revolution, a total reduction of about 1:3.23 (Gear Reduction Ratio (GRR) = 1/Gear Ratio (GR)).

Since the intermediate (idler) gear contacts directly both the smaller and the larger gear it can be removed from the calculation, also giving a ratio of 42/13 = ~3.23.

Since the number of teeth is also proportional to the circumference of the gear wheel (the bigger the wheel the more teeth it has) the gear ratio can also be expressed as the relationship between the pitch circles of both wheels (where d is the pitch diameter of the input wheel and D is the pitch diameter of the output wheel):

$$gr = \frac{\pi D}{\pi d} = \frac{D}{d}$$

Pitch circles have diameters that would give the same gear ratio, but with cylindrical surfaces that do not slip.

Since the diameter is equal to twice the radius;

$$gr = \frac{D}{d} = \frac{2R}{2r} = \frac{R}{r}$$

as well.

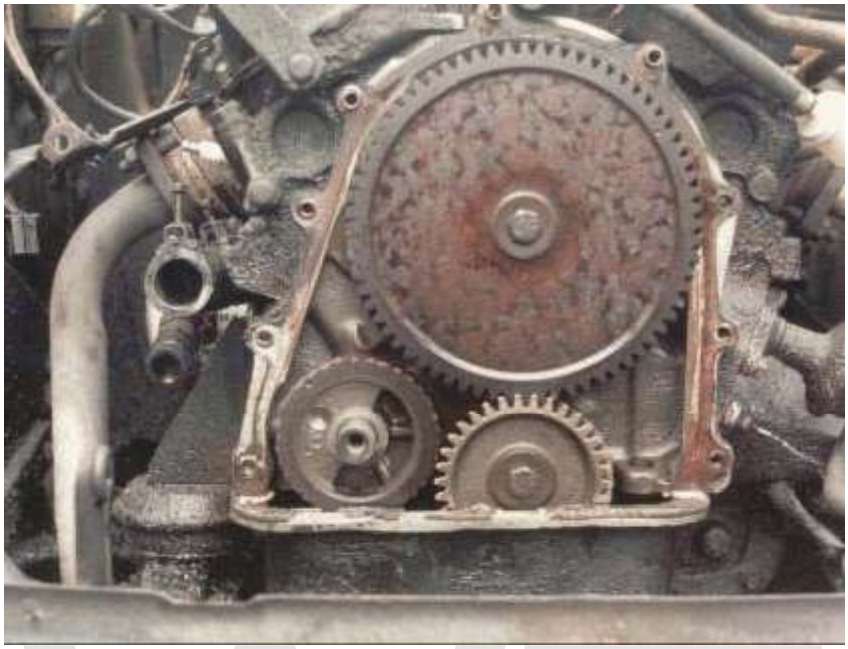
$$v_d = v_D \rightarrow \omega_d r = \omega_D R \rightarrow \frac{R}{r} = \frac{\omega_d}{\omega_D}$$

and so

$$gr = \frac{\omega_d}{\omega_D}$$

In other words, the gear ratio is proportional to ratio of the pitch circles and inversely proportional to the ratio of gear speeds.

Belts can have teeth in them also and be coupled to gear-like pulleys. Special gears called sprockets can be coupled together with chains, as on bicycles and some motorcycles. Again, exact accounting of teeth and revolutions can be applied with these machines.



Valve timing gears on a Ford Taunus V4 engine — the small gear is on the crankshaft, the larger gear is on the camshaft. The crankshaft gear has 34 teeth, the camshaft gear has 68 teeth and runs at half the crankshaft RPM.
(The small gear in the lower left is on the balance shaft.)

A belt with teeth, called the timing belt, is used in some internal combustion engines to exactly synchronize the movement of the camshaft with that of the crankshaft, so that the valves open and close at the top of each cylinder at exactly the right time relative to the movement of each piston. From the time the car is driven off the lot, to the time the belt needs replacing thousands of kilometers later, it synchronizes the two shafts exactly. A chain, called a timing chain, is used on some automobiles for this purpose, while in others, the camshaft and crankshaft are coupled directly together through meshed gears. But whichever form of drive is employed, on four-stroke engines the crankshaft/camshaft gear ratio is always 2:1, which means that for every two revolutions of the crankshaft the camshaft will rotate through one revolution.(In case of 4 stroke engines the valve cycle is repeated after every two rotations of the flywheel.)

Automobile drivetrains generally have two or more areas where gearing is used: one in the transmission, which contains a number of different sets of gearing that can be changed to allow a wide range of vehicle speeds, and another at the differential, which contains one additional set of gearing that provides further speed reduction at the wheels. As well, the differential contains further gearing that splits torque equally between the two wheels while permitting them to have different speeds when traveling a curved path. The components might be separate and connected by a driveshaft, or they might be combined into one unit called a transaxle.

A 2004 Chevrolet Corvette C5 Z06 with a six-speed manual transmission has the following gear ratios in the transmission:

Gear	Ratio
1st gear	2.97:1
2nd gear	2.07:1
3rd gear	1.43:1
4th gear	1.00:1
5th gear	0.84:1
6th gear	0.56:1
reverse	3.38:1

In 1st gear, the engine makes 2.97 revolutions for every revolution of the transmission's output. In 4th gear, the gear ratio of 1:1 means that the engine and the transmission's output are moving at the same speed. 5th and 6th gears are known as overdrive gears, in which the output of the transmission is revolving faster than the engine.

The Corvette above has a differential ratio of 3.42:1. The ratio means that for every 3.42 revolutions of the transmission's output, the wheels make one revolution. The differential ratio multiplies with the transmission ratio, so in 1st gear, the engine makes 10.16 revolutions for every revolution of the wheels.

The car's tires can almost be thought of as a third type of gearing. The example Corvette Z06 is equipped with 295/35-18 tires, which have a circumference of 82.1 inches. This means that for every complete revolution of the wheel, the car travels 82.1 inches. If the Corvette had larger tires, it would travel farther with each revolution of the wheel, which would be like a higher gear. If the car had smaller tires, it would be like a lower gear.

With the gear ratios of the transmission and differential, and the size of the tires, it becomes possible to calculate the speed of the car for a particular gear at a particular engine RPM.

For example, it is possible to determine the distance the car will travel for one revolution of the engine by dividing the circumference of the tire by the combined gear ratio of the transmission and differential.

$$d = \frac{c_t}{gr_t \times gr_d}$$

It is possible to determine a car's speed from the engine speed by multiplying the circumference of the tire by the engine speed and dividing by the combined gear ratio.

$$v_c = \frac{c_t \times v_e}{gr_t \times gr_d}$$

Gear	Distance per engine revolution	Speed per 1000 RPM
1st gear	8.1 in (210 mm)	7.7 mph (12.4 km/h)
2nd gear	11.6 in (290 mm)	11.0 mph (17.7 km/h)
3rd gear	16.8 in (430 mm)	15.9 mph (25.6 km/h)
4th gear	24.0 in (610 mm)	22.7 mph (36.5 km/h)
5th gear	28.6 in (730 mm)	27.1 mph (43.6 km/h)
6th gear	42.9 in (1,090 mm)	40.6 mph (65.3 km/h)

Wide-ratio vs. close-ratio transmission

A close-ratio transmission is a transmission in which there is a relatively little difference between the gear ratios of the gears. For example, a transmission with an engine shaft to drive shaft ratio of 4:1 in first gear and 2:1 in second gear would be considered wide-ratio when compared to another transmission with a ratio of 4:1 in first and 3:1 in second. This is because, for the wide-ratio first gear = 4/1 = 4, second gear = 2/1 = 2, so the transmission gear ratio = 4/2 = 2 (or 200%). For the close-ratio first gear = 4/1 = 4, second gear = 3/1 = 3 so the transmission gear ratio = 4/3 = 1.33 (or 133%), because 133% is less than 200%, the transmission with the 133% ratio between gears is considered close-ratio. However, not all transmissions start out with the same ratio in 1st gear or end with the same ratio in 5th gear, which makes comparing wide vs. close transmission more difficult.

Close-ratio transmissions are generally offered in sports cars, in which the engine is tuned for maximum power in a narrow range of operating speeds and the driver can be expected to enjoy shifting often to keep the engine in its power band.

Factory 4-speed or 5-speed transmission ratios are good compromises for mixed street and moderate performance use, and are "staged" or "progressive", in that the engine speed loss on shifting from 1st to 2nd is higher than the loss on shifting from 2nd to 3rd and so on. The purpose is to keep the engine in its torque range at higher vehicle speed, where wind resistance requires more power for acceleration. Wider gaps between ratios will allow a "stronger" (higher numerically, e.g. 2.90:1 instead of 2.50:1) 1st gear for better manners in traffic, but increase the RPM lost on shifting. Narrowing the gaps will increase acceleration at speed, and potentially improve top speed under certain conditions, but acceleration from stopped and operation in traffic will suffer.

The 1st gear ratio for most 4-speed transmissions is about 2.50:1, and 4th is almost always 1.00:1. The ratios of 2nd and 3rd are placed in between these two, and are discretionary to best serve the weight, intended use, speed, engine tune, and other features of the vehicle.

"Range" is the torque multiplication difference between 1st and 4th gears; wider-ratio gear-sets have more, typically between 2.8 and 3.2. This is the single most important determinant of low-speed acceleration from stopped.

"Progression" is the next factor. This is the reduction or decay in the percentage drop in engine speed in the next gear (e.g. after shifting from 1st to 2nd). Most transmissions have some degree of progression in that the RPM drop on the 1-2 shift is larger than the RPM drop on the 2-3 shift, which is in turn larger than the RPM drop on the 3-4 shift. The progression may not be linear (continuously reduced) or done in proportionate stages for various reasons, including a special need for a gear to reach a specific speed or RPM for passing, racing and so on, or simply economic necessity that the parts were available.

The two factors are not mutually exclusive, but each limits the number of options for the other. A wide range, which gives a strong torque multiplication in 1st gear for excellent manners in low-speed traffic (especially with a smaller motor, heavy chassis or numerically low axle ratio such as 2.50) mean that the progression percentages must all be high. The amount of engine speed (and therefore power) that must be lost on each up-shift is higher than would be the case in a transmission with less range (but less power in 1st gear). A numerically low 1st gear (2.00, &c.) reduces available torque in 1st gear, but allows more choices of progression.

There is no choice of ratios that gives the "best" performance at all speeds, nor is there a choice of final drive (axle) ratio that gives the "best" performance at all speeds. It simply does not exist, all ratios are compromises, and not necessarily better than the original ratios for most use.

The advantage of a close ratio gear-set lies in the fact that the RPM loss at very high speed is reduced, allowing extra power to accelerate above 100 mph. However, of necessity, the torque multiplier in the lower gears is reduced by the same proportion, and performance at low speeds is much worse. Even for road racing, the closest possible ratio is not always the best choice since many races begin with a grid start (favoring slightly wider ratios with high progression, where 1st gear acceleration is very important) and some with a flying start (favoring close ratios, where 1st gear acceleration is less important).

In general, engines with smaller displacement, very long duration cams, ported heads, large carburetors and so on don’t pull well from low rpm, and when the 3-4 shift will benefit more from close ratios in the upper gears, and even more so as the maximum speed at a specific course increases. If the shift takes place at a speed where air resistance is high (70+ mph), closer ratios are better. If your engine has been specifically designed for a tuned RPM torque peak (or if that is how the engine behaves), the transmission ratios must be chosen to ensure that after each shift during a lap the engine speed recovers to a point above this peak at that specific track. From the negative viewpoint, the ratios must be arranged to avoid dropping the engine into a "hole" on an up shift, where power falls off disproportionately.

If the widest ratio change gives a 25% loss, the shift RPM is 7,000 RPM, and there is a torque increase at 5,000 RPM you’re safe: 7,000 - 25% = 5,250, the engine will be in this desirable range on acceleration.

If the widest ratio change is 30%, shift at 7,000, and torque at 5,500: 7,000 - 30% = 4,900, far below the power range and the acceleration (and perhaps the jetting) will be weak until you reach 5,500. You will definitely benefit from a closer gear set, or at least re-arranging the progression to reduce the 30% drop to a better number. Depending on the bike and the track, adding to the drop in the previous gear pair (i.e., problem with the 2-3 shift: add some drop to the 1-2 not the 3-4) is the 1st choice but results will vary.

Individual race tracks with combination of maximum speed and corner speed will require different intermediate (2nd & 3rd) gears to allow downshifting for a specific gear to enter a turn, or to use only one gear during a turn to avoid traction loss. The key to analysis here is whether your favorite track has a spot where the engine is "flat" after shifting at an awkward moment in a turn, but better as it speeds up.

Idler gear

In a sequence of gears chained together, the ratio depends only on the number of teeth on the first and last gear. The intermediate gears, regardless of their size, do not alter the overall gear ratio of the chain. However, the addition of each intermediate gear reverses the direction of rotation of the final gear.

An intermediate gear which does not drive a shaft to perform any work is called an idler gear. Sometimes, a single idler gear is used to reverse the direction, in which case it may be referred to as a *reverse idler*. For instance, the typical automobile manual transmission engages reverse gear by means of inserting a reverse idler between two gears.

Idler gears can also transmit rotation among distant shafts in situations where it would be impractical to simply make the distant gears larger to bring them together. Not only do larger gears occupy more space, the mass and rotational inertia (moment of inertia) of a gear is proportional to the square of its radius. Instead of idler gears, a toothed belt or chain can be used to transmit torque over distance.

Chapter 6

Lift-to-Drag Ratio

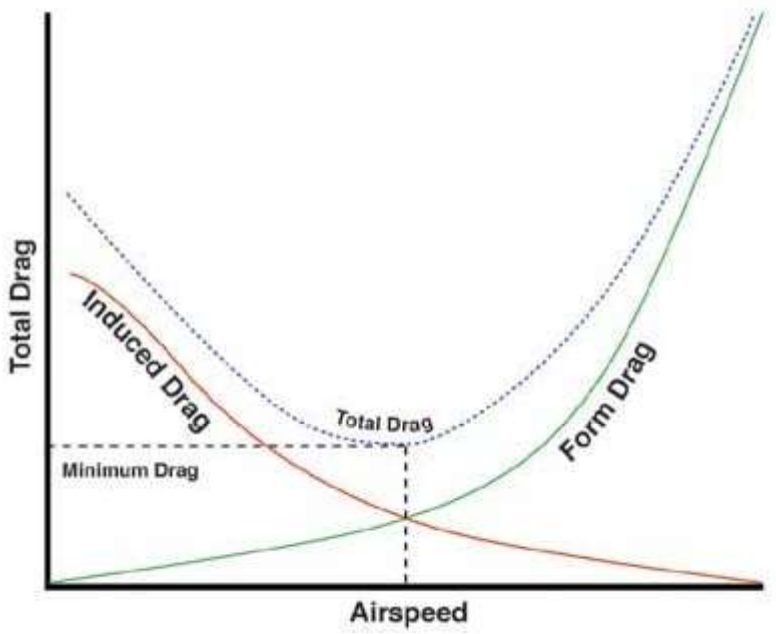
In aerodynamics, the **lift-to-drag ratio**, or **L/D ratio** ("ell-over-dee"), is the amount of lift generated by a wing or vehicle, divided by the drag it creates by moving through the air. A higher or more favorable L/D ratio is typically one of the major goals in aircraft design; since a particular aircraft's required lift is set by its weight, delivering that lift with lower drag leads directly to better fuel economy, climb performance, and glide ratio.

The term is calculated for any particular airspeed by measuring the lift generated, then dividing by the drag at that speed. These vary with speed, so the results are typically plotted on a 2D graph. In almost all cases the graph forms a U-shape, due to the two main components of drag.

Drag

Induced drag is caused by the generation of lift by the wing. Lift generated by a wing is perpendicular to the wing, but since wings typically fly at some small angle of attack, this means that a component of the force is directed to the rear. The rearward component of this force is seen as drag. At low speeds an aircraft has to generate lift with a higher angle of attack, thereby leading to greater induced drag. This term dominates the low-speed side of the L/D graph, the left side of the U.

Profile drag is caused by air hitting the wing, and other parts of the aircraft. This form of drag, also known as wind resistance, varies with the square of speed. For this reason profile drag is more pronounced at higher speeds, forming the right side of the L/D graph's U shape. Profile drag is lowered primarily by reducing cross section and streamlining.



The drag curve

The peak L/D ratio doesn't necessarily occur at the point of least total drag, as the lift produced at that speed is not high, hence a bad L/D ratio. Similarly, the speed at which the highest lift occurs does not have a good L/D ratio, as the drag produced at that speed is too high. The best L/D ratio occurs at a speed somewhere in between (usually slightly above the point of lowest drag). Designers will typically select a wing design which produces an L/D peak at the chosen cruising speed for a powered fixed-wing aircraft, thereby maximizing economy. Like all things in aeronautical engineering, the lift-to-drag ratio is not the only consideration for wing design. Performance at high angle of attack and a gentle stall are also important.

Glide ratio

As the aircraft fuselage and control surfaces will also add drag and possibly some lift, it is fair to consider the L/D of the aircraft as a whole. As it turns out, the glide ratio, which is the ratio of an (unpowered) aircraft's forward motion to its descent, is, when flown at constant speed, numerically equal to the aircraft's L/D. This is especially of interest in the design and operation of high performance sailplanes, which can have glide ratios approaching 60 to 1 (60 units of distance forward for each unit of descent) in the best cases, but with 30:1 being considered good performance for general recreational use. Achieving a glider's best L/D in practice requires precise control of airspeed and smooth and restrained operation of the controls to reduce drag from deflected control surfaces. In zero wind conditions, L/D will equal altitude lost divided by distance traveled. Achieving the maximum distance for altitude lost in wind conditions requires further modification of the best airspeed, as does alternating cruising and thermaling. To achieve high speed across country, gliders are often loaded with water ballast to increase the airspeed (allowing better penetration against a headwind). As noted below, to first order the L/D is

not dependent on speed, although the faster speed means the airplane will fly at higher Reynold's number.

Theory

Mathematically, the maximum lift-to-drag ratio can be estimated as:

$$(L/D)_{max} = \frac{1}{2}\sqrt{\frac{\pi A \epsilon}{C_{D,0}}},$$

where A is the aspect ratio, ε is the aircraft's efficiency factor, and $C_{D,0}$ is the zero-lift drag coefficient.

Supersonic/hypersonic lift to drag ratios

At very high speeds, lift to drag ratios tend to be lower. Concorde had a lift/drag ratio of around 7 at Mach 2, whereas a 747 is around 17 at about mach 0.85.

Dietrich Küchemann developed an empirical relationship for predicting L/D ratio for high Mach:

$$L/D_{max} = \frac{4(M+3)}{M}$$

where M is the Mach number. Windtunnel tests have shown this to be roughly accurate.

Examples

The following table includes some representative L/D ratios.

Flight article	Scenario	L/D ratio
Virgin Atlantic GlobalFlyer	Cruise	37
Lockheed U-2	Cruise	~28
Rutan Voyager	Cruise	27
Albatross		20
Boeing 747	Cruise	17
Common tern		12
Herring gull		10
Concorde	M2 Cruise	7.14
Cessna 150	Cruise	7
Concorde	Approach	4.35
House sparrow		4

In gliding flight, the L/D ratios are equal to the glide ratio.

Flight article	Scenario	L/D ratio / Glide ratio
Modern Sailplane	gliding	45-70 (depending on span)
Hang glider		15
Gimli glider	Boeing 767-200 with fuel exhaustion	~12
Paraglider	high performance model	11
Powered parachute	Rectangular/elliptical parachute	3.6/5.6
Space Shuttle	Approach	4.5
Wingsuit	Gliding	2.5
Northern flying squirrel	Gliding	1.98
Space Shuttle	Hypersonic	1
Apollo CM	Reentry	0.368

Chapter 7

Overall Pressure Ratio

In aeronautical engineering, the term **overall pressure ratio** is defined as the ratio of the stagnation pressure as measured at the front and rear of the compressor of a gas turbine engine. Generally speaking, a higher overall pressure ratio implies higher efficiency, but the engine will weigh more, so there is an optimum.

History of overall pressure ratios

Early jet engines had limited pressure ratios due to construction innaccuracies of the compressors and various material limits. For instance, the Junkers Jumo 004 from World War II had an overall pressure ratio 3.14:1. The immediate post-war SNECMA Atar improved this marginally to 5.2:1. Improvements in materials, compressor blades, and especially the introduction of multi-spool engines with several different rotational speeds, led to the much higher pressure ratios common today. Modern civilian engines generally operate between 30 and 40:1. The three-spool Rolls-Royce Trent 900 used on the Airbus A380, for instance, has a pressure ratio of about 39:1.

Advantages of high overall pressure ratios

A high overall pressure ratio permits a larger area ratio nozzle to be fitted on the jet engine. This means that more of the heat energy is converted to jet speed, and energetic efficiency improves. This is reflected in improvements in the engine's specific fuel consumption.

Disadvantages of high overall pressure ratios

One of the primary limiting factors on pressure ratio in modern designs is that the air heats up as it is compressed. As the air travels through the compressor stages it can reach temperatures that pose a material failure risk for the compressor blades. This is especially true for the last compressor stage, and the inlet temperature to this stage, T_3, is a common figure of merit for engine designs. For civilian engines, the pressure ratio can be adjusted as the aircraft climbs, allowing it to offset some of the heat load through the lowered

pressure and temperature of the high-altitude air. This is one of the many reasons airliners climb to high altitude as quickly as possible.

Military engines are often forced to work under conditions that maximize the heating load. For instance, the General Dynamics F-111 was required to operate at speeds of Mach 1.1 at sea level. As a side-effect of these wide operating conditions, and generally older technology in most cases, military engines typically have lower overall pressure ratios. The Pratt & Whitney TF30 used on the F-111 had a pressure ratio of about 20:1, while newer engines like the General Electric F110 and Pratt & Whitney F135 have improved this to about 30:1.

An additional issue is weight: a higher compression ratio implies a heavier engine, which in turn costs fuel to carry around. Thus, for a particular construction technology and set of flight plans an optimal overall pressure ratio can be determined.

Examples

Engine	Overall pressure ratio	notes
General Electric CF6	30.5:1	Boeing 747, A300
General Electric F110	30:1	F-14, F-15, F-16
Pratt & Whitney TF30	20:1	F-111
Rolls-Royce/Snecma Olympus 593	15.5:1	Concorde, Mach 2 engine

The Concorde's Olympus engines get additional compression from its supersonic inlet, yielding an effective overall pressure ratio of 80:1.

Differences from other similar terms

The term should not be confused with the more familiar term compression ratio applied to reciprocating engines. Compression ratio is a ratio of volumes. In the case of the reciprocating engine, the maximum expansion of the charge is limited by the mechanical movement of the pistons (or rotor), and so the compression can be measured by simply comparing the volume of the cylinder with the piston at the top and bottom of its motion. The same is not true of the "open ended" gas turbine, where operational and structural issues are the limiting factors. Nevertheless the two terms are similar in that they both offer a quick way of determining overall efficiency relative to other engines of the same class.

The broadly equivalent measure of rocket engine efficiency is chamber pressure/exit pressure, and this ratio can be over 2000 for the Space Shuttle Main Engine.

Compression ratio versus overall pressure ratio

For any given gas mix compression ratio and overall pressure ratio are interrelated as follows:

CR	1:1	3:1	5:1	10:1	15:1	20:1	25:1	35:1
PR	1:1	2:1	10:1	22:1	40:1	56:1	75:1	110:1

The reason for this difference is that compression ratio is defined via the volume reduction,

$$CR = \frac{V_1}{V_2},$$

Pressure ratio is defined as the pressure increase

$$PR = \frac{P_2}{P_1}.$$

From the combined gas law we get:

$$\frac{P_1 V_1}{T_1} = \frac{P_2 V_2}{T_2} \Rightarrow \frac{V_1}{V_2} = \frac{T_1}{T_2}\frac{P_2}{P_1} \Leftrightarrow CR = \frac{T_1}{T_2} PR$$

Since T_2 is much higher than T_1 (compressing gases puts work into them, i.e. heats them up), *CR* is much lower than *PR*.

Chapter 8

Power-to-Weight Ratio

Power-to-weight ratio (or **specific power** or **power-to-mass ratio**) is a calculation commonly applied to engines and mobile power sources to enable the comparison of one unit or design to another. Power-to-weight ratio is a measurement of actual performance of any engine or power sources. It is also used as a measurement of performance of a vehicle as a whole, with the engine's power output being divided by the weight (or mass) of the vehicle, to give a metric that is independent of the vehicle's size. Power-to-weight is often quoted by manufacturers at the peak value, but the actual value may vary in use and variations will affect performance.

The inverse of power-to-weight, weight-to-power ratio (power loading) is a calculation commonly applied to aircraft, cars, and vehicles in general, to enable the comparison of one vehicle performance to another. Power-to-weight ratio is equal to powered acceleration multiplied by the velocity of any vehicle.

Power-to-weight (specific power)

The power-to-weight ratio (Specific Power) formula for an engine (power plant) is the power generated by the engine divided by weight of the engine as follows:

$$\text{P-to-W} = P/W$$

A typical turbocharged V8 diesel engine might have an engine power of 330 horsepower (250 kW) and a weight of 835 pounds (379 kg), giving it a power-to-weight ratio of 0.65 kW/kg (0.40 hp/lb).

Examples of high power-to-weight ratios can often be found in turbines. This is because of their ability to operate at very high speeds. For example, the Space Shuttle's main engines use turbopumps (machines consisting of a pump driven by a turbine engine) to feed the propellants (liquid oxygen and liquid hydrogen) into the engine's combustion chamber. The original liquid hydrogen turbopump is similar in size to an automobile engine (weighing approximately 775 pounds (352 kg)) and produces 72,000 hp (53.6 MW) for a power-to-weight ratio of 153 kW/kg (93 hp/lb).

Physical interpretation

In classical mechanics, instantaneous power is the limiting value of the average rate of change of work done per unit time as the time interval Δt approaches zero.

$$P = \lim_{\Delta t \to 0} \frac{\Delta W(t)}{\Delta t} = \lim_{\Delta t \to 0} P_{\text{avg}}$$

If the work to be done is rectilinear motion of a body with constant mass m, whose center of mass is to be accelerated along a straight line to a speed $|\mathbf{v}(t)|$ and angle ϕ with respect to the centre and radial of a gravitational field by an onboard powerplant, then the associated kinetic energy to be delivered to the body is equal to

$$E_K = \tfrac{1}{2} m |\mathbf{v}(t)|^2$$

where:

m is mass of the body
$|\mathbf{v}(t)|$ is speed of the center of mass of the body, changing with time.

The instantaneous mechanical pushing/pulling power delivered to the body from the powerplant is then

$$P_K = \tfrac{1}{2} m 2 |\mathbf{v}(t)| \lim_{\Delta t \to 0} \frac{\Delta |\mathbf{v}(t)|}{\Delta t} = m\mathbf{a}(t) \cdot \mathbf{v}(t) = \mathbf{F}(t) \cdot \mathbf{v}(t) = \tau(t) \cdot \omega(t)$$

where:

$\mathbf{a}(t)$ is acceleration of the center of mass of the body, changing with time.
$\mathbf{F}(t)$ is linear force - or thrust - applied upon the center of mass of the body, changing with time.
$\mathbf{v}(t)$ is velocity of the center of mass of the body, changing with time.
$\tau(t)$ is torque applied upon the center of mass of the body, changing with time.
$\omega(t)$ is angular velocity of the center of mass of the body, changing with time.

Along an obstacle free positive inclined straight road steering an automobile straight ahead, positive linear acceleration will change the speed but not the direction of the vehicle.

The weight of the body is the force applied to the body to support it at rest in a uniform gravitational field, **g**. Using Newton's Second Law of Motion, then

$$\mathbf{F_W} = -m\mathbf{g}$$

where:

m is mass of the body
$\mathbf{g}$ is gravitational field (acceleration) vector

For large changes in altitude or with a body of mass significant when compared with the gravitational field source mass, then the gravitational field may no longer be considered uniform and therefore **g** also changes with time.

The torque from the powerplant is accelerating the body to a desired velocity of motion whilst lifting the weight of the body, overcoming friction through the powerplant and upon the surface of the body (e.g. rolling resistance and skin friction), and overcoming other drag from the motion of the body through fluids (e.g. air, water). The degree in which the deliverable torque associated with the body overcomes the force of gravity upon the body and yields a net positive linear climbing acceleration, or mechanical advantage, is then

$$\mathrm{MA} = \frac{m\mathbf{a}(t) - F_F - F_D}{-m\mathbf{g}} = \frac{|\mathbf{a}(t)| - \frac{1}{m}F_F(m, |\mathbf{v}(t)|) - \frac{1}{m}F_D(|\mathbf{v}(t)|)}{|\mathbf{g}|}$$

where:

m is mass of the body
$|\mathbf{v}(t)|$ is linear speed of the center of mass of the body, changing with time.
$\mathbf{a}(t)$ is *powerplant* acceleration of the center of mass of the body, changing with time.
$\mathbf{g}$ is gravitational acceleration of the center of mass of the body.
F_F is force of friction within powerplant and upon surface of the body.
F_D is force of drag.

If the friction and drag loses are negligible, then the powerplant will convert essentially all its power to either delivering kinetic energy to the body or lifting the weight of the body. To meet this ideal, techniques include low rolling resistance tyres, sufficient tyre inflation, obstacle free path, straight asphalt road, low automotive aerodynamic drag area, well lubricated powertrain and a low loss mechanical transmission to run the engine at a speed that corresponds with the engine peak output power. The mechanical advantage is then simply

$$\mathrm{MA} = \frac{|\mathbf{a}(t)|}{|\mathbf{g}|}$$

Power is only delivered if the powerplant is in motion, and is transmitted to cause the body to be in motion. It is typically assumed here that mechanical transmission allows the powerplant to operate at peak output power. This assumption allows engine tuning to

trade power band width and engine mass for transmission complexity and mass. Electric motors do not suffer from this tradeoff. The **power advantage** or **power-to-weight ratio** is then

$$\text{P-to-W} = \frac{|\mathbf{a}(t)||\mathbf{v}(t)|}{|\mathbf{g}|}$$

where:

$|\mathbf{v}(t)|$ is linear speed of the center of mass of the body.

Power-to-weight ratio is relative to a uniform gravitational field. Normalising to any arbitrary gravitational field yields the **specific power** or **power-to-mass ratio** which is then

$$\text{P-to-M} = |\mathbf{a}(t)||\mathbf{v}(t)| = \frac{|\tau(t)||\omega(t)|}{m}$$

The power-to-weight ratio is typically calculated from power and mass, although mass is usually measured as weight on a calibrated weighing scale. Values are then expressed in units power per unit force exerted on unit mass in standard gravity. Use of kg (kilogram) and lb (pound) rather than kgf (kilogram-force), SI unit N (Newton) or lbf (pound-force) is common. The value thus expressed is the power-to-mass ratio and not the power-to-weight ratio.

The actual useful power of any traction engine can be calculated using a dynamometer to measure torque and rotational speed, with peak power sustained when transmission and/or operator keeps the product of torque and rotational speed maximised. For jet engines there is often a cruise speed and power can be usefully calculated there, for rockets there is typically no cruise speed, so it is less meaningful.

Peak power of a traction engine occurs at a rotational speed higher than the speed when torque is maximised and below the maximum rated rotational speed - Max RPM. A rapidly falling torque curve would correspond with sharp torque and power curve peaks around their maxima at similar rotational speed, for example a small, lightweight engine with a large turbocharger. A slowing falling or near flat torque curve would correspond with a slowly rising power curve up to a maximum at a rotational speed close to Max RPM, for example a large, heavy multi-cylinder engine suitable for cargo/hauling. A falling torque curve could correspond with a near flat power curve across rotational speeds for smooth handling at different vehicle speeds.

Examples

Engines

Heat engines and heat pumps

Thermal energy is made up from molecular kinetic energy and latent phase energy. Heat engines are able to convert thermal energy in the form of a temperature gradient between a hot source and a cold sink into other desirable mechanical work. Heat pumps take mechanical work to regenerate thermal energy in a temperature gradient.

Heat Engine/Heat Pump type	Peak Power Output		Power-to-weight ratio		Example Use
Wärtsilä RTA96-C 14-cylinder two-stroke Turbo Diesel engine	80,080 kW	108,920 hp	0.03 kW/kg	0.02 hp/lb	Emma Mærsk container ship
Suzuki 538 cc V2 4-stroke gas (petrol) outboard Otto engine	19 kW	25 hp	0.27 kW/kg	0.16 hp/lb	Runabout boats
DOE/NASA/0032-28 Mod 2 502 cc gas (petrol) Stirling engine	62.3 kW	83.5 hp	0.30 kW/kg	0.18 hp/lb	Chevrolet Celebrity[•] 1985
GM 6.6 L Duramax LMM (LYE option) V8 Turbo Diesel engine	246 kW	330 hp	0.65 kW/kg	0.40 hp/lb	Chevrolet Kodiak[•], GMC Topkick[•]
Junkers Jumo 205A opposed-piston two-stroke Diesel engine	647 kW	867 hp	1.1 kW/kg	0.66 hp/lb	Ju 86C-1 airliner, B&V Ha 139 floatplane
GE LM2500+ marine turboshaft Brayton gas turbine	30,200 kW	40,500 hp	1.31 kW/kg	0.80 hp/lb	GTS Millennium cruiseship, QM2 ocean liner
Mazda 13B-MSP Renesis 1.3 L Wankel engine	184 kW	247 hp	1.5 kW/kg	0.92 hp/lb	Mazda RX-8[•]
PW R-4360 71.5 L 28-cylinder supercharged Radial engine	3,210 kW	4,300 hp	1.83 kW/kg	1.11 hp/lb	B-50 Superfortress, Convair B-36 C-97 Stratofreighter, C-119 Flying Boxcar Hughes H-4 Hercules "Spruce Goose"

Wright R-3350 54.57 L 18-c s/c Turbo-compound Radial engine	2,535 kW	3,400 hp	2.09 kW/kg	1.27 hp/lb	B-29 Superfortress, Douglas DC-7 C-97 S/f prototype, Kaiser-Frazer C-119F
Pattakon OPRE two stroke Diesel engine	50 kW	70 hp	2.3 kW/kg	1.4 hp/lb	
O.S. Engines 49-PI Type II 4.97 cc UAV Wankel engine	0.934 kW	1.252 hp	2.8 kW/kg	1.7 hp/lb	Model aircraft, Radio-controlled aircraft
GE LM6000 marine turboshaft Brayton gas turbine	44,700 kW	59,900 hp	5.67 kW/kg	3.38 hp/lb	Peaking power plant
GE CF6-80C2 Brayton high-bypass turbofan jet engine					Boeing 747[•], 767, Airbus A300
BMW V10 3L P84/5 2005 gas (petrol) Otto engine	690 kW	925 hp	7.5 kW/kg	4.6 hp/lb	Williams FW27 car[•], Formula One auto racing
GE90-115B Brayton turbofan jet engine	83,164 kW	111,526 hp	10.0 kW/kg	6.10 hp/lb	Boeing 777
PWR RS-24 (SSME) Block II H_2 Brayton turbopump	63,384 kW	85,000 hp	138 kW/kg	84 hp/lb	Space Shuttle (STS-110 and later) [•]
PWR RS-24 (SSME) Block I H_2 Brayton turbopump	53,690 kW	72,000 hp	153 kW/kg	93 hp/lb	Space Shuttle

- Full vehicle power-to-weight ratio shown below

Electric motors/Electromotive generators

An electric motor uses electrical energy to provide mechanical work, usually through the interaction of a magnetic field and current-carrying conductors. By the interaction of mechanical work on an electrical conductor in a magnetic field, electrical energy can be generated.

Electric Motor type	Weight		Peak Power Output		Power-to-weight ratio		Example Use
Panasonic MSMA202S1G AC servo motor	**6.5 kg**	**14.3 lb**	**2 kW**	**2.7 hp**	**0.31 kW/kg**	**0.19 hp/lb**	**Conveyor belts, Robotics**

Toshiba 660 MVA water cooled 23kV AC turbo generator	**1,342 t**	**2,959,000 lb**	**660 MW**	**885,000 hp**	**0.49 kW/kg**	**0.30 hp/lb**	**Bayswater, Eraring Coal Power stations**
Canopy Tech. Cypress 32 MW 15 kV AC PM generator	**33,557 kg**	**73,981 lb**	**32 MW**	**42,913 hp**	**0.95 kW/kg**	**0.58 hp/lb**	**Electric Power stations**
Toyota Brushless AC NdFeB PM motor	**36.3 kg**	**80.0 lb**	**50 kW**	**67 hp**	**1.37 kW/kg**	**0.84 hp/lb**	**Toyota Prius[•] 2004**
Himax HC6332-250 Brushless DC motor	**0.45 kg**	**0.99 lb**	**1.7 kW**	**2.28 hp**	**3.78 kW/kg**	**2.30 hp/lb**	**Radio controlled cars**
Hi-Pa Drive HPD40 Brushless DC wheel hub motor	**25 kg**	**55.1 lb**	**120 kW**	**161 hp**	**4.8 kW/kg**	**2.92 hp/lb**	**Mini QED HEV, Ford F150 HEV**
ElectriFly GPMG4805 Brushless DC	**1.48 kg**	**3.26 lb**	**8.4 kW**	**11.26 hp**	**5.68 kW/kg**	**3.45 hp/lb**	**Radio-controlled aircraft**

- Full vehicle power-to-weight ratio shown below

Fluid engines and fluid pumps

Fluids (liquid and gas) can be used to transmit and/or store energy using pressure and other fluid properties. Hydraulic (liquid) and pneumatic (gas) engines convert fluid pressure into other desirable mechanical or electrical work. Fluid pumps convert mechanical or electrical work into movement or pressure changes of a fluid, or storage in a pressure vessel.

Fluid Powerplant type	Dry Weight		Peak Power Output		Power-to-weight ratio	
PlatypusPower Q2/200 hydroelectric turbine	43 kg	95 lb	2 kW	2.7 hp	0.047 kW/kg	0.029 hp/lb
PlatypusPower PP20/200	330 kg	728 lb	20 kW	27 hp	0.060 kW/kg	0.037 hp/lb

hydroelectric turbine						
Atlas Copco LZL 35 pneumatic motor	20 kg	44.1 lb	5.2 kW	7 hp	0.26 kW/kg	0.16 hp/lb
Bosch 0 607 954 307 pneumatic motor	0.32 kg	0.71 lb	0.1 kW	0.13 hp	0.31 kW/kg	0.19 hp/lb
Bosch 0 607 957 307 pneumatic motor	1.7 kg	3.7 lb	0.74 kW	0.99 hp	0.44 kW/kg	0.26 hp/lb
SAI GM7 radial piston hydraulic motor	300 kg	661 lb	250 kW	335 hp	0.83 kW/kg	0.50 hp/lb
SAI GM3 radial piston hydraulic motor	15 kg	33 lb	15 kW	20 hp	1 kW/kg	0.61 hp/lb

Thermoelectric generators and electrothermal actuators

A variety of effects can be harnessed to produce thermoelectricity, thermionic emission, pyroelectricity and piezoelectricity. Electrical resistance and ferromagnetism of materials can be harnessed to generate thermoacoustic energy from an electric current.

Thermoelectric Powerplant type	Dry Weight		Peak Power Output		Power-to-weight ratio		Example Use
Teledyne Pu GPHS-RTG 1980	56 kg	123 lb	285 We	0.39 hp	5.09 W/kg	0.003 hp/lb	Galileo probe, New Horizons probe
Boeing Pu MMRTG MSL	44.1 kg	97.2 lb	123 We	0.16 hp	2.79 W/kg	0.002 hp/lb	Mars Science Laboratory

Electrochemical (galvanic) and electrostatic cell systems

(Closed cell) batteries

All electrochemical cell batteries deliver a changing voltage as their chemistry changes from "charged" to "discharged". A nominal output voltage and a cutoff voltage are typically specified for a battery by its manufacturer. The output voltage falls to the cutoff voltage when the battery becomes "discharged". The nominal output voltage is always less than the open-circuit voltage produced when the battery is "charged". The temperature of a battery can affect the power it can deliver, where lower temperatures reduce power. Total energy delivered from a single charge cycle is affected by both the battery temperature and the power it delivers. If the temperature lowers or the power demand increases, the total energy delivered at the point of "discharge" is also reduced.

Battery discharge profiles are often described in terms of a factor of battery capacity. For example a battery with a nominal capacity quoted in ampere-hours (Ah) at a C/10 rated discharge current (derived in amperes) may safely provide a higher discharge current -

and therefore higher power-to-weight ratio - but only with a lower energy capacity. Power-to-weight ratio for batteries is therefore less meaningful without reference to corresponding energy-to-weight ratio and cell temperature.

Battery type	Volts	Temp.	Energy-to-weight ratio	Power-to-weight ratio
Energizer 675 Mercury Free Zinc-air battery	1.4V	21°C	1,645 kJ/kg to 0.9 V	1.65 W/kg 2.24 mA
Panasonic R03 AAA Zinc-carbon battery	1.5 V	20±2 °C	47 kJ/kg 20 mA to 0.9 V	3.3 W/kg 20 mA
			88 kJ/kg 150 mA to 0.9 V	24 W/kg 150 mA
Eagle-Picher SAR-10081 60Ah 22-cell Nickel hydrogen battery	27.7 V	10 °C	192 kJ/kg C/2 to 22 V	23 W/kg C/2
			165 kJ/kg C/1 to 22 V	46 W/kg C/1
ClaytonPower 400Ah Lithium-ion battery	12V		617 kJ/kg	85.7 W/kg C/1 (175 A)
Energizer 522 Prismatic Zn/MnO_2 Alkaline battery	9 V	21°C	444 kJ/kg 25 mA to 4.8 V	4.9 W/kg 25 mA
			340 kJ/kg 100 mA to 4.8 V	19.7 W/kg 100 mA
			221 kJ/kg 500 mA to 4.8 V	99 W/kg 500 mA
Panasonic HHR900D 9.25Ah Nickel metal hydride battery	1.2 V	20 °C	209.65 kJ/kg to 0.7 V	11.7 W/kg C/5 58.2 W/kg C/1 116 W/kg 2C
URI 1418Ah replaceable anode Aluminium-air battery model	244.8 V	60 °C	4680 kJ/kg	130.3 W/kg (142 A)
LG Chemical/CPI E2 6Ah $LiMn_2O_4$ Lithium-ion polymer battery	3.8 V	25 °C	530.1 kJ/kg C/2 to 3.0 V	71.25 W/kg
			513 kJ/kg 1C to 3.0 V	142.5 W/kg
			581 kJ/kg C to 2.5 V	161 W/kg
Saft 45E Fe Super-Phosphate™ Lithium iron phosphate battery	3.3 V	25°C	560 kJ/kg 1.14 C to 2.0 V	183 W/kg
			0.73 kJ/kg 2.27 C to 1.5 V	367 W/kg
Energizer CH35 C 1.8Ah Nickel-cadmium battery	1.2 V	21 °C	152 kJ/kg C/10 to 1 V	4 W/kg C/10
			147.1 kJ/kg 5C	200 W/kg 5 C

			to 1 V	
		25 °C	142 kJ/kg C/10 to 7.2 V	4 W/kg C/10
Firefly Energy Oasis FF12D1-G31 6-cell 105Ah VRLA battery	12 V	-1 8°C	7 kJ/kg CCA to 7.2V	234 W/kg CCA (625A)
		0 °C	9 kJ/kg CA to 7.2 V	300 W/kg CA (800 A)
		20 °C	666 kJ/kg C/5.3 to 2.75 V	35 W/kg C/5.3
Panasonic CGA103450A 1.95Ah $LiCoO_2$ Lithium ion battery	3.7 V	0 °C	633 kJ/kg C/1 to 2.75 V	176 W/kg C/1
		20 °C	655 kJ/kg C/1 to 2.75 V	182 W/kg C/1
		20 °C	641 kJ/kg 2C to 2.75 V	356 W/kg 2C
Electric Fuel Battery Corp. UUV 120Ah Zinc-air fuel cell			630 kJ/kg	500 W/kg C/1
Sion Power 2.5Ah Li-S Lithium ion battery	2.15 V	25 °C	1260 kJ/kg	70 W/kg C/5
			1209 kJ/kg	672 W/kg 2C
Maxell / Yuasa / AIST Nickel metal hydride lab prototype		45 °C		980 W/kg
Toshiba SCiB™ cell 4.2Ah Li_2TiO_3 Lithium ion battery	2.4 V	25 °C	242 kJ/kg	67.2 W/kg C/1
			218 kJ/kg	4000 W/kg 12C
Ionix Power Systems $LiMn_2O_4$ Lithium ion battery lab model		lab	270 kJ/kg	1700 W/kg
		lab	29 kJ/kg	4900 W/kg
A123 Systems 26650 Cell 2.3Ah $LiFePO_4$ Lithium ion battery	3.3 V	-20 °C	347 kJ/kg C/1 to 2V	108 W/kg C/1
		0 °C	371 kJ/kg C/1 to 2 V	108 W/kg C/1
		25 °C	390 kJ/kg C/1 to 2 V	108 W/kg C/1
		25 °C	390 kJ/kg 27C to 2 V	3300 W/kg 27C
		25 °C	57 kJ/kg 32C to 2 V	5657 W/kg 32C
Saft VL 6Ah Lithium-ion battery	3.65 V	-20 °C	154 kJ/kg 30C to 2.5 V	41.4 W/kg 30C (180 A)
			182 kJ/kg 1C to 2.5 V	67.4 W/kg 1C
		25 °C	232 kJ/kg 1C to 2.5 V	64.4 W/kg 1C

				233 kJ/kg 58.3C to 2.5 V	3757 W/kg 58.3C (350A)
				34 kJ/kg 267C to 2.5 V	17176 W/kg 267C (1.6kA)
				4.29 kJ/kg 333C to 2.5 V	21370 W/kg 333C (2kA)

Electrostatic, electrolytic and electrochemical capacitors

Capacitors store electric charge onto two electrodes separated by an electric field semi-insulating (dielectric) medium. Electrostatic capacitors feature planar electrodes onto which electric charge accumulates. Electrolytic capacitors use a liquid electrolyte as one of the electrodes and the electric double layer effect upon the surface of the dielectric-electrolyte boundary to increase the amount of charge stored per unit volume. Electric double-layer capacitors extend both electrodes with a nanopourous material such as activated carbon to significantly increase the surface area upon which electric charge can accumulate, reducing the dielectric medium to nanopores and a very thin high permittivity separator.

Whilst capacitors tend not to be as temperature sensitive as batteries, they are significantly capacity constrained and without the strength of chemical bonds suffer from self-discharge. Power-to-weight ratio of capacitors is usually higher than batteries because charge transport units within the cell are smaller (electrons rather than ions), however energy-to-weight ratio is conversely usually lower.

Capacitor type	Capacity	Volts	Temp.	Energy-to-weight ratio	Power-to-weight ratio
ACT Premlis® Lithium ion capacitor	2000 F	4.0 V	25 °C	54 kJ/kg to 2.0 V	44.4 W/kg @ 5 A
				31 kJ/kg to 2.0 V	850 W/kg @ 10 A
Nesccap Electric double-layer capacitor	5000 F	2.7 V	25 °C	19.58 kJ/kg to 1.35 V	5.44 W/kg C/1 (1.875 A)
				5.2 kJ/kg to 1.35 V	5,200 W/kg @ 2,547A
EEStor EESU barium titanate supercapacitor	30.693 F	3500 V	85 °C	1471.98 kJ/kg	80.35 W/kg C/5
				1471.98 kJ/kg	8,035 W/kg 20 C

Fuel cell stacks and flow cell batteries

Fuel cells and flow cells, although perhaps using similar chemistry to batteries, have the distinction of not containing the energy storage medium or fuel. With a continuous flow of fuel and oxidant, available fuel cells and flow cells continue to convert the energy storage medium into electric energy and waste products. Fuel cells distinctly contain a

fixed electrolyte whereas flow cells also require a continuous flow of electrolyte. Flow cells typically have the fuel dissolved in the electrolyte.

Fuel cell type	Dry weight	Power-to-weight ratio	Example Use
Redflow Power+BOS® ZB600 10kWh ZBB	900 kg	5.6 W/kg (9.3 W/kg peak)	Rural Grid support
Ceramic Fuel Cells BlueGen MG 2.0 CHP SOFC	200 kg	10 W/kg 15 W/kg CHP	
MTU Friedrichshafen 240 kW MCFC HotModule 2006	20 t	12 W/kg	
Smart Fuel Cell Jenny 600S 25W DMFC	1.7 kg	14.7 W/kg	Portable military electronics
UTC Power PureCell 400 kW PAFC	27,216 kg	14.7 W/kg	
GEFC 50V50A-VRB Vanadium redox battery	80 kg	31.3 W/kg (125 W/kg peak)	
Ballard Power Systems Xcellsis™ HY-205 205 kW PEMFC	2,170 kg	94.5 W/kg	Mercedes-Benz Citaro O530BZ[•]
UTC Power/NASA 12 kW AFC	122 kg	98 W/kg	Space Shuttle orbiter[•]
Ballard Power Systems FCgen-1030 1.2 kW CHP PEMFC	12 kg	100 W/kg	Residential cogeneration
Ballard Power Systems FCvelocity-HD6 150 kW PEMFC	400 kg	375 W/kg	Bus and heavy duty
Honda 2003 43 kW FC Stack PEMFC[•]	43 kg	1000 W/kg	Honda FCX Clarity[•]
Lynntech, Inc. PEMFC lab prototype	347 g	1,500 W/kg	

- Full vehicle power-to-weight ratio shown below

Photovoltaics

Photovoltaic Panel type	Power-to-weight ratio
Thyssen Solartec 128W Nanocrystalline Si Triplejunction PV module	6 W/kg
Suntech/UNSW HiPerforma PLUTO220-Udm 220W Ga-F22 Polycrystalline Si PV module	13.1 W/kg STP 9.64 W/kg nominal
Global Solar PN16015A 62W CIGS polycrystalline thin film PV module	40 W/kg
Able (AEC) PUMA 6 kW GaInP2/GaAs/Ge-on-Ge Triplejunction PV array	65 W/kg

Current spacecraft grade	~77 W/kg
ITO/InP on Kapton foil	2000 W/kg

Vehicles

Power-to-weight ratios for vehicles are usually calculated using curb weight (for cars) or wet weight (for motorcycles) - in other words, excluding weight of the driver and any cargo. This could be slightly misleading, especially with regard to motorcycles, where the driver might weigh 1/3 to 1/2 as much as the vehicle itself. In the sport of competitive cycling athlete's performance is increasingly being expressed in VAMs and thus as a power-to-weight ratio in W/kg. This can be measured through the use of a bicycle powermeter or calculated from measuring incline of a road climb and the rider's time to ascend it.

Utility and practical vehicles

Most vehicles are designed to meet passenger comfort and cargo carrying requirements. Different designs trade off power-to-weight ratio to increase comfort, cargo space, fuel economy, emissions control, energy security and endurance. Reduced drag and lower rolling resistance in a vehicle design can facilitate increased cargo space without increase in the (zero cargo) power-to-weight ratio. This increases the role flexibility of the vehicle. Energy security considerations can trade off power (typically decreased) and weight (typically increased), and therefore power-to-weight ratio, for fuel flexibility or drive-train hybridisation. Some utility and practical vehicle variants such as hot hatches and sports-utility vehicles reconfigure power (typically increased) and weight to provide the perception of sports car like performance or for other psychological benefit.

Notable low ratio

Vehicle	Power	Weight	Power-to-weight ratio
Benz Patent Motorwagen 954 cc 1886	560 W / 0.75 bhp	265 kg / 584 lb	2.1 W/kg / 779 lb/hp
Stephenson's Rocket 0-2-2 steam locomotive with tender 1829	15 kW / 20 bhp	4,320 kg / 9524 lb	3.5 W/kg / 476 lb/hp
CBQ Zephyr streamliner diesel locomotive with railcars 1934	492 kW / 660 bhp	94 t / 208,000 lb	5.21 W/kg / 315 lb/hp
Alberto Contador's Verbier climb 2009 Tour de France on Specialized bike	420 W / 0.56 bhp	62 kg / 137 lb	6.7 W/kg / 245 lb/hp
Force Motors Minidor Diesel 499 cc auto rickshaw	6.6 kW / 8.8 bhp	700 kg / 1543 lb	9 W/kg / 175 lb/hp
PRR Q2 4-4-6-4 steam locomotive with tender 1944	5,956 kW / 7,987 bhp	475.9 t / 1,049,100 lb	12.5 W/kg / 131 lb/hp
Mercedes-Benz Citaro O530BZ H_2 fuel	205 kW /	14,500 kg /	14.1 W/kg /

cell bus 2002	275 bhp	32,000 lb	116 lb/hp
TGV BR Class 373 high-speed electric locomotive 1993	12,240 kW / 16,414 bhp	816 t / 1,798,972 lb	15 W/kg / 110 lb/hp
General Dynamics M1 Abrams Main battle tank 1980	1,119 kW / 1500 bhp	55.7 t / 122,800 lb	20.1 W/kg / 81.9 lb/hp
BR Class 43 high-speed diesel electric locomotive 1975	1,678 kW / 2,250 bhp	70.25 t / 154,875 lb	23.9 W/kg / 69 lb/hp
GE AC6000CW diesel electric locomotive 1996	4,660 kW / 6,250 bhp	192 t / 423,000 lb	24.3 W/kg / 68 lb/hp
BR Class 55 Napier Deltic diesel electric locomotive 1961	2,460 kW / 3,300 bhp	101 t / 222,667 lb	24.4 W/kg / 68 lb/hp
International CXT 2004	164 kW / 220 bhp	6,577 kg / 14500 lb	25 W/kg / 66 lb/hp
Ford Model T 2.9 L flex-fuel 1908	15 kW / 20 bhp	540 kg / 1,200 lb	28 W/kg / 60 lb/hp
TH!NK City 2008	30 kW / 40 bhp	1038 kg / 2,288 lb	28.9 W/kg / 56.9 lb/hp
Messerschmitt KR200 Kabinenroller 191 cc 1955	6 kW / 8.2 bhp	230 kg / 506 lb	30 W/kg / 50 lb/hp
Wright Flyer 1903	9 kW / 12 bhp	274 kg / 605 lb	33 W/kg / 50 lb/hp
Tata Nano 624 cc 2008	26 kW / 35 bhp	635 kg / 1,400 lb	41.0 W/kg / 40 lb/hp
Bombardier JetTrain high-speed gas turbine-electric locomotive 2000	3,750 kW / 5,029 bhp	90,750 kg / 200,000 lb	41.2 W/kg / 39.8 lb/hp
Suzuki MightyBoy 543 cc 1988	23 kW / 31 bhp	550 kg / 1,213 lb	42 W/kg / 39 lb/hp
Mitsubishi i MiEV 2009	47 kW / 63 bhp	1,080 kg / 2,381 lb	43.5 W/kg / 37.8 lb/hp
Holden FJ 2,160 cc 1953	44.7 kW / 60 bhp	1,021 kg / 2,250 lb	43.8 W/kg / 37.5 lb/hp
Chevrolet Kodiak/GMC Topkick LYE 6.6 L 2005	246 kW / 330 bhp	5126 kg / 11,300 lb	48 W/kg / 34.2 lb/hp
DOE/NASA/0032-28 Chevrolet Celebrity 502 cc ASE Mod II 1985	62.3 kW / 83.5 bhp	1,297 kg / 2,860 lb	48.0 W/kg / 34.3 lb/hp
Suzuki Alto 796 cc 2000	35 kW / 46 bhp	720 kg / 1,587 lb	49 W/kg / 35 lb/hp
Land Rover Defender 2.4 L 1990	90 kW / 121 bhp	1,837 kg / 4,050 lb	49 W/kg / 33 lb/hp

Common power

Vehicle	Power	Weight	Power-to-weight ratio
Toyota Prius 1.8 L 2010 (petrol only)	73 kW / 98 bhp	1,380 kg / 3,042 lb	53 W/kg / 31 lb/hp
Bajaj Platina Naked 100 cc 2006	6 kW / 8 bhp	113 kg / 249 lb	53 W/kg / 31 lb/hp
Subaru R2 type S 2003	47 kW / 63 bhp	830 kg / 1,830 lb	57 W/kg / 29 lb/hp
Ford Fiesta ECOnetic 1.6 L TDCi 5dr 2009	66 kW / 89 bhp	1,155 kg / 2,546 lb	57 W/kg / 29 lb/hp
Volvo C30 1.6D DRIVe S/S 3dr Hatch 2010	80 kW / 108 bhp	1,347 kg / 2,970 lb	59.4 W/kg / 27.5 lb/hp
Ford Focus ECOnetic 1.6 L TDCi 5dr Hatch 2009	81 kW / 108 bhp	1,357 kg / 2,992 lb	59.7 W/kg / 27 lb/hp
Ford Focus 1.8 L Zetec S TDCi 5dr Hatch 2009	84 kW / 113 bhp	1,370 kg / 3,020 lb	61 W/kg / 27 lb/hp
Honda FCX Clarity 4 kg Hydrogen 2008	100 kW / 134 bhp	1,600 kg / 3,528 lb	63 W/kg / 26 lb/hp
Hummer H1 6.6 L V8 2006	224 kW / 300 bhp	3,559 kg / 7,847 lb	63 W/kg / 26 lb/hp
Audi A2 1.4 L TDI 90 type S 2003	66 kW / 89 bhp	1,030 kg / 2,270 lb	64 W/kg / 25 lb/hp
Opel/Vauxhall/Holden/Chevrolet Astra 1.7 L CTDi 125 2010	92 kW / 123 bhp	1,393 kg / 3,071 lb	66 W/kg / 24.9 lb/hp
Mini (new) Cooper 1.6D 2007	81 kW / 108 bhp	1,185 kg / 2,612 lb	68 W/kg / 24 lb/hp
Toyota Prius 1.8 L 2010 (electric boost)	100 kW / 134 bhp	1,380 kg / 3,042 lb	72 W/kg / 23 lb/hp
Ford Focus 2.0 L Zetec S TDCi 5dr Hatch 2009	100 kW / 134 bhp	1,370 kg / 3,020 lb	73 W/kg / 23 lb/hp
Toyota Venza I4 2.7 L FWD 2009	136 kW / 182 bhp	1,706 kg / 3,760 lb	80 W/kg / 20.7 lb/hp
Ford Focus 2.0 L Zetec S 5dr Hatch 2009	107 kW / 143 bhp	1,327 kg / 2,926 lb	81 W/kg / 20 lb/hp
Fiat Grande Punto 1.6 L Multijet 120 2005	88 kW / 118 bhp	1,075 kg / 2,370 lb	82 W/kg / 20 lb/hp
Mini (classic) 1275GT 1969	57 kW / 76 bhp	686 kg / 1,512 lb	83 W/kg / 20 lb/hp
Opel/Vauxhall/Holden/Chevrolet Astra 2.0 L CTDi 160 2010	118 kW / 158 bhp	1,393 kg / 3,071 lb	85 W/kg / 19.4 lb/hp

Subaru Legacy/Liberty 2.0R 2005	121 kW / 162 bhp	1,370 kg / 3,020 lb	88 W/kg / 19 lb/hp
Subaru Outback 2.5i 2008	130.5 kW / 175 bhp	1,430 kg / 3,153 lb	91 W/kg / 18 lb/hp
Smart Fortwo 1.0 L Bradbus 2009	72 kW / 97 bhp	780 kg / 1,720 lb	92 W/kg / 18 lb/hp
Ford Focus 2.0 auto 2007	104.4 kW / 140 bhp	1,198 kg / 2,641 lb	87.1 W/kg / 19 lb/hp
Toyota Venza V6 3.5 L AWD 2009	200 kW / 268 bhp	1,835 kg / 4,045 lb	109 W/kg / 15 lb/hp
Toyota Venza I4 2.7 L FWD 2009 with Lotus mass reduction	136 kW / 182 bhp	1,210 kg / 2,667 lb	112.2 W/kg / 14.7 lb/hp
Toyota Hilux V6 DOHC 4 L 4x2 Single Cab Pickup ute 2009	175 kW / 235 bhp	1,555 kg / 3,428 lb	112.5 W/kg / 14.6 lb/hp
Toyota Venza V6 3.5 L FWD 2009	200 kW / 268 bhp	1,755 kg / 3,870 lb	114 W/kg / 14.4 lb/hp

Performance luxury, roadsters and mild sports

Some utility and practical vehicles are designed for, as BMW would say, sheer driving pleasure. Increased engine performance is a consideration, but also other features associated with luxury vehicles. Longitudinal engines are common. Bodies vary from hot hatches, sedans (saloons), coupés, convertibles and roadsters. Mid-range dual-sport and cruiser motorcycles tend to have similar power-to-weight ratios.

Vehicle	Power	Weight	Power-to-weight ratio
Mini (new) Cooper 1.6T S JCW 2008	155 kW / 208 bhp	1205 kg / 2657 lb	129 W/kg / 13 lb/hp
Mazda RX-8 1.3 L Wankel 2003	173 kW / 232 bhp	1309 kg / 2888 lb	141 W/kg / 12 lb/hp
Holden Statesman/Caprice / Buick Park Avenue / Daewoo Veritas 6 L V8 2007	270 kW / 362 bhp	1891 kg / 4170 lb	143 W/kg / 12 lb/hp
Kawasaki KLR650 Gasoline DualSport 650 cc	26 kW / 35 bhp	182 kg / 401 lb	143 W/kg / 11 lb/hp
NATO HTC M1030M1 Diesel/Jet fuel DualSport 670 cc	26 kW / 35 bhp	182 kg / 401 lb	143 W/kg / 11 lb/hp
Harley-Davidson FLSTF Softail Fat Boy Cruiser 1,584 cc 2009	47 kW / 63 bhp	324 kg / 714 lb	145 W/kg / 11.3 lb/hp
BMW 7 Series 760Li 6 L V12 2006	327 kW / 439 bhp	2250 kg / 4960 lb	145 W/kg / 11 lb/hp
Subaru Impreza WRX STi 2.0 L 2008	227 kW / 304 bhp	1530 kg / 3373 lb	148 W/kg / 11 lb/hp

Tesla Roadster 2008	185 kW / 248 bhp	1235 kg / 2723 lb	150 W/kg / 11 lb/hp
GMH HSV Clubsport / GMV VXR8 / GMC CSV CR8 / Pontiac G8 6 L V8 2006	317 kW / 425 bhp	1831 kg / 4037 lb	173 W/kg / 9.5 lb/hp

Sports vehicles and aircraft

Power-to-weight ratio is an important vehicle characteristic that affects the acceleration and handling - and therefore the driving enjoyment - of any sports vehicle. Aircraft also depend on high power-to-weight ratio to achieve sufficient lift.

Vehicle	Power	Weight	Power-to-weight ratio
Lotus Elise SC 2008	163 kW / 218 bhp	910 kg / 2006 lb	179 W/kg / 9 lb/hp
Ferrari Testarossa 1984	291 kW / 390 bhp	1506 kg / 3320 lb	193 W/kg / 9 lb/hp
Artega GT	220 kW / 300 bhp	1100 kg / 2425 lb	200 W/kg / 8 lb/hp
Lotus Exige GT3 2006	202.1 kW / 271 bhp	980 kg / 2160 lb	206 W/kg / 8 lb/hp
Bombardier Dash 8 Q400 turboprop airliner	3,781 kW / 5,071 bhp	17,185 kg / 37,888 lb	220 W/kg / 7.5 lb/hp
Chevrolet Corvette C6	321 kW / 430 bhp	1441 kg / 3177 lb	223 W/kg / 7 lb/hp
Suzuki V-Strom 650 V-twin DualSport 650 cc	50 kW / 67 bhp	194 kg / 427 lb	258 W/kg / 6.4 lb/hp
Chevrolet Corvette C6 Z06	376 kW / 505 bhp	1421 kg / 3133 lb	265 W/kg / 6 lb/hp
Porsche 911 GT2 2007	390 kW / 523 bhp	1440 kg / 3200 lb	271 W/kg / 6.1 lb/hp
Lamborghini Murciélago LP 670-4 SV 2009	493 kW / 661 bhp	1550 kg / 3417 lb	318 W/kg / 5.1 lb/hp
McLaren F1 GT 1997	467.6 kW / 627 bhp	1220 kg / 2690 lb	403 W/kg / 4 lb/hp
Supermarine Spitfire Fighter aircraft 1936	1,096 kW / 1,470 bhp	2,309 kg / 5,090 lb	475 W/kg / 3.46 lb/hp
Messerschmitt Bf 109 Fighter aircraft 1935	1,085 kW / 1,455 bhp	2,247 kg / 4,954 lb	483 W/kg / 3.40 lb/hp
Thunderbolt Land speed record car	3504 kW / 4700 bhp	7 t / 15432 lb	500 W/kg / 3.28 lb/hp
Ferrari FXX 2005	597 kW / 801 bhp	1155 kg /	517 W/kg /

		2546 lb	3.2 lb/hp
Polaris Industries Assault Snowmobile 2009	115 kW / 154 bhp	221 kg / 487 lb	523 W/kg / 3.16 lb/hp
Ultima GTR 720 2006	536.9 kW / 720 bhp	920 kg / 2183 lb	583 W/kg / 3 lb/hp
Honda CBR1000RR 2009	133 kW / 178 bhp	199 kg / 439 lb	668 W/kg / 2.5 lb/hp
KillaCycle Drag racing electric motorcycle	260 kW / 350 bhp	281 kg / 619 lb	925 W/kg / 1.77 lb/hp
MTT Turbine Superbike 2008	213.3 kW / 286 bhp	227 kg / 500 lb	940 W/kg / 1.75 lb/hp
Vyrus 987 C3 4V V supercharged motorcycle 2010	157.3 kW / 211 bhp	158 kg / 348.3 lb	996 W/kg / 1.65 lb/hp
BMW Williams FW27 Formula One 2005	690 kW / 925 bhp	600 kg / 1323 lb	1150 W/kg / 1.43 lb/hp
Honda RC211V MotoGP 2004-6	176.73 kW / 237 bhp	148 kg / 326 lb	1194 W/kg / 1.37 lb/hp
Boeing 747-300 at Mach 0.84 cruise, 35,000 ft altitude	245 MW / 328,656 bhp	178.1 t / 392,800 lb	1376 W/kg / 1.20 lb/hp
John Force Racing Funny Car NHRA Drag Racing 2008	5,963.60 kW / 8,000 bhp	1043 kg / 2,300 lb	5717 W/kg / 0.30 lb/hp

Supersonic vehicles

Some sports and aerospace vehicles are capable of exceeding the speed of sound. Vehicles in this class must account for transonic wave drag, shock waves and aerodynamic heating. Turbojet, turbofan, ramjet, afterburner and rocket propulsion is common.

Vehicle	Power	Weight	Power-to-weight ratio
Space Shuttle Endeavour (OV-105)	190 MW / 255,000 bhp	78 t / 172,000 lb	2,437 W/kg / 0.7 lb/hp
Aérospatiale/BAC Concorde 1969 at Mach 2.02 supercruise full thrust	330 MW / 443,143 bhp	78.7 t / 173,500 lb	4,199 W/kg / 0.39 lb/hp
Thrust Super Sonic Car	82 MW / 110,000 bhp	10.5 t / 23,149 lb	7,812 W/kg / 0.21 lb/hp
F-35 Lightning II Multirole combat aircraft 2006 at Mach 1.67 full thrust	110 MW / 146,922 bhp	13,300 kg / 29,300 lb	8,238 W/kg / 0.20 lb/hp
Sukhoi Su-35BM Multirole combat aircraft 2008 at Mach 2.25 full thrust	188.5 MW / 252,842 bhp	18,400 kg / 40,500 lb	10,247 W/kg / 0.16 lb/hp
Lockheed SR-71 Blackbird Surveillance aircraft 1966 at Mach	318.3 MW / 426,853 bhp	30,600 kg / 67,461 lb	10,402 W/kg / 0.16 lb/hp

3.2, 290 kN thrust			
Sukhoi T-50 Stealth multirole fighter aircraft 2010 at Mach 2, 294 kN thrust	201 MW / 270,620 bhp	18,500 kg / 40,786 lb	10,908 W/kg / 0.15 lb/hp
F-22 Raptor Air superiority fighter aircraft 1990 at Mach 2.25, 312 kN thrust	241 MW / 323,088 bhp	19,700 kg / 43,430 lb	12,230 W/kg / 0.13 lb/hp
Rockwell X-30 scramjet SSTO spaceplane 1986 at Mach 30, 1.4 MN thrust (proposed)	14 GW / 19,300,000 bhp	136 t / 300,000 lb	105,740 W/kg / 0.016 lb/hp
Space Shuttle Endeavour (OV-105) with SLWT and 2 SRBs	33 GW / 44,000,000 bhp	280 t / 616,500 lb	118,000 W/kg / 0.014 lb/hp

Chapter 9

Signal-to-Noise Ratio

Signal-to-noise ratio (often abbreviated **SNR** or **S/N**) is a measure used in science and engineering to quantify how much a signal has been corrupted by noise. It is defined as the ratio of signal power to the noise power corrupting the signal. A ratio higher than 1:1 indicates more signal than noise. While SNR is commonly quoted for electrical signals, it can be applied to any form of signal (such as isotope levels in an ice core or biochemical signaling between cells).

In less technical terms, signal-to-noise ratio compares the level of a desired signal (such as music) to the level of background noise. The higher the ratio, the less obtrusive the background noise is.

"Signal-to-noise ratio" is sometimes used informally to refer to the ratio of useful information to false or irrelevant data in a conversation or exchange. For example, in online discussion forums and other online communities, off-topic posts and spam are regarded as "noise" that interferes with the "signal" of appropriate discussion.

Definition

Signal-to-noise ratio is defined as the power ratio between a signal (meaningful information) and the background noise (unwanted signal):

$$\mathrm{SNR} = \frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}},$$

where P is average power. Both signal and noise power must be measured at the same or equivalent points in a system, and within the same system bandwidth. If the signal and the noise are measured across the same impedance, then the SNR can be obtained by calculating the square of the amplitude ratio:

$$\mathrm{SNR} = \frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}} = \left(\frac{A_{\mathrm{signal}}}{A_{\mathrm{noise}}}\right)^2,$$

where A is root mean square (RMS) amplitude (for example, RMS voltage). Because many signals have a very wide dynamic range, SNRs are often expressed using the logarithmic decibel scale. In decibels, the SNR is defined as

$$\mathrm{SNR_{dB}} = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right) = P_{\mathrm{signal,dB}} - P_{\mathrm{noise,dB}},$$

which may equivalently be written using amplitude ratios as

$$\mathrm{SNR_{dB}} = 10\log_{10}\left(\frac{A_{\mathrm{signal}}}{A_{\mathrm{noise}}}\right)^2 = 20\log_{10}\left(\frac{A_{\mathrm{signal}}}{A_{\mathrm{noise}}}\right).$$

The concepts of signal-to-noise ratio and dynamic range are closely related. Dynamic range measures the ratio between the strongest un-distorted signal on a channel and the minimum discernable signal, which for most purposes is the noise level. SNR measures the ratio between an arbitrary signal level (not necessarily the most powerful signal possible) and noise. Measuring signal-to-noise ratios requires the selection of a representative or *reference* signal. In audio engineering, the reference signal is usually a sine wave at a standardized nominal or alignment level, such as 1 kHz at +4 dBu (1.228 V_{RMS}).

SNR is usually taken to indicate an *average* signal-to-noise ratio, as it is possible that (near) instantaneous signal-to-noise ratios will be considerably different. The concept can be understood as normalizing the noise level to 1 (0 dB) and measuring how far the signal 'stands out'.

Alternative definition

An alternative definition of SNR is as the reciprocal of the coefficient of variation, i.e., the ratio of mean to standard deviation of a signal or measurement:

$$\mathrm{SNR} = \frac{\mu}{\sigma}$$

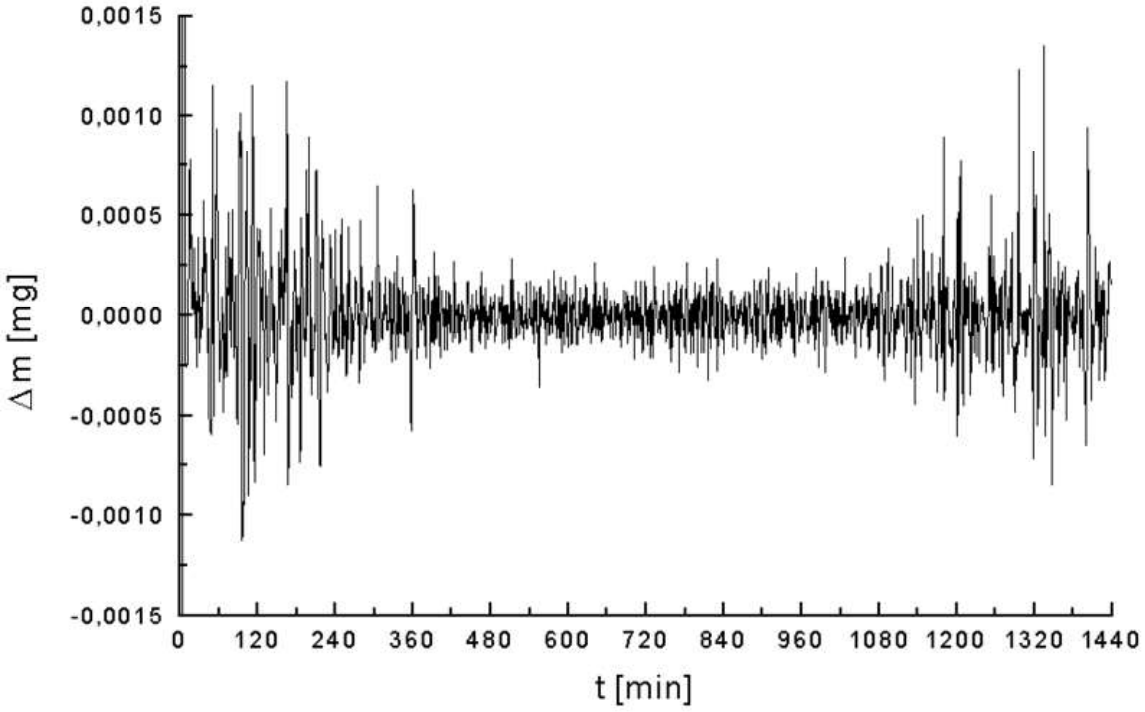
where μ is the signal mean or expected value and σ is the standard deviation of the noise, or an estimate thereof. Notice that such an alternative definition is only useful for variables that are always positive (such as photon counts and luminance). Thus it is commonly used in image processing, where the SNR of an image is usually calculated as the ratio of the mean pixel value to the standard deviation of the pixel values over a given neighborhood. Sometimes SNR is defined as the square of the alternative definition above.

The *Rose criterion* (named after Albert Rose) states that an SNR of at least 5 is needed to be able to distinguish image features at 100% certainty. An SNR less than 5 means less than 100% certainty in identifying image details.

Yet another alternative, very specific and distinct definition of SNR is employed to characterize sensitivity of imaging systems.

Related measures are the "contrast ratio" and the "contrast-to-noise ratio".

Improving SNR in practice



Recording of the noise of a thermogravimetric analysis device that is poorly isolated from a mechanical point of view; the middle of the curve shows a lower noise, due to a lesser surrounding human activity at night.

All real measurements are disturbed by noise. This includes electronic noise, but can also include external events that affect the measured phenomenon — wind, vibrations, gravitational attraction of the moon, variations of temperature, variations of humidity, etc., depending on what is measured and of the sensitivity of the device. It is often possible to reduce the noise by controlling the environment. Otherwise, when the characteristics of the noise are known and are different from the signals, it is possible to filter it or to process the signal. When the signal is constant or periodic and the noise is random, it is possible to enhance the SNR by averaging the measurement.

Digital signals

When a measurement is digitised, the number of bits used to represent the measurement determines the maximum possible signal-to-noise ratio. This is because the minimum possible noise level is the error caused by the quantization of the signal, sometimes called Quantization noise. This noise level is non-linear and signal-dependent; different calculations exist for different signal models. Quantization noise is modeled as an analog error signal summed with the signal before quantization ("additive noise").

This theoretical maximum SNR assumes a perfect input signal. If the input signal is already noisy (as is usually the case), the signal's noise may be larger than the quantization noise. Real analog-to-digital converters also have other sources of noise that further decrease the SNR compared to the theoretical maximum from the idealized quantization noise, including the intentional addition of dither.

Although noise levels in a digital system can be expressed using SNR, it is more common to use E_b/N_o, the energy per bit per noise power spectral density.

The modulation error ratio (MER) is a measure of the SNR in a digitally modulated signal.

Fixed point

For *n*-bit integers with equal distance between quantization levels (uniform quantization) the dynamic range (DR) is also determined.

Assuming a uniform distribution of input signal values, the quantization noise is a uniformly-distributed random signal with a peak-to-peak amplitude of one quantization level, making the amplitude ratio $2^n/1$. The formula is then:

$$\mathrm{DR_{dB}} = \mathrm{SNR_{dB}} = 20\log_{10}(2^n) \approx 6.02 \cdot n$$

This relationship is the origin of statements like "16-bit audio has a dynamic range of 96 dB". Each extra quantization bit increases the dynamic range by roughly 6 dB.

Assuming a full-scale sine wave signal (that is, the quantizer is designed such that it has the same minimum and maximum values as the input signal), the quantization noise approximates a sawtooth wave with peak-to-peak amplitude of one quantization level and uniform distribution. In this case, the SNR is approximately

$$\mathrm{SNR_{dB}} \approx 20\log_{10}(2^n\sqrt{3/2}) \approx 6.02 \cdot n + 1.761$$

Floating point

Floating-point numbers provide a way to trade off signal-to-noise ratio for an increase in dynamic range. For n bit floating-point numbers, with n-m bits in the mantissa and m bits in the exponent:

$$\mathrm{DR_{dB}} = 6.02 \cdot 2^m$$
$$\mathrm{SNR_{dB}} = 6.02 \cdot (n - m)$$

Note that the dynamic range is much larger than fixed-point, but at a cost of a worse signal-to-noise ratio. This makes floating-point preferable in situations where the dynamic range is large or unpredictable. Fixed-point's simpler implementations can be used with no signal quality disadvantage in systems where dynamic range is less than 6.02m. The very large dynamic range of floating-point can be a disadvantage, since it requires more forethought in designing algorithms.

Optical SNR

Optical signals have a carrier frequency, which is much higher than the modulation frequency (about 200 THz and more). This way the noise bandwidth covers a bandwidth which is much wider than the signal itself. The resulting signal influence relies mainly on the filtering of the noise. To describe the signal quality without taking the receiver into account the optical SNR (OSNR) is used. The OSNR is the ratio between the signal power and the noise power in a given bandwidth. Most commonly a reference bandwidth of 0.1 nm is used. This bandwidth is independent from the modulation format, the frequency and the receiver. For instance a OSNR of 20dB/0.1nm could be given, even the signal of 40 GBit DPSK would not fit in this bandwidth. OSNR is measured with a Optical Spectrum Analyzer

Chapter 10

Standing Wave Ratio

In telecommunications, **standing wave ratio** (**SWR**) is the ratio of the amplitude of a partial standing wave at an antinode (maximum) to the amplitude at an adjacent node (minimum), in an electrical transmission line.

The SWR is usually defined as a voltage ratio called the **VSWR**, for *voltage standing wave ratio*. For example, the VSWR value 1.2:1 denotes a maximum standing wave amplitude that is 1.2 times greater than the minimum standing wave value. It is also possible to define the SWR in terms of current, resulting in the ISWR, which has the same numerical value. The *power standing wave ratio* (PSWR) is defined as the square of the VSWR.

SWR is used as an efficiency measure for transmission lines, electrical cables that conduct radio frequency signals, used for purposes such as connecting radio transmitters and receivers with their antennas, and distributing cable television signals. A problem with transmission lines is that impedance mismatches in the cable tend to reflect the radio waves back toward the source end of the cable, preventing all the power from reaching the destination end. SWR measures the relative size of these reflections. An ideal transmission line would have an SWR of 1:1, with all the power reaching the destination and no reflected power. An infinite SWR represents complete reflection, with all the power reflected back down the cable. The SWR of a transmission line is measured with an instrument called an SWR meter, and checking the SWR is a standard part of installing and maintaining transmission lines.

Relationship to the reflection coefficient

The voltage component of a standing wave in a uniform transmission line consists of the forward wave (with amplitude V_f) superimposed on the reflected wave (with amplitude V_r).

Reflections occur as a result of discontinuities, such as an imperfection in an otherwise uniform transmission line, or when a transmission line is terminated with other than its characteristic impedance. The reflection coefficient Γ is defined thus:

$$\Gamma = \frac{V_r}{V_f}.$$

Γ is a complex number that describes both the magnitude and the phase shift of the reflection. The simplest cases, when the imaginary part of Γ is zero, are:

- Γ = − 1: maximum negative reflection, when the line is short-circuited,
- Γ = 0: no reflection, when the line is perfectly matched,
- Γ = + 1: maximum positive reflection, when the line is open-circuited.

For the calculation of VSWR, only the magnitude of Γ, denoted by ρ, is of interest. Therefore, we define

$$\rho = | \Gamma | .$$

At some points along the line the two waves interfere constructively, and the resulting amplitude V_{max} is the sum of their amplitudes:

$$V_{\max} = V_f + V_r = V_f + \rho V_f = V_f(1+\rho).$$

At other points, the waves interfere destructively, and the resulting amplitude V_{min} is the difference between their amplitudes:

$$V_{\min} = V_f - V_r = V_f - \rho V_f = V_f(1-\rho).$$

The voltage standing wave ratio is then equal to:

$$VSWR = \frac{V_{\max}}{V_{\min}} = \frac{1+\rho}{1-\rho}.$$

As ρ, the magnitude of Γ, always falls in the range [0,1], the VSWR is always ≥ +1.

The SWR can also be defined as the ratio of the maximum amplitude of the electric field strength to its minimum amplitude, i.e. E_{max} / E_{min}.

Further analysis

To understand the standing wave ratio in detail, we need to calculate the voltage (or, equivalently, the electrical field strength) at any point along the transmission line at any moment in time. We can begin with the forward wave, whose voltage as a function of time *t* and of distance *x* along the transmission line is:

$$V_f(x,t) = A\sin(\omega t - kx),$$

where A is the amplitude of the forward wave, ω is its angular frequency and k is the wave number (equal to ω divided by the speed of the wave). The voltage of the reflected wave is a similar function, but spatially reversed (the sign of x is inverted) and attenuated by the reflection coefficient ρ:

$$V_r(x,t) = \rho A \sin(\omega t + kx).$$

The total voltage V_t on the transmission line is given by the superposition principle, which is just a matter of adding the two waves:

$$V_t(x,t) = A \sin(\omega t - kx) + \rho A \sin(\omega t + kx).$$

Using standard trigonometric identities, this equation can be converted to the following form:

$$V_t(x,t) = A\sqrt{4\rho \cos^2 kx + (1-\rho)^2}\cos(\omega t + \phi),$$

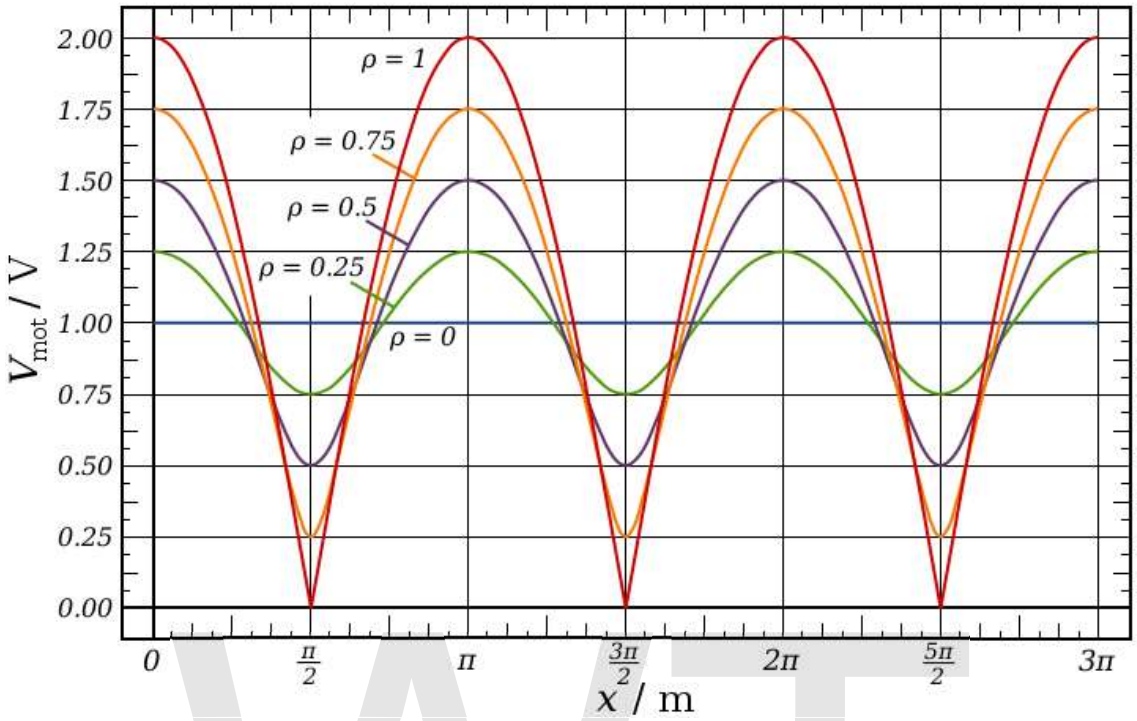
where

$$\tan\phi = \frac{(1+\rho)}{(1-\rho)}\cot(kx).$$

This form of the equation shows, if we ignore some of the details, that the maximum voltage over time V_{mot} at a distance x from the transmitter is the periodic function

$$V_{mot} = A\sqrt{4\rho \cos^2 kx + (1-\rho)^2}.$$

This varies with x from a minimum of $A(1-\rho)$ to a maximum of $A(1+\rho)$, as we saw in the earlier, simplified discussion. A graph of V_{mot} against x, in the case when $\rho = 0.5$, is shown below. The maximum and minimum V_{mot} in a periods are V_{min} and V_{max} and are the values used to calculate the SWR.



Standing wave ratio for a range of ρ. In this graph, A and k are set to unity.

It is important to note that this graph does *not* show the instantaneous voltage profile along the transmission line. It only shows the *maximum amplitude* of the oscillation at each point. The instantaneous voltage is a function of both time and distance, so could only be shown fully by a three-dimensional or animated graph.

Practical implications of SWR

The most common case for measuring and examining SWR is when installing and tuning transmitting antennas. When a transmitter is connected to an antenna by a feed line, the impedance of the antenna and feed line must match exactly for maximum energy transfer from the feed line to the antenna to be possible. The impedance of the antenna varies based on many factors including: the antenna's natural resonance at the frequency being transmitted, the antenna's height above the ground, and the size of the conductors used to construct the antenna.

When an antenna and feedline do not have matching impedances, some of the electrical energy cannot be transferred from the feedline to the antenna. Energy not transferred to the antenna is reflected back towards the transmitter. It is the interaction of these reflected waves with forward waves which causes standing wave patterns. Reflected power has three main implications in radio transmitters: Radio Frequency (RF) energy losses increase, distortion on transmitter due to reflected power from load and damage to the transmitter can occur.

Matching the impedance of the antenna to the impedance of the feed line is typically done using an antenna tuner. The tuner can be installed between the transmitter and the feed line, or between the feed line and the antenna. Both installation methods will allow the transmitter to operate at a low SWR, however if the tuner is installed at the transmitter, the feed line between the tuner and the antenna will still operate with a high SWR, causing additional RF energy to be lost through the feedline.

Many amateur radio operators consider any impedance mismatch a serious matter. Power loss will increase as the SWR increases. For example, a dipole antenna tuned to operate at 3.75 MHz—the center of the 80 meter amateur radio band—will exhibit an SWR of about 6:1 at the edges of the band. However, if the antenna is fed with 250 feet of RG-8A coax, the loss due to standing waves is only 2.2dB, which may seem like a small loss, but is on a logarithmic scale. If running a typical 100W transmitter on the HF band, 2.2dB of loss would reduce the output power to 60W. That is a 40% reduction in power. Feed line loss typically increases with frequency, so VHF and above antennas must be matched closely to the feedline. The same 6:1 mismatch to 250 feet of RG-8A coax would incur 10.8dB of loss at 146 MHz. However, a length of 250 feet would not likely be used for 2m VHF radios. Antennas for the 80m band frequently involve large or complex designs typically mounted on a tall tower with great distances needed between buildings and thus the transmitter. VHF requires a much smaller antenna, and unless being used on a high powered repeater, does not have a very tall tower. The most common usage of 2m band is mobile single or dual band VHF or VHF/UHF mobiles. Also in part due to the typical output power of a VHF band is 50W, due to the FCC requirement of RF exposure evaluations needing to be conducted on power greater than 50W in the 2m band. This 50W with the 250 feet of cable would be reduced to a tiny 5W with 10dB of loss. On the more rare occasion where a long transmission line is needed for 146Mhz, a higher quality low loss transmission line would be used instead of the relatively cheap RG-8A.

Chapter 11

Stroke Ratio

In a reciprocating piston engine, the **stroke ratio**, defined by either **bore/stroke ratio** or **stroke/bore ratio**, is a term which is used to describe the ratio between the diameter of the cylinder bore and the length of the piston stroke within its cylinders. This can be used for either an internal combustion engine, such as a petrol- or diesel engine, where the fuel is burned within the cylinders of the engine, or external combustion engine, such as a steam engine, where the combustion of the fuel takes place *outside* the working cylinders of the engine.

While the stroke ratio can provide insight into the goals of an engine's designer, it has no direct effect on the speed at which an engine reaches maximum torque: holding displacement constant, lengthening the crank throw reduces the piston area by an exactly corresponding amount.

Conventions

In a piston engine, there are two different ways of describing the *stroke ratio* of its cylinders, and these are often mixed together causing confusion. These are: *bore/stroke* ratio, and *stroke/bore* ratio.

Bore/stroke ratio

Bore/stroke is the most commonly used term, which is mainly used in the North America, Europe, United Kingdom, Asia, Australia, and some other countries.

The diameter of the cylinder bore is divided by the length of the piston stroke to give the ratio.

Stroke/bore ratio

Stroke/bore ratio is generally more rare than **bore/stroke** ratio, but is used in some countries, like in Finland for example.

The length of the piston stroke is divided by the diameter of the cylinder bore to give the ratio.

Stroke/bore ratio is similar to the bore/stroke ratio with the following exception:

When stroke/bore value is over 1:1 the engine is long-stroke or undersquare and when the stroke/bore value is under 1:1 the engine is short-stroke or oversquare. The square engine has a value of 1:1 in both cases.

For example an engine with 110 millimetres (4.33 in) stroke and 80 millimetres (3.15 in) bore, stroke/bore value 1.375, is an undersquare or long-stroke engine. An engine that has 70 millimetres (2.76 in) stroke and 100 millimetres (3.94 in) bore, stroke/bore value 0.7, is oversquare or short-stroke.

Square, undersquare and oversquare engines

The following terms describe the naming conventions for the various configurations of the relationship ratio between the diameter of the cylinder bore and the length of the piston stroke within the cylinders of a piston engine.

Square engine

An engine is described as a **square engine** when it has equal bore *and* stroke dimensions, giving a bore/stroke value of exactly 1:1.

For example an engine which has 95 millimetres (3.74 in) bore, and an identical 95 millimetres (3.74 in) stroke, has a bore/stroke value of:

95 mm / 95 mm = 1.00

Usually engines that have a bore/stroke ratio of 0.95 to 1.04 are referred as square engines.

Square engine examples

The Volkswagen Group W16 engine as used in the Bugatti Veyron is an example of a square engine - with an identical bore and stroke of 86.0 millimetres (3.39 in). Another example of a square engine is the 1970s Ford 400M with a 4.00" bore and stroke.

The Mercedes-Benz M117 engine with a displacement of 5547 cubic centimeters is an example of a nearly square engine with a bore of 96.5 millimeters and a stroke of 94.8 millimeters.

The Cadillac 500-V8 manufactured from 1970-1976 is a nearly square engine with a 4.300 inch bore and a 4.304 inch stroke.

Nissan's SR20DE is a square engine, with a bore and stroke of 86mm.

The 1973-1976 Kawasaki Z-1 and KZ(Z)900 had a 66 mm bore and a 66 mm stroke, also making it a square engine.

Oversquare, or short-stroke engine

An engine is described as **oversquare** or **short-stroke** if its cylinders have a greater bore diameter than its stroke length - giving a ratio value of greater than 1:1.

For example an engine which has 100 millimetres (3.94 in) bore and 80 millimetres (3.15 in) stroke has a bore/stroke value of:

100 mm / 80 mm = 1.25:1

An oversquare engine allows for more and larger valves in the head of the cylinder, lower friction losses (due to the reduced distance travelled during each engine rotation) and lower crank stress (due to the lower peak piston speed relative to engine speed). Because these characteristics favor higher engine speeds, oversquare engines are often tuned to develop peak torque at a relatively high speed.

The reduced stroke length allows for a shorter cylinder and sometimes a shorter connecting rod, generally making oversquare engines less tall than undersquare engines of similar engine displacement but wider and longer (for engines with vertical cylinder axes).

By changing the crankshaft and modifying the connecting rod(s), piston(s) and/or engine block an engine can be "de-stroked". This reduces the displacement and consequently the torque of the engine, but can allow it to run at higher speeds and in fact develop greater peak power.

Oversquare engine examples

Oversquare engines are extremely common, including both Chevrolet and Ford small block V8s. Most Boxer (horizontally-opposed) engines (such as those built by Volkswagen, Porsche, and Subaru) feature oversquare designs since any increase in stroke length would result in twice the increase in overall engine size.

This is particularly crucial in Subaru's front-engine layout, where the steering angle of the front wheels is limited largely by the size of the engine. Although oversquare engines have a reputation for being high-strung, low-torque machines, the Subaru EJ engine develops peak torque at speeds as low as 3200 RPM.

Extreme examples of oversquare engine designs are found in Formula One race cars, whose rules tightly limit displacement and thereby require that power be achieved

through high engine speeds. Stroke ratios of 2.5:1 are typical, with engines capable of 19,000 RPM.

Undersquare, or long-stroke engine

An engine is described as **undersquare** or **long-stroke** if its cylinders have a smaller bore (width, diameter) than its stroke (length of piston travel) - giving a ratio value of less than 1:1.

For example an engine which has 90 millimetres (3.54 in) bore and 120 millimetres (4.72 in) stroke has a bore/stroke value of:

90 mm / 120 mm = 0.75:1

At a given engine speed, a longer stroke increases engine friction (since the piston travels a greater distance per stroke) and increases stress on the crankshaft (due to the higher peak piston speed). The smaller bore also reduces the area available for valves in the cylinder head, requiring them to be smaller or fewer in number. Because these factors favor lower engine speeds, undersquare engines are most often tuned to develop peak torque at relatively low speeds.

An undersquare engine will typically be more compact in the directions perpendicular to piston travel but larger in the direction parallel to piston travel.

An engine can be "stroked" by replacing the crankshaft with a so-called "stroker" crankshaft and modifying the connecting rod(s), piston(s) or engine block to accommodate the increased piston travel. This increases the displacement and therefore the torque of the engine, but may reduce the peak speed at which it is safe to run.

Undersquare engine examples

Many inline engines, particularly those mounted transversely in front-wheel-drive cars, utilize an undersquare design. The smaller bore allows for a shorter engine that increases room available for the front wheels to steer. Examples of this include many Volkswagen, Honda, and Mazda engines. Some rear-wheel-drive cars that borrow engines from front-wheel-drive cars (such as the Mazda Miata) use an undersquare design.

Despite their reputation as low-speed torque machines, some undersquare engines are designed for quite high speeds. The Honda Integra Type R's B18C5 engine has one of the highest redlines of any production engine, yet features an undersquare design. The 2011 Ford Coyote engine is a modern undersquare engine with a 7,000 rpm redline.

Many British automobile companies used undersquare designs through the 1950s, largely because of a motor tax system that taxed cars by their cylinder bore. This includes the Austin A-Series engine, and many Nissan derivatives.

The Chrysler Slant-6 engine, in its most common 225 cubic inch (3.7 litre) version, is a massively undersquare engine, with a 86 millimetres (3.39 in) bore and a 105 millimetres (4.13 in) stroke, producing most of its power right on the peak of its torque curve. The Achilles heel of this engine, otherwise known for its exceptional durability, is being over-revved by inexperienced drivers. Red line for a factory engine is under 4,500 revolutions per minute (rpm); red line with aftermarket connecting rods is about 5,500 rpm. On the other hand, a well-maintained Slant-6 can be made to idle as low as 75 rpm (though this is *not* a recommended speed - neither the alternator nor the oil pump will function adequately). In some circles, the Slant-6 is nicknamed "The Stump-Puller" for its diesel engine-like low-speed torque. Appropriate gearing and driving skill is required for performance use.

Willys also used mostly undersquare engines; in fact the L134 and F134 engines, with their fairly small 79.4 millimetres (3.13 in) bore and 111.1 millimetres (4.37 in) stroke, are probably the most undersquare engines ever built (for Jeeps).

The Dodge Power Wagon, among other vehicles, used a straight-six Chrysler Flathead engine of 230 cubic inches (3.8 litre) with a bore of 83 millimetres (3.27 in) and a stroke of 117 millimetres (4.61 in), yielding a substantially under-square stroke ratio of 0.70.

Virtually all piston aircraft engines used in military aircraft were long stroke engines. The PW R-2800, Wright R-3350, PW R-4360, Rolls-Royce Merlin (1650), Allison V-1710, and Hispano-Suiza 12Y-Z are only a few of more than a hundred examples.

Chapter 12

Thrust-to-Weight Ratio

Thrust-to-weight ratio is a ratio of thrust to weight of a rocket, jet engine, propeller engine, or a vehicle propelled by such an engine. It is a dimensionless quantity and is an indicator of the performance of the engine or vehicle.

The instantaneous thrust-to-weight ratio of a vehicle varies continually during operation due to progressive consumption of fuel or propellant. The thrust-to-weight ratio based on initial thrust and weight is often published and used as a figure of merit for quantitative comparison of the initial performance of vehicles.

Calculation

The **thrust-to-weight ratio** can be calculated by dividing the thrust (in SI units – in newtons) by the weight (in newtons) of the engine or vehicle. It is a dimensionless quantity.

For valid comparison of the initial thrust-to-weight ratio of two or more engines or vehicles, thrust must be measured under controlled conditions.

Aircraft

The **thrust-to-weight ratio** and wing loading are the two most important parameters in determining the performance of an aircraft. For example, the thrust-to-weight ratio of a combat aircraft is a good indicator of the manoeuvrability of the aircraft.

The thrust-to-weight ratio varies continually during a flight. Thrust varies with throttle setting, airspeed, altitude and air temperature. Weight varies with fuel burn and changes of payload. For aircraft, the quoted thrust-to-weight ratio is often the maximum static thrust at sea-level divided by the maximum takeoff weight.

In cruising flight, the thrust-to-weight ratio of an aircraft is the inverse of the lift-to-drag ratio because thrust is equal to drag, and weight is equal to lift.

$$\left(\frac{T}{W}\right)_{cruise} = \frac{1}{\left(\frac{L}{D}\right)_{cruise}}$$

Propeller-driven aircraft

For propeller-driven aircraft, the thrust-to-weight ratio can be calculated as follows:

$$\frac{T}{W} = \left(\frac{\eta_p}{V}\right)\left(\frac{P}{W}\right)$$

where η_p is propulsive efficiency at true airspeed V

P is engine power

Rockets

The **thrust-to-weight ratio** of a rocket, or rocket-propelled vehicle, is an indicator of its acceleration expressed in multiples of gravitational acceleration g.

Rockets and rocket-propelled vehicles operate in a wide range of gravitational environments, including the *weightless* environment. It is customary to calculate the thrust-to-weight ratio using initial gross weight at sea-level on earth. This is sometimes called *Thrust-to-Earth-weight ratio*. The thrust-to-Earth-weight ratio of a rocket, or rocket-propelled vehicle, is an indicator of its acceleration expressed in multiples of earth's gravitational acceleration, g_0.

The thrust-to-weight ratio of an engine is larger for the bare engine than for the whole launch vehicle. The thrust-to-weight ratio of a bare engine is of use since it determines the maximum acceleration that any vehicle using that engine could theoretically achieve with minimum propellant and structure attached.

For a takeoff from the surface of the earth using thrust and no aerodynamic lift, the thrust-to-weight ratio for the whole vehicle has to be *more than* **one**. In general, the thrust-to-weight ratio is numerically equal to the *g-force* that the vehicle can generate. Provided the vehicle's *g-force* exceeds local gravity (expressed as a multiple of g_0) then takeoff can occur.

Many factors affect a thrust-to-weight ratio, and it typically varies over the flight with the variations of thrust due to speed and altitude, and the weight due to the remaining propellant and payload mass. The main factors that affect thrust include freestream air temperature, pressure, density, and composition. Depending on the engine or vehicle under consideration, the actual performance will often be affected by buoyancy and local gravitational field strength.

Examples

The Russian-made RD-180 rocket engine (which powers Lockheed Martin's Atlas V) produces 3,820 kN of sea-level thrust and has a dry mass of 5,307 kg. Using the Earth surface gravitational field strength of 9.807 m/s², the sea-level thrust-to-weight ratio is computed as follows: (1 kN = 1000 N = 1000 kg · m/s²)

$$\frac{T}{W} = \frac{3,820 \text{ kN}}{(5,307 \text{ kg})(9.807 \text{ m/s}^2)} = 0.07340 \frac{\text{kN}}{\text{N}} = 73.40 \frac{\text{N}}{\text{N}} = 73.40$$

Aircraft

Vehicle	T/W	Scenario
Concorde	.373	
English Electric Lightning	0.63	maximum takeoff weight, no reheat
F-15 Eagle	1.04	nominally loaded
F-16 Fighting Falcon	1.096	
Hawker Siddeley Harrier	1.1	
Dassault Rafale	1.13	
Mikoyan MiG-29	1.01	
Eurofighter Typhoon	1.25	
English Electric Lightning	~1.2	light weight, full reheat
Space Shuttle	1.5	Take-off
F-15 Eagle	~1.6	light weight, full afterburner
Space Shuttle	3	Peak (throttled back for astronaut comfort)

Note that the above duct engined aircraft do not have a thrust-to-weight ratio greater than one at maximum take-off weight, whereas rockets do.

Jet and Rocket Engines

Jet or Rocket engine	Mass, kg	Jet or rocket thrust, kN	Thrust-to-weight ratio
RD-0410 nuclear rocket engine	2000	35.2	1.8
J-58 (SR-71 Blackbird jet engine)	2722	150	5.2
Concorde's Rolls-Royce/Snecma Olympus 593 turbojet with reheat	3175	169.2	5.4
RD-0750 rocket engine, three-propellant mode	4621	1413	31.2
RD-0146 rocket engine	260	98	38.5
Space Shuttle's SSME rocket engine	3177	2278	73.2

RD-180 rocket engine	5393	4152	78.6
F-1 (Saturn V first stage)	8391	7740.5	94.1
NK-33 rocket engine	1222	1638	136.8

Fighter Aircraft

Table a: Thrust To Weight Ratios, Fuels Weights, and Weights of Different Fighter Planes

Specifications / Fighters	F-15K	F-15C	MiG-29K	MiG-29B	JF-17	J-10	F-35A	F-35B	F-35C	F-22
Engine(s) Thrust Maximum (lbf)	58,320 (2)	46,900 (2)	39,682 (2)	36,600 (2)	18,300 (1)	27,557 (1)	39,900 (1)	39,900 (1)	39,900 (1)	70,000 (2)
Aircraft Weight Empty (lb)	37,500	31,700	28,050	24,030	14,520	20,394	29,300	32,000	34,800	43,340
Aircraft Weight Full fuel (lb)	51,023	45,574	39,602	31,757	19,650	28,760	47,780	46,003	53,800	61,340
Aircraft Weight Max Take-off load (lb)	81,000	68,000	49,383	40,785	28,000	42,500	70,000	60,000	70,000	83,500
Total fuel weight (lb)	13,523	13,874	11,552	07,727	05,130	08,366	18,480	14,003	19,000	18,000
T/W ratio (Thrust / AC weight full fuel)	1.14	1.03	1.00	1.15	0.93	0.96	0.84	0.87	0.74	1.14

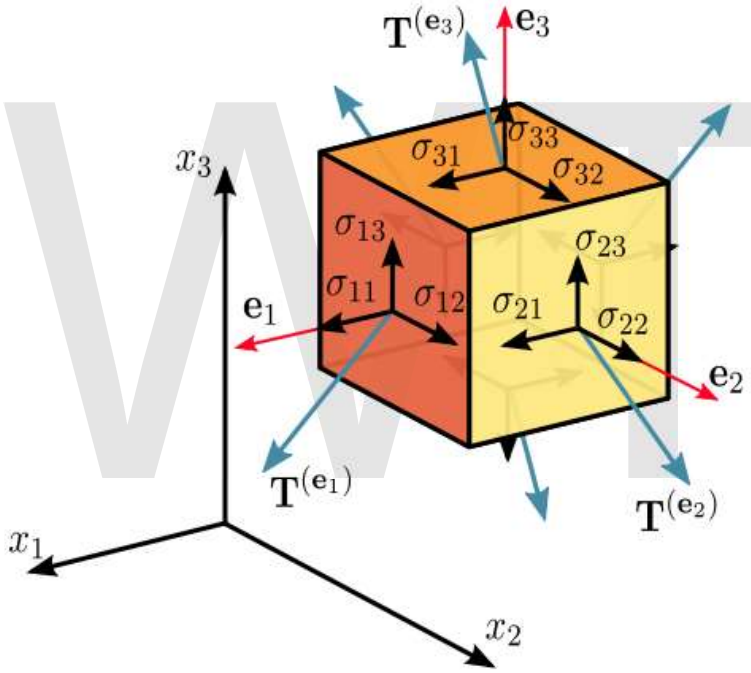
Table b: Thrust To Weight Ratios, Fuels Weights, and Weights of Different Fighter Planes (In International System)

In International System	F-15K	F-15C	MiG-29K	MiG-29B	JF-17	J-10	F-35A	F-35B	F-35C	F-22
Engine(s) Thrust Maximum (kgf)	26,456 (2)	21,274 (2)	18,000 (2)	16,600 (2)	08,300 (1)	12,500 (1)	18,098 (1)	18,098 (1)	18 098 (1)	31,764 (2)
Aircraft Weight Empty (kg)	17,010	14,379	12,723	10,900	06,586	09,250	13,290	14,515	15,785	19,673
Aircraft Weight Full fuel (kg)	23,143	20,671	17,963	14,405	08,886	13,044	21,672	20,867	24,403	27,836
Aircraft Weight Max Take-off load (kg)	36,741	30,845	22,400	18,500	12,700	19,277	31,752	27,216	31,752	37,869
Total fuel weight (kg)	06,133	06,292	05,240	03,505	02,300	03,794	08,382	06,352	08,618	08,163
T/W ratio (Thrust / AC weight full fuel)	1.14	1.03	1.00	1.15	0.93	0.96	0.84	0.87	0.74	1.14

- Fuel density used in calculations = 0.803 Kilograms/Liter
- The Number inside ( ) brackets is the Number of Engine(s).
- Engines powering F-15K are the Pratt & Whitney Engines, not General Electric's.
- MiG-29K's empty weight is an estimate.
- JF-17's Engine rating is of RD-93.
- JF-17 if mated with its engine WS-13, and if that engine gets its promised 18,969 lb then the T/W ratio becomes 0.97
- J-10's empty weight & fuel weight is an estimate.
- J-10's Engine rating is of AL-31FN.
- J-10 if mated with its engine WS-10A, and if that engine gets its promised 132 KN(29,674 lbf) then the T/W ratio becomes 1.03

Chapter 13

Introduction to Tensor



Stress, a second-order tensor. The tensor's components, in a three-dimensional Cartesian coordinate system, form the matrix $\sigma = \left[\mathbf{T}^{(\mathbf{e}_1)}\mathbf{T}^{(\mathbf{e}_2)}\mathbf{T}^{(\mathbf{e}_3)}\right] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$ whose columns are the forces acting on the $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ faces of the cube.

Tensors are geometric entities introduced into mathematics and physics to extend the notion of scalars, geometric vectors, and matrices to higher orders. Tensors were first conceived by Tullio Levi-Civita and Gregorio Ricci-Curbastro, who continued the earlier work of Bernhard Riemann and Elwin Bruno Christoffel and others, as part of the

absolute differential calculus. The concept enabled an alternative formulation of the intrinsic differential geometry of a manifold in the form of the Riemann curvature tensor.

Many physical quantities are naturally regarded not as vectors themselves, but as correspondences between one set of vectors and another. For example, the stress tensor **T** takes a direction **v** as input and produces the stress $\mathbf{T}^{(v)}$ on the surface normal to this vector as output and so expresses a relationship between these two vectors. Because they express a relationship between vectors, tensors themselves are independent of a particular choice of coordinate system. It is possible to represent a tensor by examining what it does to a coordinate basis or frame of reference; the resulting quantity is then an organized multi-dimensional array of numerical values. The coordinate-independence of a tensor then takes the form of a "covariant" transformation law that relates the array computed in one coordinate system to that computed in another one.

The order (or degree) of a tensor is the dimensionality of the array needed to represent it. A number is a 0-dimensional array, so it is sufficient to represent a scalar, a 0th-order tensor. A coordinate vector, or 1-dimensional array, can represent a vector, a 1st-order tensor. A 2-dimensional array, or square matrix, is then needed to represent a 2nd-order tensor. In general, an order-*k* tensor can be represented as a *k*-dimensional array of components. The order of a tensor is the number of indices necessary to refer unambiguously to an individual component of a tensor.

History

The concepts of later tensor analysis arose from the work of C. F. Gauss in differential geometry, and the formulation was much influenced by the theory of algebraic forms and invariants developed in the middle of the nineteenth century. The word "tensor" itself was introduced in 1846 by William Rowan Hamilton to describe something different from what is now meant by a tensor. The contemporary usage was brought in by Woldemar Voigt in 1898.

Tensor calculus was developed around 1890 by Gregorio Ricci-Curbastro (also called just Ricci) under the title *absolute differential calculus*, and originally presented by Ricci in 1892. It was made accessible to many mathematicians by the publication of Ricci and Tullio Levi-Civita's 1900 classic text *Méthodes de calcul différentiel absolu et leurs applications* (Methods of absolute differential calculus and their applications) (Ricci & Levi-Civita 1900) (in French; translations followed).

In the 20th century, the subject came to be known as *tensor analysis*, and achieved broader acceptance with the introduction of Einstein's theory of general relativity, around 1915. General relativity is formulated completely in the language of tensors. Einstein had learned about them, with great difficulty, from the geometer Marcel Grossmann. Levi-Civita then initiated a correspondence with Einstein to correct mistakes Einstein had made in his use of tensor analysis. The correspondence lasted 1915–17, and was characterized by mutual respect, with Einstein at one point writing:

> I admire the elegance of your method of computation; it must be nice to ride through these fields upon the horse of true mathematics while the like of us have to make our way laboriously on foot.

Tensors were also found to be useful in other fields such as continuum mechanics. Some well-known examples of tensors in differential geometry are quadratic forms such as metric tensors, and the Riemann curvature tensor. The exterior algebra of Hermann Grassmann, from the middle of the nineteenth century, is itself a tensor theory, and highly geometric, but it was some time before it was seen, with the theory of differential forms, as naturally unified with tensor calculus. The work of Élie Cartan made differential forms one of the basic kinds of tensor fields used in mathematics.

From about the 1920s onwards, it was realised that tensors play a basic role in algebraic topology (for example in the Künneth theorem). Correspondingly there are types of tensors at work in many branches of abstract algebra, particularly in homological algebra and representation theory. Multilinear algebra can be developed in greater generality than for scalars coming from a field, but the theory is then certainly less geometric, and computations more technical and less algorithmic. Tensors are generalized within category theory by means of the concept of monoidal category, from the 1960s.

Definition

There are several different approaches to defining tensors. Although seemingly different, the approaches just describe the same geometric concept using different languages and at different levels of abstraction.

... as multidimensional arrays

Just as a scalar is described by a single number and a vector can be described by a list of numbers, tensors in general can be considered as a multidimensional array of numbers, which are known as its "scalar components" or simply "components." The entries of such an array are symbolically denoted by the name of the tensor with indices giving the position in the array. The total number of indices is equal to the dimension of the array and is called the *order* or the *rank* of the tensor. For example, the entries (also called *components*) of an order 2 tensor T would be denoted T_{ij}, where i and j are indices running from 1 to the dimension of the related vector space.

Just like the components of a vector change when we change the basis of the vector space, the entries of a tensor also change under such a transformation. Recall that the components of a vector can respond in two distinct ways to a change of basis,

$$\hat{e}_i = \sum_j R_i^j e_j = R_i^j e_j,$$

where R_{ji} is a matrix and in the second expression the summation sign was suppressed (a notational convenience introduced by Einstein that will be used throughout). The components, v^i, of a regular (or column) vector, **v**, transform with the inverse of the matrix *R*,

$$\hat{v}^i = (R^{-1})^i_j v^j,$$

where the hat denotes the components in the new basis. While the components, w_i, of a covector or (row vector), **w** transform with the matrix R itself,

$$\hat{w}_i = R^j_i w_j.$$

The components of a tensor transform in a similar manner with a transformation matrix for each index. If an index transforms like a vector with the inverse of the basis transformation, it is called *contravariant* and is traditionally denoted with an upper index, while an index that transforms with the basis transformation itself is called *covariant* and is denoted with a lower index. The "transformation law" for a rank *m* tensor with *n* contravariant indices and *m-n* covariant indices is thus given as,

$$\hat{T}^{i_1,\ldots,i_n}_{i_{n+1},\ldots,i_m} = (R^{-1})^{i_1}_{j_1} \cdots (R^{-1})^{i_n}_{j_n} R^{j_{n+1}}_{i_{n+1}} \cdots R^{j_m}_{i_m} T^{j_1,\ldots,j_n}_{j_{n+1},\ldots,j_m}.$$

Such a tensor is said to be of order or *type* (*n*,*m-n*).

The definition of a tensor as a multidimensional array satisfying a "transformation law" traces back to the work of Ricci. Nowadays, this definition is still popular in physics and engineering text books.

Tensor fields

In many applications, especially in differential geometry and physics, it is natural to consider the components of a tensor to be functions. This was, in fact, the setting of Ricci's original work. In modern mathematical terminology such an object is called a tensor field, but they are often simply referred to as tensors themselves.

In this context the defining transformation law takes a different form. The "basis" for the tensor field is determined by the coordinates of the underlying space, and the defining transformation law is expressed in terms of partial derivatives of the coordinate functions, $\bar{x}_i(x_1, \ldots, x_k)$, defining a coordinate transformation,

$$\hat{T}^{i_1\ldots i_n}_{i_{n+1}\ldots i_m}(\bar{x}_1, \ldots, \bar{x}_k) = \frac{\partial \bar{x}^{i_1}}{\partial x^{j_1}} \cdots \frac{\partial \bar{x}^{i_n}}{\partial x^{j_n}} \frac{\partial x^{j_{n+1}}}{\partial \bar{x}^{i_{n+1}}} \cdots \frac{\partial x^{j_m}}{\partial \bar{x}^{i_m}} T^{j_1\ldots j_n}_{j_{n+1}\ldots j_m}(x_1, \ldots, x_k).$$

... as multilinear maps

A downside to the definition of a tensor using the multidimensional array approach is that it is not apparent from the definition that the defined object is indeed basis independent, as is expected from an intrinsically geometric object. Although it is possible to show that transformation laws indeed ensure independence from the basis, sometimes a more intrinsic definition is preferred. One approach is to define a tensor as a multilinear map. In that approach a type (*n*,*m*) tensor *T* is defined as a map,

$$T: \underbrace{V^* \times \cdots \times V^*}_{\text{n copies}} \times \underbrace{V \times \cdots \times V}_{\text{m copies}} \to \mathbf{R},$$

where *V* is a vector space and *V** is the corresponding dual space of covectors, which is linear in each of its arguments.

By applying a multilinear map *T* of type (*n*,*m*) to a basis $\{e_j\}$ for *V* and a canonical cobasis $\{\varepsilon^i\}$ for *V**,

$$T^{i_1 \dots i_n}_{j_1 \dots j_m} \equiv T(\varepsilon^{i_1}, \dots, \varepsilon^{i_n}, e_{j_1}, \dots, e_{j_m}),$$

a *n*+*m* dimensional array of components can be obtained. A different choice of basis will yield different components. But, because *T* is linear in all of its arguments, the components satisfy the tensor transformation law used in the multilinear array definition. The multidimensional array of components of *T* thus form a tensor according to that definition. Moreover, such an array can be realised as the components of some multilinear map *T*. This motivates viewing multilinear maps as the intrinsic objects underlying tensors.

This approach, defining tensors as multilinear maps, is popular in modern differential geometry textbooks and more mathematically inclined physics textbooks.

Abstract index notation

The abstract index notation is a way to write tensors such that the indices are no longer thought of as numerical, but rather are indeterminates. The abstract index notation captures the expressiveness of indices and the basis-independence of index-free notation.

Operations

There are a number of basic operations that may be conducted on tensors that again produce a tensor. The linear nature of tensor implies that two tensors of the same type may be added together, and that tensors may be multiplied by a scalar with results analogous to the scaling of a vector. On components, this operations are simply performed component for component. These operations do not change the type of the tensor, however there also exist operations that change the type of the tensors.

Raising or lowering an index

When a vector space is equipped with an inner product (or *metric* as it often called in this context), there exist operations that convert a contravariant (upper) index into a covariant (lower) index and vice versa. A metric itself is a (symmetric) (0,2)-tensor, it is thus possible to contract an upper index of a tensor with one of lower indices of the metric. This produces a new tensor with the same index structure as the previous, but with lower index in the position of the contracted upper index. This operation is quite graphically known as *lowering an index*.

Conversely, a metric has an inverse which is a (2,0)-tensor. This inverse metric can be contracted with a lower index to produce an upper index. This operation is called *raising an index*.

Applications

Tensors are important in physics and engineering. In the field of diffusion tensor imaging, for instance, a tensor quantity that expresses the differential permeability of organs to water in varying directions is used to produce scans of the brain; in this technique tensors are in effect made visible. That application is of a tensor of second order. While such uses of tensors are the most frequent, tensors of higher order also matter in many fields.

Continuum mechanics

Important examples are provided by continuum mechanics. The stresses inside a solid body or fluid are described by a tensor. The stress tensor and strain tensor are both second order tensors, and are related in a general linear elastic material by a fourth-order elasticity tensor. In detail, the tensor quantifying stress in a 3-dimensional solid object has components that can be conveniently represented as a 3×3 array. The three faces of a cube-shaped infinitesimal volume segment of the solid are each subject to some given force. The force's vector components are also three in number. Thus, 3×3, or 9 components are required to describe the stress at this cube-shaped infinitesimal segment. Within the bounds of this solid is a whole mass of varying stress quantities, each requiring 9 quantities to describe. Thus, a second order tensor is needed.

If a particular surface element inside the material is singled out, the material on one side of the surface will apply a force on the other side. In general, this force will not be orthogonal to the surface, but it will depend on the orientation of the surface in a linear manner. This is described by a tensor of type (2,0), in linear elasticity, or more precisely by a tensor field of type (2,0), since the stresses may vary from point to point.

Applications of tensors of order > 2

The concept of a tensor of order two is often conflated with that of a matrix. Tensors of higher order do however capture ideas important in science and engineering, as has been

shown successively in numerous areas as they develop. This happens, for instance, in the field of computer vision, with the trifocal tensor generalizing the fundamental matrix.

The field of nonlinear optics studies the changes to material polarization density under extreme electric fields. The polarization waves generated are related to the generating electric fields through the nonlinear susceptibility tensor. If the polarization **P** is not linearly proportional to the electric field **E**, the medium is termed *nonlinear*. To a good approximation (for sufficiently weak fields, assuming no permanent dipole moments are present), **P** is given by a Taylor series in **E** whose coefficients are the nonlinear susceptibilities:

$$\frac{P_i}{\varepsilon_0} = \sum_j \chi_{ij}^{(1)} E_j + \sum_{jk} \chi_{ijk}^{(2)} E_j E_k + \sum_{jk\ell} \chi_{ijk\ell}^{(3)} E_j E_k E_\ell + \cdots$$

Here $\chi^{(1)}$ is the linear susceptibility, $\chi^{(2)}$ gives the Pockels effect and second harmonic generation, and $\chi^{(3)}$ gives the Kerr effect. This expansion shows the way higher-order tensors arise naturally in the subject matter.

Generalizations

Tensor densities

It is also possible for a tensor field to have a "density". A tensor with density *r* transforms as an ordinary tensor under coordinate transformations, except that it is also multiplied by the determinant of the Jacobian to the r^{th} power. Invariantly, in the language of multilinear algebra, one can think of tensor densities as multilinear maps taking their values in a density bundle such as the (1-dimensional) space of *n*-forms (where *n* is the dimension of the space), as opposed to taking their values in just **R**. Higher "weights" then just correspond to taking additional tensor products with this space in the range.

In the language of vector bundles, the determinant bundle of the tangent bundle is a line bundle that can be used to 'twist' other bundles *r* times. While locally the more general transformation law can indeed be used to recognise these tensors, there is a global question that arises, reflecting that in the transformation law one may write either the Jacobian determinant, or its absolute value. Non-integral powers of the (positive) transition functions of the bundle of densities make sense, so that the weight of a density, in that sense, is not restricted to integer values.

Restricting to changes of coordinates with positive Jacobian determinant is possible on orientable manifolds, because there is a consistent global way to eliminate the minus signs; but otherwise the line bundle of densities and the line bundle of *n*-forms are distinct.

Chapter 14

Einstein Notation and Abstract Index Notation

Einstein notation

In mathematics, especially in applications of linear algebra to physics, the **Einstein notation** or **Einstein summation convention** is a notational convention useful when dealing with coordinate formulae. It was introduced by Albert Einstein in 1916.

According to this convention, when an index variable appears twice in a single term, once in an upper (superscript) and once in a lower (subscript) position, it implies that we are summing over all of its possible values. In typical applications, the index values are 1,2,3 (representing the three dimensions of physical Euclidean space), or 0,1,2,3 or 1,2,3,4 (representing the four dimensions of space-time, or Minkowski space), but they can have any range, even (in some applications) an infinite set. Thus in three dimensions

$$y = c_i x^i$$

actually means

$$y = \sum_{i=1}^{3} c_i x^i = c_1 x^1 + c_2 x^2 + c_3 x^3.$$

The upper indices are not exponents, but instead different axes. Thus, for example, x^2 should be read as "x-two", not "x squared", and corresponds to the traditional y-axis.

Abstract index notation is a way of presenting the summation convention so that it is made clear that it is independent of coordinates.

In general relativity, the Greek alphabet and the Roman alphabet are used to distinguish whether summing over 1,2,3 or 0,1,2,3 (usually Roman, *i*, *j*, ... for 1,2,3 and Greek, μ, ν, ... for 0,1,2,3). As in sign conventions, the convention used in practice varies: Roman and Greek may be reversed.

When there is a fixed basis, one can work with only subscripts, but in general one must distinguish between superscripts and subscripts; see below.

In some fields, Einstein notation is referred to simply as index notation, or indicial notation. The use of the implied summation of repeated indices is also referred to as the *Einstein Sum Convention.*

Introduction

The basic idea of Einstein notation is that a covector and a vector can form a scalar:

$$y = c_1 x^1 + c_2 x^2 + c_3 x^3 + \cdots + c_n x^n$$

This is typically written as an explicit sum:

$$y = \sum_{i=1}^{n} c_i x^i$$

A scalar is invariant under transformations of basis. When the basis is changed, the components of a vector change by a linear transformation described by a matrix, while the covector changes by the inverse matrix. This is designed to guarantee that the linear function associated with the covector, the sum above, is the same no matter what the basis is. Since it is only this sum which is invariant under changes of basis, not the individual terms in the sum, this led Einstein to propose the convention that repeated indices imply the sum:

$$y = c_i x^i$$

In Einstein notation, covector indices are subscripts and vector indices are superscripts. The position of the index has a specific meaning. It is important, of course, not to confuse a superscript with an exponent—all the relations with superscripts and subscripts are linear, they involve no power higher than the first. Here, the superscripted *i* above the symbol *x* represents an integer-valued index running from 1 to *n*.

The virtue of Einstein notation is that it represents the invariant quantities with a simple notation.

Vector representations

First, we can use Einstein notation in linear algebra to distinguish easily between vectors and covectors: upper indices a^i are used to label *components* (coordinates) of vectors, while lower indices a_i are used to label components of covectors. However, vectors themselves (not their components) have lower indices e_i, and covectors have upper indices e^i. Given a vector space V and its dual space V^*, one indexes vectors (elements of V) with subscripts, as in $v_i \in V$, and covectors with superscripts, as in $w^i \in V^*$.

However, the coordinates of vectors and covectors follow the opposite convention: if e_i is a basis for V and e^i is the dual basis for V^*, then vectors are expressed as:

$$v = a^i e_i = \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^n \end{bmatrix}$$

and covectors are expressed as

$$w = a_i e^i = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

This is because a component of a vector (one of its coordinates, in some basis) is the *value* of a *covector*: the coefficient of e_i is the value of the corresponding covector in the dual basis: $a^i = e^i(v)$. Note that e^i is a covector, but a^i is a scalar. In other words, since basis vectors are given lower indices and coordinates are labeled with upper indices, summation notation suggests pairing them (in the obvious way) to express the vector.

In terms of covariance and contravariance of vectors, lower indices represent 'components' of covariant vectors (covectors), while upper indices represent components of contravariant vectors (vectors): they transform covariantly (resp., contravariantly) with respect to change of basis.

A particularly confusing notation is to use the same letter both for a (co)vector and its components, as in:

$$v = v^i e_i = \begin{bmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{bmatrix}$$
$$w = w_i e^i = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}$$

Here v^i does not mean "the covector v", but rather, "the components of the vector v".

Mnemonics

- "**Up**per indices go **up** to down; **l**ower indices go **l**eft to right"
- You can stack vectors (column matrices) side-by-side:

$$\begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix}.$$

Hence the lower index indicates which *column* you are in.

- You can stack covectors (row matrices) top-to-bottom:

$$\begin{bmatrix} w^1 \\ \vdots \\ w^k \end{bmatrix}$$

Hence the upper index indicates which *row* you are in.

Superscripts and subscripts vs. only subscripts

In the presence of a non-degenerate form (an isomorphism $V \to V^*$, for instance a Riemannian metric or Minkowski metric), one can raise and lower indices.

A basis gives such a form (via the dual basis), hence when working on $\mathbf{R}^n$ with a fixed basis, one can work with just subscripts.

However, if one changes coordinates, the way that coefficients change depends on the variance of the object, and one cannot ignore the distinction.

Common operations in this notation

In Einstein notation, the usual element reference $\mathbf{A}_{mn}$ for the *m*th row and *n*th column of matrix $\mathbf{A}$ becomes $\mathbf{A}_n^m$. We can then write the following operations in Einstein notation as follows.

Inner product

Given a row vector *v* and a column vector *u* of the same size, we can take the inner product $v_i u^i$, which is a scalar: it's evaluating the covector on the vector.

Multiplication of a vector by a matrix

Given a matrix A_j^i and a (column) vector v^j, the coefficients of the product $\mathbf{A}v$ are given by $A_j^i v^j$.

Similarly, $v^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$ is equivalent to $A_i^j v_j$.

Matrix multiplication

We can represent matrix multiplication as:

$$C^i_k = A^i_j B^j_k$$

This expression is equivalent to the more conventional (and less compact) notation:

$$\mathbf{C}_{ik} = (\mathbf{AB})_{ik} = \sum_{j=1}^{N} A_{ij} B_{jk}$$

Trace

Given a square matrix A^i_j, summing over a common index A^i_i yields the trace.

Outer product

The outer product of the column vector **u** by the row vector **v** yields an $M \times N$ matrix **A**:

$$\mathbf{A} = \mathbf{u}\,\mathbf{v}$$

In Einstein notation, we have:

$$A^i_j = u^i v_j = (uv)^i_j$$

Since *i* and *j* represent two *different* indices, and in this case over two different ranges *M* and *N* respectively, the indices are not eliminated by the multiplication. Both indices survive the multiplication to become the two indices of the newly-created matrix *A*.

Coefficients on tensors and related

Given a tensor field and a basis (of linearly independent vector fields), the coefficients of the tensor field in a basis can be computed by evaluating on a suitable combination of the basis and dual basis, and inherits the correct indexing. We list notable examples.

Throughout, let e_i be a basis of vector fields (a moving frame).

- (covariant) metric tensor

  $g_{ij} = g(e_i, e_j)$

- (contravariant) metric tensor

  $g^{ij} = g(e^i, e^j)$

- Torsion tensor (using the below)

$$T^c_{ab} = \Gamma^c_{ab} - \Gamma^c_{ba} - \gamma^c_{ab},$$

which follows from the formula

$$T = \nabla_X Y - \nabla_Y X - [X, Y].$$

- Riemann curvature tensor

$$R^\rho{}_{\sigma\mu\nu} = dx^\rho(R(\partial_\mu, \partial_\nu)\partial_\sigma)$$

This also applies for some operations that are not tensorial, for instance:

- Christoffel symbols

$$\nabla_i e_j = \Gamma^k_{ij} e_k$$

where $\nabla_i e_j$ is the covariant derivative. Equivalently,

$$\Gamma^k_{ij} = e^k \nabla_i e_j$$

- commutator coefficients

$$[e_i, e_j] = \gamma^k_{ij} e_k$$

where $[e_i,e_j]$ is the Lie bracket. Equivalently,

$$\gamma^k_{ij} = e^k [e_i, e_j].$$

Vector dot product

In mechanics and engineering, vectors in 3D space are often described in relation to orthogonal unit vectors **i**, **j** and **k**.

$$\mathbf{u} = u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k}$$

If the basis vectors **i**, **j**, and **k** are instead expressed as $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$, a vector can be expressed in terms of a summation:

$$\mathbf{u} = u^1 \mathbf{e}_1 + u^2 \mathbf{e}_2 + u^3 \mathbf{e}_3 = \sum_{i=1}^{3} u^i \mathbf{e}_i$$

In Einstein notation, the summation symbol is omitted since the index *i* is repeated once as an upper index and once as a lower index, and we simply write

$$\mathbf{u} = u^i \mathbf{e}_i$$

Using $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ instead of **i**, **j**, and **k**, together with Einstein notation, we obtain a concise algebraic presentation of vector and tensor equations. For example,

$$\mathbf{u} \cdot \mathbf{v} = \left(\sum_{i=1}^{3} u^i \mathbf{e}_i \right) \cdot \left(\sum_{j=1}^{3} v^j \mathbf{e}_j \right) = (u^i \mathbf{e}_i) \cdot (v^j \mathbf{e}_j) = u^i v^j (\mathbf{e}_i \cdot \mathbf{e}_j).$$

Since

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$$

where δ_{ij} is the Kronecker delta, which is equal to 1 when $i = j$, and 0 otherwise, we find

$$\mathbf{u} \cdot \mathbf{v} = u^i v^j \delta_{ij}.$$

One can use δ_{ij} to lower indices of the vectors; namely, $u_i = \delta_{ij} u^j$ and $v_i = \delta_{ij} v^j$. Then

$$\mathbf{u} \cdot \mathbf{v} = u^i v^j \delta_{ij} = u^i v_i = u_j v^j$$

Note that, despite $u^i = u_i$ for any fixed i, it is incorrect to write

$$\mathbf{u} \cdot \mathbf{v} = u^i v^i,$$

since on the right hand side the index i is repeated both times as an upper index and so there is no summation over i according to the Einstein convention. Rather, one should explicitly write the summation:

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{3} u^i v^i.$$

Vector cross product

For the cross product,

$$\mathbf{u} \times \mathbf{v} = \left(\sum_{j=1}^{3} u^j \mathbf{e}_j \right) \times \left(\sum_{k=1}^{3} v^k \mathbf{e}_k \right) = (u^j \mathbf{e}_j) \times (v^k \mathbf{e}_k)$$
$$= u^j v^k (\mathbf{e}_j \times \mathbf{e}_k) = u^j v^k \epsilon^i_{jk} \mathbf{e}_i$$

where $\mathbf{e}_j \times \mathbf{e}_k = \epsilon^i_{jk}\mathbf{e}_i$ and $\epsilon^i_{jk} = \delta^{il}\epsilon_{ljk}$, with ε_{ijk} the Levi-Civita symbol defined by:

$$\epsilon_{ijk} = \begin{cases} 0 & \text{unless } i,j,k \text{ are distinct} \\ +1 & \text{if } (i,j,k) \text{ is an even permutation of } (1,2,3) \\ -1 & \text{if } (i,j,k) \text{ is an odd permutation of } (1,2,3) \end{cases}$$

One then recovers

$$\mathbf{u} \times \mathbf{v} = (u^2v^3 - u^3v^2)\mathbf{e}_1 + (u^3v^1 - u^1v^3)\mathbf{e}_2 + (u^1v^2 - u^2v^1)\mathbf{e}_3$$

from

$$\mathbf{u} \times \mathbf{v} = \epsilon^i_{jk}u^jv^k\mathbf{e}_i = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\epsilon^i_{jk}u^jv^k\mathbf{e}_i$$

.

In other words, if $\mathbf{w} = \mathbf{u} \times \mathbf{v}$, then $w^i\mathbf{e}_i = \epsilon^i_{jk}u^jv^k\mathbf{e}_i$, so that $w^i = \epsilon^i_{jk}u^jv^k$.

Abstract definitions

In the traditional usage, one has in mind a vector space *V* with finite dimension *n*, and a specific basis of *V*. We can write the basis vectors as $\mathbf{e}_1$, $\mathbf{e}_2$, ..., $\mathbf{e}_n$. Then if **v** is a vector in *V*, it has coordinates $v^1, \ldots, v^n$ relative to this basis.

The basic rule is:

$$\mathbf{v} = v^i\mathbf{e}_i.$$

In this expression, it was assumed that the term on the right side was to be summed as *i* goes from 1 to *n*, because the index *i* does not appear on both sides of the expression. (Or, using Einstein's convention, because the index *i* appeared twice.)

An index that is summed over is a *summation index*. Here, the *i* is known as a *summation index*. It is also known as a *dummy index* since the result is not dependent on it; thus we could also write, for example:

$$\mathbf{v} = v^j\mathbf{e}_j.$$

An index that is not summed over is a *free index* and should be found in each term of the equation or formula. Compare dummy indices and free indices with free variables and bound variables.

The value of the Einstein convention is that it applies to other vector spaces built from V using the tensor product and duality. For example, $V \otimes V$, the tensor product of V with itself, has a basis consisting of tensors of the form $\mathbf{e}_{ij} = \mathbf{e}_i \otimes \mathbf{e}_j$. Any tensor $\mathbf{T}$ in $V \otimes V$ can be written as:

$$\mathbf{T} = T^{ij}\mathbf{e}_{ij}.$$

V^*, the dual of V, has a basis $\mathbf{e}^1$, $\mathbf{e}^2$, ..., $\mathbf{e}^n$ which obeys the rule

$$\mathbf{e}^i(\mathbf{e}_j) = \delta^i_j.$$

Here δ is the Kronecker delta, so δ^i_j is 1 if $i = j$ and 0 otherwise.

As

$$\mathrm{Hom}(V, W) = V^* \otimes W$$

the row-column coordinates on a matrix correspond to the upper-lower indices on the tensor product.

Examples

Einstein summation is clarified with the help of a few simple examples. Consider four-dimensional spacetime, where indices run from 0 to 3:

$$a^\mu b_\mu = a^0 b_0 + a^1 b_1 + a^2 b_2 + a^3 b_3$$
$$a^{\mu\nu} b_\mu = a^{0\nu} b_0 + a^{1\nu} b_1 + a^{2\nu} b_2 + a^{3\nu} b_3.$$

The above example is one of contraction, a common tensor operation. The tensor $a^{\mu\nu} b_\mu$ becomes a new tensor by summing over the first upper index and the lower index. Typically the resulting tensor is renamed with the contracted indices removed:

$$s^\nu = a^{\mu\nu} b_\mu.$$

For a familiar example, consider the dot product of two vectors **a** and **b**. The dot product is defined simply as summation over the indices of **a** and **b**:

$$\mathbf{a} \cdot \mathbf{b} = a^\alpha b_\alpha = a^0 b_0 + a^1 b_1 + a^2 b_2 + a^3 b_3,$$

which is our familiar formula for the vector dot product. Remember it is sometimes necessary to change the components of **a** in order to lower its index; however, this is not necessary in Euclidean space, or any space with a metric equal to its inverse metric (e.g., flat spacetime).

Abstract index notation

Abstract index notation is a mathematical notation for tensors and spinors that uses indices to indicate their types, rather than their components in a particular basis. The indices are mere placeholders, not related to any fixed basis and, in particular, are non-numerical. The notation was introduced by Roger Penrose as a way to use the formal aspects of the Einstein summation convention in order to compensate for the difficulty in describing contractions and covariant differentiation in modern abstract tensor notation, while preserving the explicit covariance of the expressions involved.

Let V be a vector space, and V^* its dual. Consider, for example, a rank-2 covariant tensor $h \in V^* \otimes V^*$. Then h can be identified with a bilinear form on V. In other words, it is a function of two arguments in V which can be represented as a pair of *slots*:

$$h = h(-, -).$$

Abstract index notation is merely a *labelling* of the slots by Latin letters, which have no significance apart from their designation as labels of the slots (i.e., they are non-numerical):

$$h = h_{ab}.$$

A contraction between two tensors is represented by the repetition of an index label, where one label is contravariant (an *upper index* corresponding to a tensor in V) and one label is covariant (a *lower index* corresponding to a tensor in V^*). Thus, for instance,

$$t_{ab}{}^{b}$$

is the trace of a tensor $t = t_{ab}{}^{c}$ over its last two slots. This manner of representing tensor contractions by repeated indices is formally similar to the Einstein summation convention. However, as the indices are non-numerical, it does not imply summation: rather it corresponds to the abstract basis-independent trace operation (or duality pairing) between tensor factors of type V and those of type V^*.

Abstract indices and tensor spaces

A general homogeneous tensor is an element of a tensor product of copies of V and V^*, such as

$$V \otimes V^* \otimes V^* \otimes V \otimes V^*.$$

Label each factor in this tensor product with a Latin letter in a raised position for each contravariant V factor, and in a lowered position for each covariant V^* position. In this way, write the product as

$$V^a V_b V_c V^d V_e$$

or, simply

$$V^a{}_{bc}{}^d{}_e.$$

It is important to remember that these last two expressions signify precisely the same object as the first. We shall denote tensors of this type by the same sort of notation, for instance

$$h^a{}_{bc}{}^d{}_e \in V^a{}_{bc}{}^d{}_e = V \otimes V^* \otimes V^* \otimes V \otimes V^*.$$

Contraction

In general, whenever one contravariant and one covariant factor occur in a tensor product of spaces, there is an associated *contraction* (or *trace*) map. For instance,

$$\mathrm{Tr}_{12} : V \otimes V^* \otimes V^* \otimes V \otimes V^* \to V^* \otimes V \otimes V^*$$

is the trace on the first two spaces of the tensor product.

$$\mathrm{Tr}_{15} : V \otimes V^* \otimes V^* \otimes V \otimes V^* \to V^* \otimes V^* \otimes V$$

is the trace on the first and last space.

These trace operations are signified on tensors by the repetition of an index. Thus the first trace map is given by

$$\mathrm{Tr}_{12} : h^a{}_{bc}{}^d{}_e \mapsto h^a{}_{ac}{}^d{}_e$$

and the second by

$$\mathrm{Tr}_{15} : h^a{}_{bc}{}^d{}_e \mapsto h^a{}_{bc}{}^d{}_a$$

Braiding

To any tensor product, there are associated braiding maps. For example, the braiding map

$$\tau_{(12)} : V \otimes V \to V \otimes V$$

interchanges the two tensor factors (so that its action on simple tensors is given by $\tau(v \otimes w) = w \otimes v$). In general, the braiding maps are in one-to-one correspondence with elements of the symmetric group, acting by permuting the tensor factors. Here, we

use τ_σ to denote the braiding map associated to the permutation σ (represented as a product of disjoint cyclic permutations).

Braiding maps are important in differential geometry, for instance, in order to express the Bianchi identity. Here let R denote the Riemann tensor, regarded as a tensor in $V^* \otimes V^* \otimes V^* \otimes V$. The first Bianchi identity then asserts that

$$R + \tau_{(123)}R + \tau_{(132)}R = 0.$$

Abstract index notation handles braiding as follows. On a particular tensor product, an ordering of the abstract indices is fixed (usually this is a lexicographic ordering). The braid is then represented in notation by permuting the labels of the indices. Thus, for instance, with the Riemann tensor

$$R = R_{abc}{}^{d} \in V_{abc}{}^{d} = V^* \otimes V^* \otimes V^* \otimes V,$$

the Bianchi identity becomes

$$R_{abc}{}^{d} + R_{cab}{}^{d} + R_{bca}{}^{d} = 0.$$

Chapter 15

Tensor Product and Tensor Contraction

Tensor product

In mathematics, the **tensor product**, denoted by $\otimes$, may be applied in different contexts to vectors, matrices, tensors, vector spaces, algebras, topological vector spaces, and modules, among many other structures or objects. In each case the significance of the symbol is the same: the most general bilinear operation. In some contexts, this product is also referred to as **outer product**. The term "tensor product" is also used in relation to monoidal categories.

Tensor product of vector spaces

The tensor product $V \otimes W$ of two vector spaces V and W over a field K can be defined by the method of *generators and relations*.

To construct $V \otimes W$, one begins with the set of ordered pairs in the Cartesian product $V \times W$. For the purposes of this construction, regard this Cartesian product as a set rather than a vector space. The free vector space F on $V \times W$ is defined by taking the vector space in which the elements of $V \times W$ are a basis. In set-builder notation,

$$F(V \times W) = \left\{ \sum_{i=1}^{n} \alpha_i e_{(v_i, w_i)} \;\middle|\; n \in \mathbb{N}, \alpha_i \in K, (v_i, w_i) \in V \times W \right\},$$

where we have used the symbol $e_{(v,w)}$ to emphasize that these are taken to be linearly independent *by definition* for distinct $(v,w) \in V \times W$.

The tensor product arises by defining the following four equivalence relations in $F(V \times W)$:

$$e_{(v_1+v_2,w)} \sim e_{(v_1,w)} + e_{(v_2,w)}$$
$$e_{(v,w_1+w_2)} \sim e_{(v,w_1)} + e_{(v,w_2)}$$
$$ce_{(v,w)} \sim e_{(cv,w)} \sim e_{(v,cw)}$$

where v_1, v_2, and w_1, w_2 are vectors from V and W (respectively), and c is from the underlying field K. Denoting by R the space generated by these four equivalence relations, the **tensor product of the two vector spaces** V and W is then the quotient space

$$V \otimes W = F(V \times W)/R.$$

It is also called the **tensor product space** of V and W and is a vector space (which can be verified by directly checking the vector space axioms). The **tensor product of two elements** v and w is the equivalence class $(e_{(v,w)} + R)$ of $e_{(v,w)}$ in $V \otimes W$, denoted $v \otimes w$. This notation can somewhat obscure the fact that tensors are always cosets: manipulations performed via the representatives (v,w) must always be checked that they do not depend on the particular choice of representative.

The space R is mapped to zero in $V \otimes W$, so that the above three equivalence relations become equalities in the tensor product space:

$$(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w;$$
$$v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2;$$
$$cv \otimes w = v \otimes cw = c(v \otimes w).$$

Given bases $\{v_i\}$ and $\{w_i\}$ for V and W respectively, the tensors $\{v_i \otimes w_j\}$ form a basis for $V \otimes W$ (generally ordered so that $v_i \otimes w_{j+1}$ comes before $v_{i+1} \otimes w_j$). The dimension of the tensor product therefore is the product of dimensions of the original spaces; for instance $\mathbf{R}^m \otimes \mathbf{R}^n$ will have dimension mn.

Elements of $V \otimes W$ are sometimes referred to as **tensors**, although this term refers to many other related concepts as well. An element of $V \otimes W$ of the form $v \otimes w$ is called a **pure** or **simple tensor**. In general, an element of the tensor product space is not a pure tensor, but rather a finite linear combination of pure tensors. That is to say, if v_1 and v_2 are linearly independent, and w_1 and w_2 are also linearly independent, then $v_1 \otimes w_1 + v_2 \otimes w_2$ cannot be written as a pure tensor. The number of simple tensors required to express an element of a tensor product is called the **tensor rank,** and for linear operators or matrices, thought of as (1,1) tensors (elements of the space $V \otimes V^*$), it agrees with matrix rank.

Characterization by a universal property

The tensor product is universal, in that any other bilinear map from $V \times W$ to some vector space Z induces a *unique* linear map from $V \otimes W$ to Z. That is, the tensor product captures the entirety of the concept of a bilinear map: all bilinear maps are isomorphic to the natural embedding map

$$\begin{aligned} \varphi : V \times W &\to V \otimes W \\ (u, w) &\mapsto u \otimes w \end{aligned}$$

Specifically, any bilinear map $\psi \in L^2(V \times W, Z)$ induces a *unique* (up to isomorphism) linear map $T \in L(V \otimes W, Z)$, given by $\psi = T \circ \varphi$.

In the category of bilinear maps ψ from $V \times W$, the morphisms are the linear maps T. Given a bilinear map ψ, and a linear morphism T, this gives the bilinear map T o ψ.

The natural bilinear embedding map φ is an initial object of that category. The tensor product $V \otimes W$ can be defined as the image of any initial object ψ, which is isomorphic to φ.

As a functor

The tensor product also operates on linear maps between vector spaces. Specifically, given two linear maps $S : V \to X$ and $T : W \to Y$ between vector spaces, the **tensor product of the two linear maps** S and T is a linear map

$$S \otimes T : V \otimes W \to X \otimes Y$$

defined by

$$(S \otimes T)(v \otimes w) = S(v) \otimes T(w).$$

In this way, the tensor product becomes a bifunctor from the category of vector spaces to itself, covariant in both arguments.

The Kronecker product of two matrices is the matrix of the tensor product of the two corresponding linear maps under a standard choice of bases of the tensor products.

More than two vector spaces

The construction and the universal property of the tensor product can be extended to allow for more than two vector spaces. For example, suppose that V_1, V_2, and V_3 are three vector spaces. The tensor product $V_1 \otimes V_2 \otimes V_3$ is defined along with a trilinear mapping from the direct product

$$\varphi : V_1 \times V_2 \times V_3 \to V_1 \otimes V_2 \otimes V_3$$

so that, any trilinear map F from the direct product to a vector space W

$$F : V_1 \times V_2 \times V_3 \to W$$

factors uniquely as

$$F = L \circ \varphi$$

where L is a linear map. The tensor product is uniquely characterized by this property, up to a unique isomorphism.

This construction is related to repeated tensor products of two spaces. For example, if V_1, V_2, and V_3 are three vector spaces, then there are (natural) isomorphisms

$$V_1 \otimes V_2 \otimes V_3 \cong V_1 \otimes (V_2 \otimes V_3) \cong (V_1 \otimes V_2) \otimes V_3.$$

More generally, the tensor product of an arbitrary indexed family V_i, $i \in I$, is defined to be universal with respect to multilinear mappings of the direct product $\prod_{i \in I} V_i$.

Tensor powers and braiding

Let n be a non-negative integer. The nth **tensor power** of the vector space V is the n-fold tensor product of V with itself. That is

$$V^{\otimes n} \stackrel{\text{def}}{=} \underbrace{V \otimes \cdots \otimes V}_{n}.$$

A permutation σ of the set {1,2,…,n} determines a mapping of the nth Cartesian power of V

$$\sigma : V^n \to V^n$$

defined by

$$\sigma(v_1, v_2, \ldots, v_n) = (v_{\sigma 1}, v_{\sigma 2}, \ldots, v_{\sigma n}).$$

Let

$$\varphi : V^n \to V^{\otimes n}$$

be the natural multilinear embedding of the Cartesian power of V into the tensor power of V. Then, by the universal property, there is a unique isomorphism

$$\tau_\sigma : V^{\otimes n} \to V^{\otimes n}$$

such that

$$\varphi \circ \sigma = \tau_\sigma \circ \varphi.$$

The isomorphism τ_σ is called the **braiding map** associated to the permutation σ.

Tensor product of two tensors

A tensor on V is an element of a vector space of the form

$$T^r_s(V) = \underbrace{V \otimes \cdots \otimes V}_{r} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{s} = V^{\otimes r} \otimes V^{*\otimes s}$$

for non-negative integers r and s. There is a general formula for the components of a (tensor) product of two (or more) tensors. For example, if F and G are two covariant tensors of rank m and n (respectively) (i.e. $F \in T_m^0$, and $G \in T_n^0$), then the components of their tensor product are given by

$$(F \otimes G)_{i_1 i_2 \ldots i_{m+n}} = F_{i_1 i_2 \ldots i_m} G_{i_{m+1} i_{m+2} i_{m+3} \ldots i_{m+n}}.$$

In this example, it is assumed that there is a chosen basis B of the vector space V, and the basis on any tensor space T_s^r is tacitly assumed to be the standard one (this basis is described in the Kronecker products). Thus, the components of the tensor product of two tensors are the ordinary product of the components of each tensor.

Note that in the tensor product, the factor F consumes the first rank(F) indices, and the factor G consumes the next rank(G) indices, so

$$\operatorname{rank}(F \otimes G) = \operatorname{rank}(F) + \operatorname{rank}(G).$$

The tensor $T^r_s(V)$ may be naturally viewed as a module for the Lie algebra $\operatorname{End}(V)$ by means of the diagonal action: for simplicity let us assume $r=s=1$, then, for each $u \in \operatorname{End}(V)$,

$$u(a \otimes b) = u(a) \otimes b - a \otimes u^*(b),$$

where $u^* \in \operatorname{End}(V^*)$ is the transpose of u, that is, in terms of the obvious pairing on $V \otimes V^*$, $\langle u(a), b\rangle = \langle a, u^*(b)\rangle$.

There is a canonical isomorphism $T^1_1(V) \to \operatorname{End}(V)$ given by

$$(a \otimes b)(x) = \langle x, b\rangle a.$$

Under this isomorphism, every $u \in \mathrm{End}(V)$ may be first viewed as an endomorphism of $T^1_1(V)$ and then viewed as an endomorphism of $\mathrm{End}(V)$. In fact it is the adjoint representation $\mathrm{ad}(u)$ of End(V).

Example

Let **U** be a tensor of type (1,1) with components $U^{\alpha}{}_{\beta}$, and let **V** be a tensor of type (1,0) with components V^{γ}. Then

$$U^{\alpha}{}_{\beta}V^{\gamma} = (U \otimes V)^{\alpha}{}_{\beta}{}^{\gamma}$$

and

$$V^{\mu}U^{\nu}{}_{\sigma} = (V \otimes U)^{\mu\nu}{}_{\sigma}.$$

The tensor product inherits all the indices of its factors.

Kronecker product of two matrices

With matrices this operation is usually called the *Kronecker product*, a term used to make clear that the result has a particular block structure imposed upon it, in which each element of the first matrix is replaced by the second matrix, scaled by that element. For matrices U and V this is:

$$U \otimes V = \begin{bmatrix} u_{1,1}V & u_{1,2}V & \cdots \\ u_{2,1}V & u_{2,2}V & \\ \vdots & & \ddots \end{bmatrix} = \begin{bmatrix} u_{1,1}v_{1,1} & u_{1,1}v_{1,2} & \cdots & u_{1,2}v_{1,1} & u_{1,2}v_{1,2} & \cdots \\ u_{1,1}v_{2,1} & u_{1,1}v_{2,2} & & u_{1,2}v_{2,1} & u_{1,2}v_{2,2} & \\ \vdots & & \ddots & & & \\ u_{2,1}v_{1,1} & u_{2,1}v_{1,2} & & & & \\ u_{2,1}v_{2,1} & u_{2,1}v_{2,2} & & & & \\ \vdots & & & & & \end{bmatrix}.$$

For example, the tensor product of two two-dimensional square matrices:

$$\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \otimes \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} = \begin{bmatrix} a_{1,1}\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} & a_{1,2}\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \\ a_{2,1}\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} & a_{2,2}\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} a_{1,1}b_{1,1} & a_{1,1}b_{1,2} & a_{1,2}b_{1,1} & a_{1,2}b_{1,2} \\ a_{1,1}b_{2,1} & a_{1,1}b_{2,2} & a_{1,2}b_{2,1} & a_{1,2}b_{2,2} \\ a_{2,1}b_{1,1} & a_{2,1}b_{1,2} & a_{2,2}b_{1,1} & a_{2,2}b_{1,2} \\ a_{2,1}b_{2,1} & a_{2,1}b_{2,2} & a_{2,2}b_{2,1} & a_{2,2}b_{2,2} \end{bmatrix}.$$

The resultant rank is at most 4, and the resultant dimension 16. Here rank denotes the tensor rank (number of requisite indices), while the matrix rank counts the number of degrees of freedom in the resulting array.

A representative case is the Kronecker product of any two rectangular arrays, considered as matrices. A dyadic product is the special case of the tensor product between two vectors of the same dimension.

Tensor product of multilinear maps

Given multilinear maps $f(x_1,...,x_k)$ and $g(x_1,...,x_m)$ their tensor product is the multilinear function

$$(f \otimes g)(x_1, \ldots, x_{k+m}) = f(x_1, \ldots, x_k) g(x_{k+1}, \ldots, x_{k+m}).$$

Relation with the dual space

In the discussion on the universal property, replacing X by the underlying scalar field of V and W yields that the space $(V \otimes W)^*$ (the dual space of $V \otimes W$, containing all linear functionals on that space) is naturally identified with the space of all bilinear functionals on $V \times W$. In other words, every bilinear functional is a functional on the tensor product, and vice versa.

Whenever V and W are finite dimensional, there is a natural isomorphism between $V^* \otimes W^*$ and $(V \otimes W)^*$, whereas for vector spaces of arbitrary dimension we only have an inclusion $V^* \otimes W^* \subset (V \otimes W)^*$. So, the tensors of the linear functionals are bilinear functionals. This gives us a new way to look at the space of bilinear functionals, as a tensor product itself.

Types of tensors

Linear subspaces of the bilinear operators (or in general, multilinear operators) determine natural quotient spaces of the tensor space, which are frequently useful. See wedge product for the first major example. Another would be the treatment of algebraic forms as symmetric tensors.

Over more general rings

The notation $\otimes_R$ refers to a tensor product of modules over a ring R.

Tensor product for computer programmers

Array programming languages

Array programming languages may have this pattern built in. For example, in APL the tensor product is expressed as $\circ.\times$ (for example $A \circ.\times B$ or $A \circ.\times B \circ.\times C$). In J the tensor product is the dyadic form of ***/** (for example **a */ b** or **a */ b */ c**).

Note that J's treatment also allows the representation of some tensor fields (as **a** and **b** may be functions instead of constants—the result is then a derived function, and if **a** and **b** are differentiable, then **a*/b** is differentiable).

However, these kinds of notation are not universally present in array languages. Other array languages may require explicit treatment of indices (for example, MATLAB),

and/or may not support higher-order functions such as the Jacobian derivative (for example, Fortran/APL).

Tensor contraction

In multilinear algebra, a **tensor contraction** is an operation on one or more tensors that arises from the natural pairing of a finite-dimensional vector space and its dual. In components, it is expressed as a sum of products of scalar components of the tensor(s) caused by applying the summation convention to a pair of dummy indices which are bound to each other in an expression. The contraction of a single mixed tensor occurs when a pair of literal indices (one a subscript, the other a superscript) of the tensor are set equal to each other and summed over. In the Einstein notation this summation is built into the notation. The result is another tensor with rank (or **order**) reduced by 2.

Tensor contraction can be seen as a generalization of the trace and of matrix multiplication.

Abstract formulation

Let V be a vector space over a field k. The core of the contraction operation, and the simplest case, is the natural pairing of V with its dual vector space V^*. The pairing is the linear transformation from the tensor product of these two spaces to the field k:

$$C : V^* \otimes V \to k$$

corresponding to the bilinear form

$$\langle f, v \rangle = f(v)$$

where f is in V^* and v is in V. The map C defines the contraction operation on a tensor of type (1,1), which is an element of $V^* \otimes V$. Note that the result is a scalar (an element of k). Using the natural isomorphism between $V^* \otimes V$ and the space of linear transformations from V to V, one obtains a basis-free definition of the trace.

In general, a tensor of type (m, n) (with $m \geq 1$ and $n \geq 1$) is an element of the vector space

$$V \otimes \ldots \otimes V \otimes V^* \otimes \ldots \otimes V^*$$

(where there are m V factors and n V^* factors). Applying the natural pairing to the kth V factor and the lth V^* factor, and using the identity on all other factors, defines the (k, l) contraction operation, which is a linear map which yields a tensor of type (m-1, n-1). By analogy with the (1,1) case, the general contraction operation is sometimes called the trace.

Contraction in index notation

In abstract index notation, the basic contraction of a vector and a dual vector is denoted by

$$\tilde{f}(\vec{v}) = f_\gamma v^\gamma$$

which is shorthand for the explicit coordinate summation

$$f_\gamma v^\gamma = f_1 v^1 + f_2 v^2 + \ldots + f_n v^n$$

(where v^i are the components of v in a particular basis and f_i are the components of f in the corresponding dual basis).

Since a general mixed dyadic tensor is a linear combination of decomposable tensors of the form $f \otimes v$, the explicit formula for the dyadic case follows: let

$$\mathbf{T} = T^i{}_j \mathbf{e_i} \mathbf{e^j}$$

be a mixed dyadic tensor. Then its contraction is

$$T^i{}_j \mathbf{e_i} \cdot \mathbf{e^j} = T^i{}_j \delta_i{}^j = T^j{}_j = T^1{}_1 + \cdots + T^n{}_n.$$

A general contraction is denoted by labeling one covariant index and one contravariant index with the same letter, summation over that index being implied by the summation convention. The resulting contracted tensor inherits the remaining indices of the original tensor. For example, contracting a tensor T of type (2,2) on the second and third indices to create a new tensor U of type (1,1) is written as

$$T^{ab}{}_{bc} = \sum_b T^{ab}{}_{bc} = T^{a1}{}_{1c} + T^{a2}{}_{2c} + \ldots + T^{an}{}_{nc} = U^a{}_c.$$

By contrast, let

$$\mathbf{T} = \mathbf{e^i} \mathbf{e^j}$$

be an unmixed dyadic tensor. This tensor does not contract; if its base vectors are dotted the result is the contravariant metric tensor,

$$g^{ij} = \mathbf{e^i} \cdot \mathbf{e^j},$$

whose rank is 2.

Metric contraction

As in the previous example, contraction on a pair of indices that are either both contravariant or both covariant is not possible in general. However, in the presence of an inner product (also known as a metric) g, such contractions are possible. One uses the

metric to raise or lower one of the indices, as needed, and then one uses the usual operation of contraction. The combined operation is known as *metric contraction*.

Application to tensor fields

Contraction is often applied to tensor fields over spaces (e.g. Euclidean space, manifolds, or schemes). Since contraction is a purely algebraic operation, it can be applied pointwise to a tensor field, e.g. if T is a (1,1) tensor field on Euclidean space, then in any coordinates, its contraction (a scalar field) U at a point x is given by

$$U(x) = \sum_i T_i^i(x)$$

Since the role of x is not complicated here, it is often suppressed, and the notation for tensor fields becomes identical to that for purely algebraic tensors.

Over a Riemannian manifold, a metric (field of inner products) is available, and both metric and non-metric contractions are crucial to the theory. For example, the Ricci tensor is a non-metric contraction of the Riemann curvature tensor, and the scalar curvature is the unique metric contraction of the Ricci tensor.

One can also view contraction of a tensor field in the context of modules over an appropriate ring of functions on the manifold or the context of sheaves of modules over the structure sheaf.

Tensor divergence

As an application of the contraction of a tensor field, let V be a vector field on a Riemannian manifold (for example, Euclidean space). Let ${V^\alpha}_{;\beta}$ be the covariant derivative of V (in some choice of coordinates). In the case of Cartesian coordinates in Euclidean space, one can write

$${V^\alpha}_{;\beta} = \frac{\partial V^\alpha}{\partial x^\beta}.$$

Then changing index β to α causes the pair of indices to become bound to each other, so that the derivative contracts with itself to obtain the following sum:

$${V^\alpha}_{;\alpha} = {V^0}_{;0} + \ldots + {V^n}_{;n}$$

which is the divergence div V. Then

$$\mathrm{div} V = {V^\alpha}_{;\alpha} = 0$$

is a continuity equation for V.

In general, one can define various divergence operations on higher-rank tensor fields, as follows. If T is a tensor field with at least one contravariant index, taking the covariant differential and contracting the chosen contravariant index with the new covariant index corresponding to the differential results in a new tensor of rank one lower than that of T.

Contraction of a pair of tensors

One can generalize the core contraction operation (vector with dual vector) in a slightly different way, by considering a pair of tensors T and U. The tensor product $T \otimes U$ is a new tensor, which, if it has at least one covariant and one contravariant index, can be contracted. The case where T is a vector and U is a dual vector is exactly the core operation introduced first here.

In abstract index notation, to contract two tensors with each other, one places them side by side (juxtaposed) as factors of the same term. This implements the tensor product, yielding a composite tensor. Contracting two indices in this composite tensor implements the desired contraction of the two tensors.

For example, matrices can be represented as tensors of type (1,1) with the first index being contravariant and the second index being covariant. Let $\Lambda^{\alpha}{}_{\beta}$ be the components of one matrix and let $M^{\beta}{}_{\gamma}$ be the components of a second matrix. Then their multiplication is given by the following contraction, an example of the contraction of a pair of tensors:

$$\Lambda^{\alpha}{}_{\beta} M^{\beta}{}_{\gamma} = N^{\alpha}{}_{\gamma}.$$

Also, the interior product of a vector with a differential form is a special case of the contraction of two tensors with each other.

More general algebraic contexts

Let R be a commutative ring and let M be a finite free module over R. Then contraction operates on the full (mixed) tensor algebra of M in exactly the same way as it does in the case of vector spaces over a field. (The key fact is that the natural pairing is still perfect in this case.)

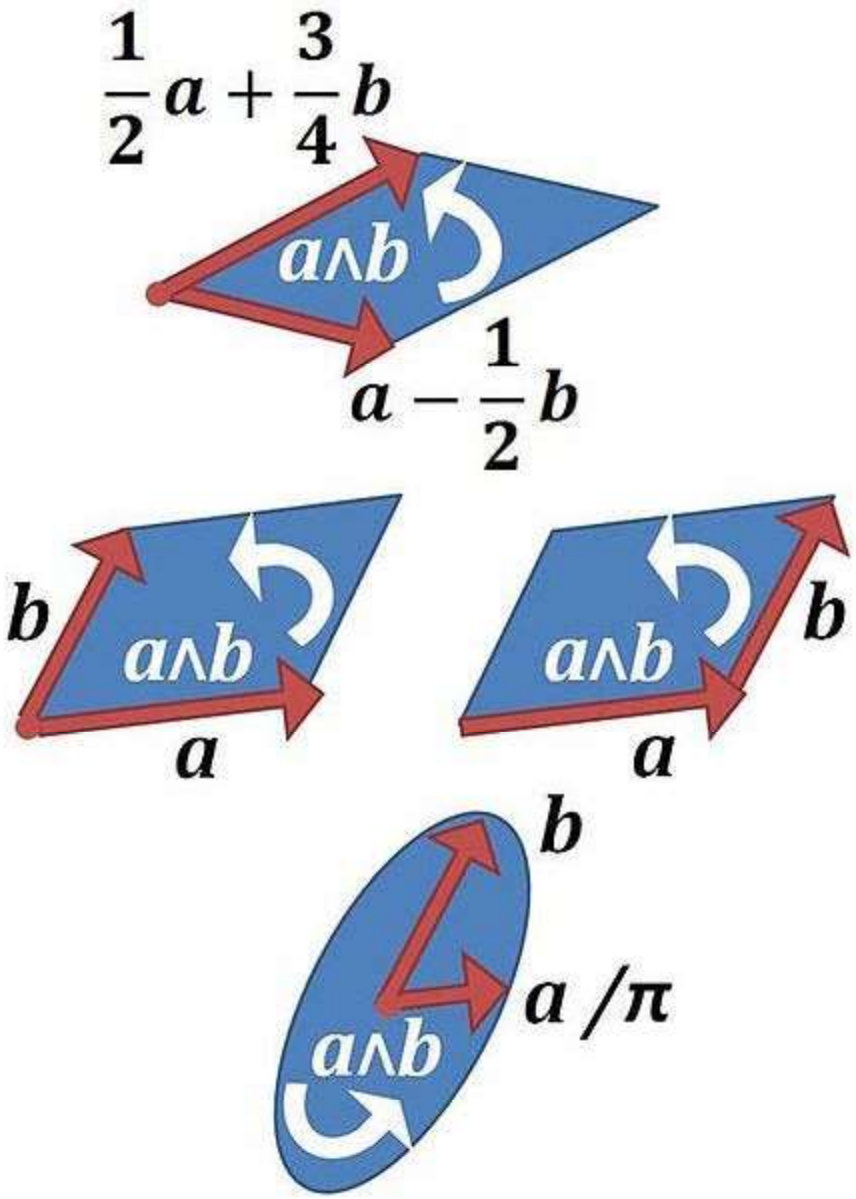
More generally, let O_X be a sheaf of commutative rings over a topological space X, e.g. O_X could be the structure sheaf of a complex manifold, analytic space, or scheme. Let M be a locally free sheaf of modules over O_X of finite rank. Then the dual of M is still well-behaved and contraction operations make sense in this context.

Chapter 16

Bivector

In mathematics, a **bivector** or **2-vector** is a quantity in geometric algebra or exterior algebra that generalises the idea of a vector. If a scalar is considered a zero dimensional quantity, and a vector is a one dimensional quantity, then a bivector can be thought of as two dimensional. Bivectors have applications in many areas of mathematics and physics. They are related to complex numbers in two dimensions and to both pseudovectors and quaternions in three dimensions. They can be used to generate rotations in any dimension, and are a useful tool for classifying such rotations. They also are used in physics, tying together a number of otherwise unrelated quantities.

Bivectors are generated by the exterior product on vectors – given two vectors **a** and **b** their exterior product **a** ∧ **b** is a bivector. But not all bivectors can be generated this way, and in higher dimensions a sum of exterior products is often needed. More precisely a bivector that requires only a single exterior product is *simple*; in two and three dimensions all bivectors are simple, but in higher dimensions this is not generally the case. The exterior product is antisymmetric, so **b** ∧ **a** negates the bivector, producing a rotation with the opposite sense, and **a** ∧ **a** is the zero bivector.



Parallel plane segments with the same orientation and area corresponding to the same bivector **a** ∧ **b**.

Geometrically, a simple bivector can be interpreted as an oriented plane segment, much as vectors can be thought of as directed line segments. Specifically for the bivector **a** ∧ **b**, its *magnitude* is the area of the parallelogram with edges **a** and **b**, its *attitude* that of any plane specified by **a** and **b**, and its *orientation* the sense of the rotation that would align **a** with **b**. It does not have a definite location or position.

History

The bivector was first defined in 1844 by German mathematician Hermann Grassmann in exterior algebra, as the result of the exterior product. Around the same time in 1843 in Ireland William Rowan Hamilton discovered quaternions. It was not until English

mathematician William Kingdon Clifford in 1888 added the geometric product to Grassmann's algebra, incorporating the ideas of both Hamilton and Grassmann, and founded Clifford algebra, that the bivector as it is known today was fully understood.

Around this time Josiah Willard Gibbs and Oliver Heaviside developed vector calculus which included separate cross product and dot products, derived from quaternion multiplication. The success of vector calculus, and of the book Vector Analysis by Gibbs and Wilson, meant the insights of Hamilton and Clifford were overlooked for a long time, as much of 20th century mathematics and physics was formulated in vector terms. Gibbs instead described bivectors as vectors, and used "bivector" to describe an unrelated quantity, a use that has sometimes been copied.

Today the bivector is largely studied as a topic in geometric algebra, a more restricted Clifford algebra over real or complex vector spaces with nondegenerate quadratic form. Its resurgence was led by David Hestenes who, along with others, discovered a range of new applications in physics for geometric algebra.

Formal definition

For this chapter the bivector will be considered only in real geometric algebras. This in practice is not much of a restriction, as all useful applications are drawn from such algebras. Also unless otherwise stated all examples have a Euclidian metric and so a quadratic form with signature 1.

Geometric algebra and the geometric product

The bivector arises from the definition of the geometric product over a vector space. For vectors **a**, **b** and **c** the geometric product on vectors is defined as follows:

Associativity:

$$(\mathbf{ab})\mathbf{c} = \mathbf{a}(\mathbf{bc})$$

Left and right distributivity:

$$\mathbf{a}(\mathbf{b} + \mathbf{c}) = \mathbf{ab} + \mathbf{ac}$$
$$(\mathbf{b} + \mathbf{c})\mathbf{a} = \mathbf{ba} + \mathbf{ca}$$

Contraction:

$$\mathbf{a}^2 = Q(\mathbf{a}) = \epsilon_{\mathbf{a}}|\mathbf{a}|^2$$

Where Q is the quadratic form, $|\mathbf{a}|$ is the magnitude of **a** and $\epsilon_{\mathbf{a}}$ is the signature of the vector. For a space with Euclidian metric $\epsilon_{\mathbf{a}}$ is 1 so can be omitted, and the contraction condition becomes:

$$\mathbf{a}^2 = |\mathbf{a}|^2$$

The interior product

From associativity **a**(**ab**) = **a**^{2}**b**, a scalar times **b**. So **ab** cannot be a scalar. But

$$\frac{1}{2}(\mathbf{ab} + \mathbf{ba}) = \frac{1}{2}((\mathbf{a}+\mathbf{b})^2 - \mathbf{a}^2 - \mathbf{b}^2)$$

is a sum of scalars and so a scalar. From the law of cosines on the triangle formed by the vectors its value is $|\mathbf{a}||\mathbf{b}|\cos\theta$, where θ is the angle between the vectors. It is therefore identical to the interior product between two vectors, and is written the same way,

$$\mathbf{a} \cdot \mathbf{b} = \frac{1}{2}(\mathbf{ab} + \mathbf{ba}).$$

It is symmetric, scalar valued, and can be used to determine the angle between two vectors: in particular if **a** and **b** are orthogonal it is zero.

The exterior product

In the same way another quantity can be written down:

$$\frac{1}{2}(\mathbf{ab} - \mathbf{ba})$$

This is called the exterior product, **a** ∧ **b**. It is antisymmetric in **a** and **b**, that is

$$\mathbf{a} \wedge \mathbf{b} = \frac{1}{2}(\mathbf{ab} - \mathbf{ba}) = -\mathbf{b} \wedge \mathbf{a}$$

By addition:

$$\mathbf{a} \cdot \mathbf{b} + \mathbf{a} \wedge \mathbf{b} = \frac{1}{2}(\mathbf{ab} + \mathbf{ba}) + \frac{1}{2}(\mathbf{ab} - \mathbf{ba}) = \mathbf{ab}$$

That is the geometric product is the sum of the symmetric interior product and antisymmetric exterior product.

To calculate **a** ∧ **b** consider the sum

$$(\mathbf{a} \cdot \mathbf{b})^2 - (\mathbf{a} \wedge \mathbf{b})^2$$

Expanding using the geometric product and simplifying gives

$$(\mathbf{a}\cdot\mathbf{b})^2 - (\mathbf{a}\wedge\mathbf{b})^2 = \mathbf{a}^2\mathbf{b}^2$$

so using the Pythagorean trigonometric identity:

$$(\mathbf{a}\wedge\mathbf{b})^2 = (\mathbf{a}\cdot\mathbf{b})^2 - \mathbf{a}^2\mathbf{b}^2 = |\mathbf{a}|^2\,|\mathbf{b}|^2\,(\cos^2\theta - 1) = -\,|\mathbf{a}|^2\,|\mathbf{b}|^2\sin^2\theta$$

With a negative square it cannot be a scalar or vector quantity, so it is a new sort of object, a **bivector**. It has magnitude $|\mathbf{a}||\mathbf{b}|\sin\theta$, where θ is the angle between the vectors, and so is zero for parallel vectors.

To distinguish them from vectors bivectors are written here with bold capitals, for example:

$$\mathbf{A} = \mathbf{a}\wedge\mathbf{b} = -\mathbf{b}\wedge\mathbf{a}\ .$$

Although other conventions are used, in particular as vectors and bivectors are both elements of the geometric algebra.

Properties

The space $\Lambda^2\mathbb{R}^n$

The algebra generated by the geometric product is the geometric algebra over the vector space. For a Euclidean vector space it is written $\mathcal{G}_n$ or $C\ell_n(\mathbb{R})$, where n is the dimension of the vector space $\mathbb{R}^n$. $C\ell_n$ is both a vector space and an algebra, generated by all the products between vectors in $\mathbb{R}^n$, so it contains all vectors and bivectors. More precisely as a vector space it contains the vectors and bivectors as subspaces. The space of all bivectors is written $\Lambda^2\mathbb{R}^n$. Unlike $\mathbb{R}^n$ it is not a Euclidean subspace; nor is it a subalgebra.

The even subalgebra

The subalgebra generated by the bivectors is the *even subalgebra* of the geometric agebra, written $C\ell^+_n$. This algebra results from considering all products of scalars and bivectors generated by the geometric product. It has dimension 2^{n-1}, and contains $\Lambda^2\mathbb{R}^n$ as a linear subspace with dimension $\frac{1}{2}n(n-1)$ (a triangular number). In two and three dimensions the even subalgebra contains only scalars and bivectors, and each is of particular interest. In two dimensions the even subalgebra is isomorphic to the complex numbers, $\mathbb{C}$, while in three it is isomorphic to the quaternions, $\mathbb{H}$. More generally the even subalgebra can be used to generate rotations in any dimension, and can be generated by bivectors in the algebra.

Magnitude

As noted in the previous section the magnitude of a simple bivector, that is one that is the exterior product of two vectors **a** and **b**, is $|\mathbf{a}||\mathbf{b}|\sin\theta$, where θ is the angle between the vectors. It is written $|\mathbf{B}|$, where **B** is the bivector.

For general bivectors the magnitude can be calculated by taking the norm of the bivector considered as a vector in the space $\Lambda^2\mathbb{R}^n$. If the magnitude is zero then all the bivector's components are zero, and the bivector is the zero bivector which as an element of the geometric algebra equals the scalar zero.

Unit bivectors

A unit bivector is one with unit magnitude. It can be derived from any non-zero bivector by dividing the bivector by its magnitude, that is

$$\frac{\mathbf{B}}{|\mathbf{B}|}.$$

Of particular interest are the unit bivectors formed from the products of the Standard basis. If $\mathbf{e}_i$ and $\mathbf{e}_j$ are distinct basis vectors then the product $\mathbf{e}_i \wedge \mathbf{e}_j$ is a bivector. As the vectors are orthogonal this is just $\mathbf{e}_i\mathbf{e}_j$, written $\mathbf{e}_{ij}$, with unit magnitude as the vectors are unit vectors. The set of all such bivectors form a basis for $\Lambda_2\mathbb{R}_n$. For instance in four dimensions the basis for $\Lambda_2\mathbb{R}_4$ is ($\mathbf{e}_1\mathbf{e}_2$, $\mathbf{e}_1\mathbf{e}_3$, $\mathbf{e}_1\mathbf{e}_4$, $\mathbf{e}_2\mathbf{e}_3$, $\mathbf{e}_2\mathbf{e}_4$, $\mathbf{e}_3\mathbf{e}_4$) or ($\mathbf{e}_{12}$, $\mathbf{e}_{13}$, $\mathbf{e}_{14}$, $\mathbf{e}_{23}$, $\mathbf{e}_{24}$, $\mathbf{e}_{34}$).

Simple bivectors

The exterior product of two vectors is a bivector, but not all bivectors are exterior products of two vectors. For example in four dimensions the bivector

$$\mathbf{B} = \mathbf{e}_1 \wedge \mathbf{e}_2 + \mathbf{e}_3 \wedge \mathbf{e}_4 = \mathbf{e}_1\mathbf{e}_2 + \mathbf{e}_3\mathbf{e}_4 = \mathbf{e}_{12} + \mathbf{e}_{34}$$

cannot be written as the exterior product of two vectors. A vector that can be written as the exterior product of two vectors is simple. In two and three dimensions all bivectors are simple, but not in four or more dimensions; in four dimensions every bivectors is the sum of at most two exterior products. A bivector has a real square if and only if it is simple, and only simple bivectors can be represented geometrically by a oriented plane area.

Product of two bivectors

The geometric product of two bivectors, **A** and **B**, is

$$\mathbf{AB} = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \times \mathbf{B} + \mathbf{A} \wedge \mathbf{B}.$$

The quantity $\mathbf{A} \cdot \mathbf{B}$ is the scalar valued interior product, while $\mathbf{A} \wedge \mathbf{B}$ is the grade 4 exterior product that arises in four or more dimensions. The quantity $\mathbf{A} \times \mathbf{B}$ is the bivector valued commutator product, given by

$$\mathbf{A} \times \mathbf{B} = \frac{1}{2}(\mathbf{AB} - \mathbf{BA}),$$

The space of bivectors $\Lambda^2\mathbb{R}^n$ are a Lie algebra over $\mathbb{R}$, with the commutator product as the Lie bracket. The full geometric product of bivectors generates the even subalgebra.

Of particular interest is the product of a bivector with itself. As the commutator product is antisymmetric the product simplifies to

$$\mathbf{AA} = \mathbf{A} \cdot \mathbf{A} + \mathbf{A} \wedge \mathbf{A}.$$

If the bivector is *simple* the last term is zero and the product is the scalar valued $\mathbf{A} \cdot \mathbf{A}$, which can be used as a check for simplicity. In particular the exterior product of bivectors only exists in four or more dimensions, so all bivectors in two and three dimensions are simple.

Two dimensions

When working with coordinates in geometric algebra it is usual to write the basis vectors as ($\mathbf{e}_1$, $\mathbf{e}_2$, ...), a convention that will be used here.

A vector in real two dimensional space $\mathbb{R}^2$ can be written $\mathbf{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2$, where a_1 and a_2 are real numbers, $\mathbf{e}_1$ and $\mathbf{e}_2$ are orthonormal basis vectors. The geometric product of two such vectors is

$$\begin{aligned}\mathbf{ab} &= (a_1\mathbf{e}_1 + a_2\mathbf{e}_2)(b_1\mathbf{e}_1 + b_2\mathbf{e}_2) \\ &= a_1b_1\mathbf{e}_1\mathbf{e}_1 + a_1b_2\mathbf{e}_1\mathbf{e}_2 + a_2b_1\mathbf{e}_2\mathbf{e}_1 + a_2b_2\mathbf{e}_2\mathbf{e}_2 \\ &= a_1b_1 + a_2b_2 + (a_1b_2 - a_2b_1)\mathbf{e}_1\mathbf{e}_2.\end{aligned}$$

This can be split into the symmetric, scalar valued, interior product and an antisymmetric, bivector valued exterior product:

$$\begin{aligned}\mathbf{a} \cdot \mathbf{b} &= a_1b_1 + a_2b_2, \\ \mathbf{a} \wedge \mathbf{b} &= (a_1b_2 - a_2b_1)\mathbf{e}_1\mathbf{e}_2 = (a_1b_2 - a_2b_1)\mathbf{e}_{12}.\end{aligned}$$

All bivectors in two dimensions are of this form, that is multiples of the bivector $\mathbf{e}_1\mathbf{e}_2$, written $\mathbf{e}_{12}$ to emphasise it is a bivector rather than a vector. The magnitude of $\mathbf{e}_{12}$ is 1, with

$$\mathbf{e}_{12}^2 = -1,$$

so it is called the **unit bivector**. The term unit bivector can be used in other dimensions but it is only uniquely defined in two dimensions and all bivectors are multiples of $\mathbf{e}_{12}$. As the highest grade element of the algebra $\mathbf{e}_{12}$ is also the pseudoscalar which is given the symbol i.

Complex numbers

With the properties of negative square and unit magnitude the unit bivector can be identified with the imaginary unit from complex numbers. The bivectors and scalars together form the even subalgebra of the geometric algebra, which is isomorphic to the complex numbers $\mathbb{C}$. The even subalgebra has basis $(1, \mathbf{e}_{12})$, the whole algebra has basis $(1, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_{12})$.

The complex numbers are usually identified with the coordinate axes and two dimensional vectors, which would mean associating them with the vector elements of the geometric algebra. There is no contradiction in this, as to get from a general vector to a complex number an axis needs to be identified as the real axis, $\mathbf{e}_1$ say. This multiplies by all vectors to generate the elements of even subalgebra.

All the properties of complex numbers can be derived from bivectors, but two are of particular interest. First as with complex numbers products of bivectors and so the even subalgebra are commutative. This is only true in two dimensions, so properties of the bivector in two dimensions that depend on commutativity do not usually generalise to higher dimensions.

Second a general bivector can be written

$$\theta \mathbf{e}_{12} = i\theta,$$

where θ is a real number. Putting this into the Taylor series for the exponential map and using the property $\mathbf{e_{12}}^2 = -1$ results in a bivector version of Euler's formula,

$$e^{\theta \mathbf{e}_{12}} = e^{i\theta} = \cos\theta + i\sin\theta,$$

which when multiplied by any vector rotates it through an angle θ about the origin:

$$(x'\mathbf{e}_1 + y'\mathbf{e}_2) = (x\mathbf{e}_1 + y\mathbf{e}_2)e^{i\theta}.$$

The product of a vector with a bivector in two dimensions is anticommutative, so the following products all generate the same rotation

$$\mathbf{v}' = \mathbf{v}e^{i\theta} = e^{-i\theta}\mathbf{v} = e^{\frac{-i\theta}{2}}\mathbf{v}e^{\frac{i\theta}{2}}.$$

Of these the last product is the one that generalises into higher dimensions. The quantity needed is called a rotor and is given the symbol R, so in two dimensions a rotor that rotates through angle θ can be written

$$R = e^{\frac{-i\theta}{2}} = e^{\frac{-\theta \mathbf{e}_{12}}{2}},$$

and the rotation it generates is

$$\mathbf{v}' = R\mathbf{v}R^{-1}.$$

Three dimensions

In three dimensions the geometric product of two vectors is

$$\begin{aligned}\mathbf{ab} &= (a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3)(b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + b_3\mathbf{e}_3) \\ &= a_1b_1{\mathbf{e}_1}^2 + a_2b_2{\mathbf{e}_2}^2 + a_3b_3{\mathbf{e}_3}^2 + (a_2b_3 - a_3b_2)\mathbf{e}_2\mathbf{e}_3 + (a_3b_1 - a_1b_3)\mathbf{e}_3\mathbf{e}_1 + (a_1b_2 - a_2b_1)\mathbf{e}_1\mathbf{e}_2.\end{aligned}$$

This can be split into the symmetric, scalar valued, interior product and the antisymmetric, bivector valued, exterior product:

$$\begin{aligned}\mathbf{a} \cdot \mathbf{b} &= a_1b_1 + a_2b_2 + a_3b_3 \\ \mathbf{a} \wedge \mathbf{b} &= (a_2b_3 - a_3b_2)\mathbf{e}_{23} + (a_3b_1 - a_1b_3)\mathbf{e}_{31} + (a_1b_2 - a_2b_1)\mathbf{e}_{12}.\end{aligned}$$

In three dimensions all bivectors are simple and so the result of an exterior product. The unit bivectors $\mathbf{e}_{23}$, $\mathbf{e}_{31}$ and $\mathbf{e}_{12}$ form a basis for the space of bivectors $\Lambda^2\mathbb{R}^3$, which itself a three dimensional linear space. So if a general bivector is:

$$\mathbf{A} = A_{23}\mathbf{e}_{23} + A_{31}\mathbf{e}_{31} + A_{12}\mathbf{e}_{12},$$

they can be added like vectors

$$\mathbf{A} + \mathbf{B} = (A_{23} + B_{23})\mathbf{e}_{23} + (A_{31} + B_{31})\mathbf{e}_{31} + (A_{12} + B_{12})\mathbf{e}_{12}.$$

while when multiplied they produce the following

$$\mathbf{AB} = -A_{23}B_{23} - A_{31}B_{31} - A_{12}B_{12} + (A_{12}B_{31} - A_{31}B_{12})\mathbf{e}_{23} + (A_{23}B_{12} - A_{12}B_{23})\mathbf{e}_{31} + (A_{31}B_{23} - A_{23}B_{31})\mathbf{e}_{12}$$

which can be split into symmetric scalar and antisymmetric bivector parts as follows

$$\begin{aligned}\mathbf{A} \cdot \mathbf{B} &= -A_{12}B_{12} - A_{31}B_{31} - A_{23}B_{23} \\ \mathbf{A} \times \mathbf{B} &= (A_{23}B_{13} - A_{13}B_{23})\mathbf{e}_{12} + (A_{12}B_{23} - A_{23}B_{12})\mathbf{e}_{13} + (A_{13}B_{12} - A_{12}B_{13})\mathbf{e}_{23}.\end{aligned}$$

The exterior product of two bivectors in three dimensions is zero.

A bivector **B** can be written as the product of its magnitude and a unit bivector, so writing β for $|\mathbf{B}|$ and using the Taylor series for the exponential map it can be shown that

$$e^{\mathbf{B}} = e^{\beta \frac{\mathbf{B}}{\beta}} = \cos\beta + \frac{\mathbf{B}}{\beta}\sin\beta.$$

This is another version of Euler's formula, but with a general bivector in three dimensions. Unlike in two dimensions bivectors are not commutative so properties that depend on commutativity do not apply in three dimensions. For example in general $e^{\mathbf{A}+\mathbf{B}} \neq e^{\mathbf{A}}e^{\mathbf{B}}$ in three (or more) dimensions.

The full geometric algebra in three dimensions, $C\ell_3(\mathbb{R})$, has basis (1, $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, $\mathbf{e}_{23}$, $\mathbf{e}_{31}$, $\mathbf{e}_{12}$, $\mathbf{e}_{123}$). The element $\mathbf{e}_{123}$ is a trivector and the pseudoscalar for the geometry. Bivectors in three dimensions are sometimes identified with pseudovectors to which they are related, as discussed below.

Quaternions

Bivectors are not closed under the geometric product, but the even subalgebra is. In three dimensions it consists of all scalar and bivector elements of the geometric algebra, so a general element can be written for example $a + \mathbf{A}$, where a is the scalar part and $\mathbf{A}$ is the bivector part. It is written $C\ell^+_3$ and has basis (1, $\mathbf{e}_{23}$, $\mathbf{e}_{31}$, $\mathbf{e}_{12}$). The product of two general elements of the even subalgebra is

$$(a + \mathbf{A})(b + \mathbf{B}) = ab + a\mathbf{B} + b\mathbf{A} + \mathbf{A}\cdot\mathbf{B} + \mathbf{A}\times\mathbf{B}.$$

The even subalgebra, that is the algebra consisting of scalars and bivectors, is isomorphic to the quaternions, $\mathbb{H}$. This can be seen by comparing the basis to the quaternion basis, or from the above product which is identical to the quaternion product, except for a change of sign which relates to the negative products in the bivector interior product $\mathbf{A} \cdot \mathbf{B}$. Other quaternion properties can be similarly related to or derived from geometric algebra.

This suggests that the usual split of a quaternion into scalar and vector parts would be better represented as a split into scalar and bivector parts; if this is done there is no special quaternion product, there is just the normal geometric product on the elements. It also relates quaternions in three dimensions to complex numbers in two, as each is isomorphic to the even subalgebra for the dimension, a relationship that generalises to higher dimensions.

Rotation vector

The rotation vector, from the axis angle representation of rotations, is a compact way of representing rotations in three dimensions. In its most compact form it consists of a vector, the product of the a unit vector that is the axis of rotation and the angle of rotation, so the magnitude of the vector is the rotation angle.

In geometric algebra this vector is classified as a bivector. This can be seen in its relation to quaternions. If the axis is ω and the angle of rotation is θ then the rotation vector is $\omega\theta$ quaternion associated with the rotation is

$$q = \left(\cos\left(\frac{\theta}{2}\right), \omega \sin\left(\frac{\theta}{2}\right)\right)$$

but this is just the exponent of half of the bivector $\mathbf{\Omega}\theta$, that is

$$e^{\frac{\mathbf{\Omega}\theta}{2}} = \cos\left(\frac{\theta}{2}\right) + \mathbf{\Omega} \sin\left(\frac{\theta}{2}\right)$$

So rotation vectors are bivectors, just as quaternions are elements of the geometric algebra, and they are related by the exponential map in that algebra.

Rotors

The bivector $\mathbf{\Omega}\theta$ generates a rotation through the exponential map. The even elements generated rotate a general vector in three dimensions in the same way as quaternions:

$$\mathbf{v}' = e^{\frac{\mathbf{\Omega}\theta}{2}} \mathbf{v} e^{-\frac{\mathbf{\Omega}\theta}{2}}.$$

As to two dimensions the quantity $e^{\mathbf{\Omega}\theta}$ is called a rotor and written R. The quantity $e^{-\mathbf{\Omega}\theta}$ is then R^{-1}, and they generate rotations as follows

$$\mathbf{v}' = R\mathbf{v}R^{-1}.$$

This is identical to two dimensions, except here rotors are four-dimensional objects isomorphic to the quaternions. This can be generalised to all dimensions, with rotors, elements of the even subalgebra with unit magnitude, being generated by the exponential map from bivectors. They form a double cover over the rotation group, so the rotors R and $-R$ represent the same rotation.

Matrices

Bivectors are isomorphic to skew-symmetric matrices; the general bivector $B_{23}\mathbf{e}_{23} + B_{31}\mathbf{e}_{31} + B_{12}\mathbf{e}_{12}$ maps to the matrix

$$M_B = \begin{pmatrix} 0 & -B_{12} & B_{31} \\ B_{12} & 0 & -B_{23} \\ -B_{31} & B_{23} & 0 \end{pmatrix}.$$

This multiplied by vectors on both sides gives the same vector as the product of a vector and bivector; an example is the angular velocity tensor.

Skew symmetric matrices generate orthogonal matrices with determinant 1 through the exponential map. In particular the exponent of a bivector associated with a rotation is a rotation matrix, that is the rotation matrix M_R given by the above skew-symmetric matrix is

$$M_R = e^{M_B}.$$

The rotation described by M_R is the same as that described by the rotor R given by

$$R = e^{\frac{B}{2}},$$

and the matrix M_R can be also calculated directly from rotor R:

$$M_R = \begin{pmatrix} (R\mathbf{e}_1R^{-1})\cdot\mathbf{e}_1 & (R\mathbf{e}_2R^{-1})\cdot\mathbf{e}_1 & (R\mathbf{e}_3R^{-1})\cdot\mathbf{e}_1 \\ (R\mathbf{e}_1R^{-1})\cdot\mathbf{e}_2 & (R\mathbf{e}_2R^{-1})\cdot\mathbf{e}_2 & (R\mathbf{e}_3R^{-1})\cdot\mathbf{e}_2 \\ (R\mathbf{e}_1R^{-1})\cdot\mathbf{e}_3 & (R\mathbf{e}_2R^{-1})\cdot\mathbf{e}_3 & (R\mathbf{e}_3R^{-1})\cdot\mathbf{e}_3 \end{pmatrix}.$$

Bivectors are related to the eigenvalues of a rotation matrix. Given a rotation matrix M the eigenvalues can calculated by solving the characteristic equation for that matrix $0 = \det(M - \lambda I)$. By the fundamental theorem of algebra this has three roots, but only one real root as there is only one eigenvector, the axis of rotation. The other roots must be a complex conjugate pair. They have unit magnitude so purely imaginary logarithms, equal to the magnitude of the bivector associated with the rotation, which is also the angle of rotation. The eigenvectors associated with the complex eigenvalues are in the plane of the bivector, so the exterior product of two non-parallel eigenvectors result in the bivector, or at least a multiple of it.

Axial vectors

The rotation vector is an example of an axial vector. Axial vectors or pseudovectors are vectors that undergo a sign change compared to normal or polar vectors under inversion, that is when reflected or otherwise inverted. Examples include quantities like torque, angular momentum and vector magnetic fields. Such quantities can be described as bivectors in geometric algebra; that is quantities that might use axial vectors in vector algebra are better represented by bivectors in geometric algebra. More precisely, the Hodge dual gives the isomorphism between axial vectors and bivectors, so each axial vector is associated with a bivector and vice-versa; that is

$$\mathbf{A} = *\mathbf{a}\ ; \quad \mathbf{a} = *\mathbf{A}$$

where * indicates the Hodge dual. Alternately, using the unit pseudoscalar in $C\ell_3(\mathbb{R})$, $i = \mathbf{e}_1\mathbf{e}_2\mathbf{e}_3$ gives

$$\mathbf{A} = \mathbf{a}i\ ; \quad \mathbf{a} = -\mathbf{A}i.$$

This is easier to use as the product is just the geometric product. But it is antisymmetric because (as in two dimensions) the unit pseudoscalar i squares to -1, so a negative is needed in one of the products.

This relationship extends to operations like the vector valued cross product and bivector valued exterior product, as when written as determinants they are calculated in the same way:

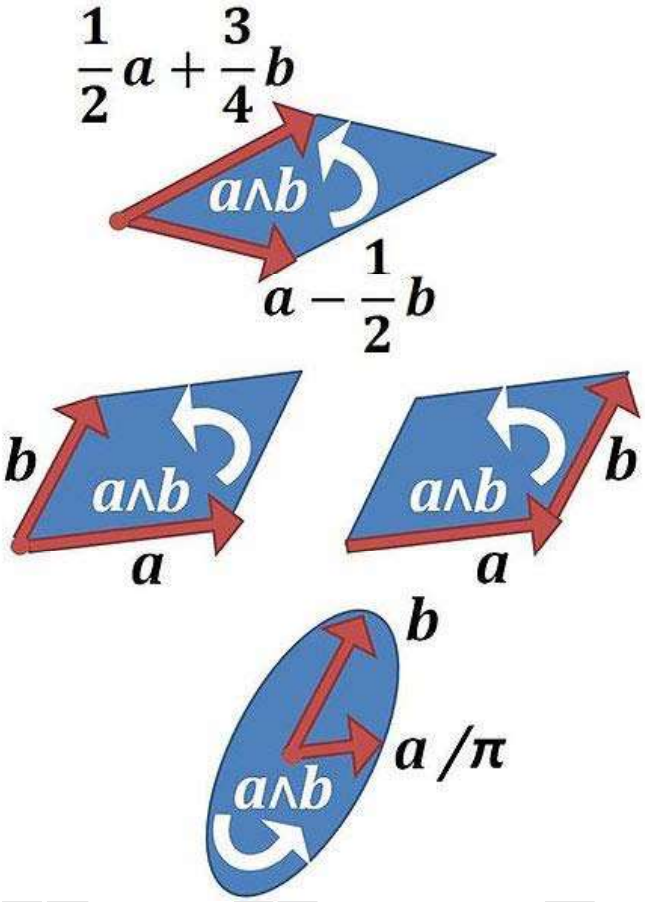
$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}; \quad \mathbf{a} \wedge \mathbf{b} = \begin{vmatrix} \mathbf{e}_{23} & \mathbf{e}_{31} & \mathbf{e}_{12} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix},$$

so are related by the Hodge dual:

$$*(\mathbf{a} \wedge \mathbf{b}) = \mathbf{a} \times \mathbf{b}\ ; \quad *(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \wedge \mathbf{b}.$$

Bivectors have a number of advantages over axial vectors. They better disambiguate axial and polar vectors, that is the quantities represented by them, so it is clearer which operations are allowed and what their results are. For example the inner product of a polar vector and a axial vector resulting from the cross product in the triple product should result in a pseudoscalar, a result which is more obvious if the calculation is framed as the exterior product of a vector and bivector. They generalises to other dimensions; in particular bivectors can be used to describe quantities like torque and angular momentum in two as well as three dimensions. Also, they closely match geometric intuition in a number of ways, as seen in the next section.

Geometric interpretation



Parallel plane segments with the same orientation and area corresponding to the same bivector **a** ∧ **b**.

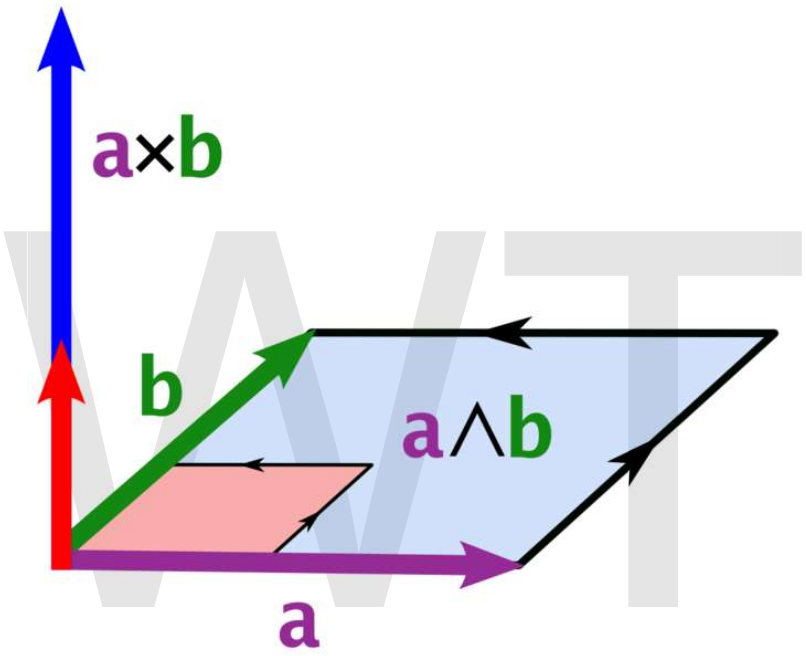
As suggested by their name and that of the algebra, one of the attractions of bivectors is that they have a natural geometric interpretation. This can be described in any dimension but is best done in three where parallels can be drawn with more familiar objects, before being applied to higher dimensions. In two dimensions the geometric interpretation is trivial, as the space is two dimensional so has only one plane and all bivectors are associated with it differing only by a scale factor.

All bivectors can be interpreted as planes, or more precisely as directed plane segments. In three dimensions there are three properties of a bivector that can be interpreted geometrically:

- The arrangement of the plane in space, precisely the attitude of the plane (or alternately the rotation, geometric orientation or gradient of the plane), is associated with the ratio of the bivector components. In particular the three basis bivectors, $\mathbf{e}_{23}$, $\mathbf{e}_{31}$ and $\mathbf{e}_{12}$, or scalar multiples of them, are associated with the *yz*-plane, *xz*-plane and *xy*-plane respectively.
- The magnitude of the bivector is associated with the area of the plane segment. The area does not have a particular shape so any shape can be used. It can even be

represented in other ways, such as by an angular measure. But if the vectors are interpreted as lengths the bivector is usually interpreted as an area with the same units, as follows.

- Like the direction of a vector a plane associated with a bivector has a direction, a circulation or a sense of rotation in the plane, which takes two values seen as clockwise and counterclockwise when viewed from viewpoint not in the plane. This is associated with a change of sign in the bivector, that is if the direction is reversed the bivector is negated. Alternately if two bivectors have the same attitude and magnitude but opposite directions then one is the negative of the other.



The cross product **a** × **b** is orthogonal to the bivector **a** ∧ **b**.

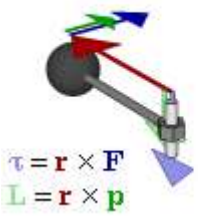
In three dimensions all bivectors can be generated by the exterior product of two vectors. If the bivector **B** = **a** ∧ **b** then the magnitude of **B** is

$$|\mathbf{B}| = |\mathbf{a}|\,|\mathbf{b}| \sin\theta,$$

where θ is the angle between the vectors. This is the area of the parallelogram with edges **a** and **b**, as shown in the diagram. One interpretation is that the area is swept out by **b** as it moves along **a**. The exterior product is antisymmetric, so reversing the order of **a** and **b** to make **a** move along **b** results in a bivector with the opposite direction that is the negative of the first. The plane of bivector **a** ∧ **b** contains both **a** and **b** so they are both parallel to the plane.

Bivectors and axial vectors are related by Hodge dual. In a real vector space the Hodge dual relates a subspace to its orthogonal complement, so if a bivector is represented by a

plane then the axial vector associated with it is simply the plane's surface normal. The plane has two normals, one on each side, giving the two possible orientations for the plane and bivector.



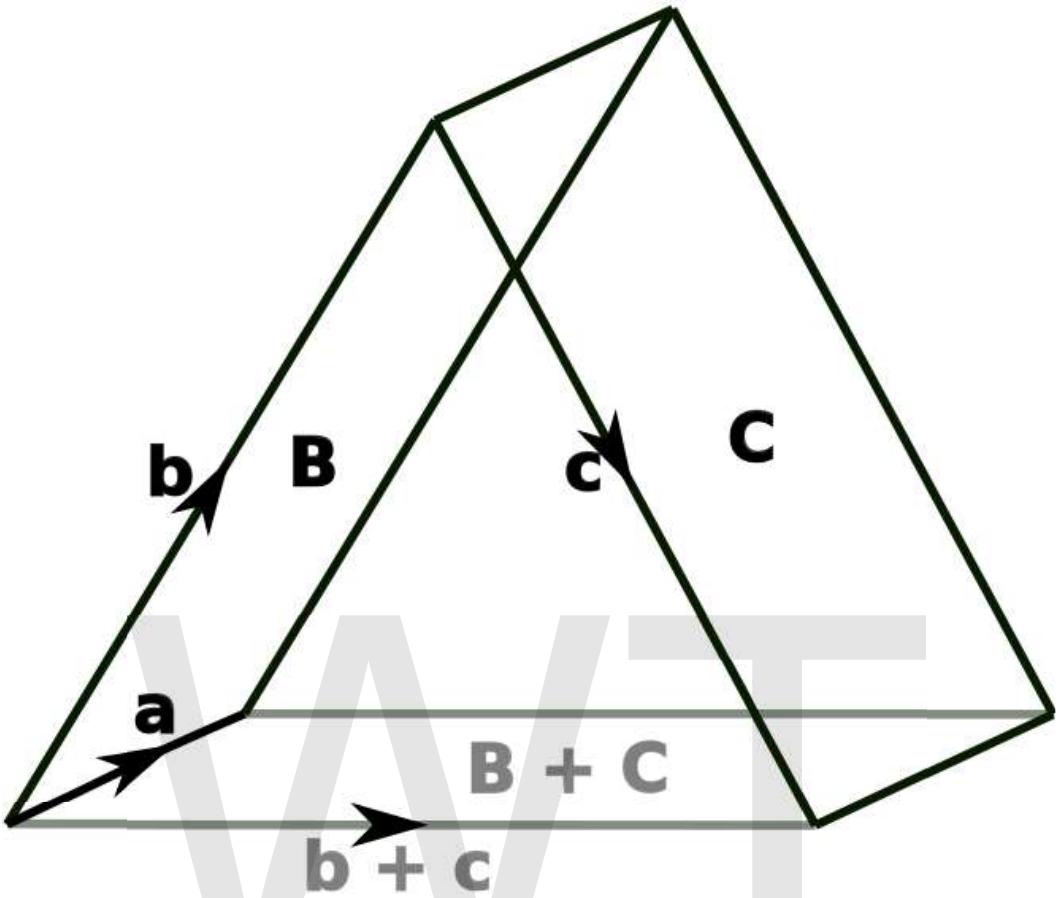
Relationship between force **F**, torque **τ**, linear momentum **p**, and angular momentum **L**.

This relates the cross product to the exterior product. It can also be used to represent physical quantities, like torque and angular momentum. In vector algebra they are usually represented by vectors, perpendicular to the plane of the force, linear momentum or displacement that they are calculated from. But if a bivector is used instead the plane is the plane of the bivector, so is a more natural way to represent the quantities and the way they act. It also unlike the vector representation generalises into other dimensions.

The product of two bivectors has a geometric interpretation. For non-zero bivectors **A** and **B** the product can be split into symmetric and antisymmetric parts as follows:

$$\mathbf{AB} = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \times \mathbf{B}.$$

Like vectors these have magnitudes $|\mathbf{A} \cdot \mathbf{B}| = |\mathbf{A}||\mathbf{B}| \cos \theta$ and $|\mathbf{A} \times \mathbf{B}| = |\mathbf{A}||\mathbf{B}| \sin \theta$, where θ is the angle between the planes. In three dimensions it is the same as the angle between the normal vectors dual to the planes, and it generalises to some extent in higher dimensions.



Two bivectors, two of the non-parallel sides of a prism, being added to give a third bivector

Bivectors can be added together as areas. Given two non-zero bivectors **B** and **C** in three dimensions it is always possible to find a vector that is contained in both, **a** say, so the bivectors can be written as exterior products involving **a**:

$$\mathbf{B} = \mathbf{a} \wedge \mathbf{b}$$
$$\mathbf{C} = \mathbf{a} \wedge \mathbf{c}$$

This can be interpreted geometrically as seen in the diagram: the two areas sum to give a third, with the three areas forming faces of a prism with **a**, **b**, **c** and **b** + **c** as edges. This corresponds to the two ways of calculating the area using the distributivity of the exterior product:

$$\begin{aligned}\mathbf{B} + \mathbf{C} &= \mathbf{a} \wedge \mathbf{b} + \mathbf{a} \wedge \mathbf{c} \\ &= \mathbf{a} \wedge (\mathbf{b} + \mathbf{c}).\end{aligned}$$

This only works in three dimensions as it is the only dimension where a vector parallel to both bivectors must exist. In higher dimensions bivectors generally are not associated

with a single plane, or if they are (simple bivectors) two bivectors may have no vector in common, and so sum to a non-simple bivector.

Four dimensions

In four dimensions the basis elements for the space $\Lambda^2\mathbb{R}^4$ of bivectors are ($\mathbf{e}_{12}$, $\mathbf{e}_{13}$, $\mathbf{e}_{14}$, $\mathbf{e}_{23}$, $\mathbf{e}_{24}$, $\mathbf{e}_{34}$), so a general bivector is of the form

$$\mathbf{A} = a_{12}\mathbf{e}_{12} + a_{13}\mathbf{e}_{13} + a_{14}\mathbf{e}_{14} + a_{23}\mathbf{e}_{23} + a_{24}\mathbf{e}_{24} + a_{34}\mathbf{e}_{34}.$$

Orthogonality

In four dimensions bivectors are orthogonal to bivectors. That is the dual of a bivector is a bivector, and the space $\Lambda^2\mathbb{R}^4$ is dual to itself in $C\ell_4(\mathbb{R})$. Normal vectors are not unique, instead every plane is orthogonal to all the vectors in its dual space. This can be used to partition the bivectors into two 'halves', for example into two sets of three unit bivectors each. There are only four distinct ways to do this, and whenever its done one vector is in only one of the two halves, for example ($\mathbf{e}_{12}$, $\mathbf{e}_{13}$, $\mathbf{e}_{14}$) and ($\mathbf{e}_{23}$, $\mathbf{e}_{24}$, $\mathbf{e}_{34}$).

Simple bivectors in 4D

In four dimensions bivectors are generated by the exterior product of vectors in $\mathbb{R}^4$, but with one important difference from $\mathbb{R}^3$ and $\mathbb{R}^2$. In four dimensions not all bivectors are simple. There are bivectors such as $\mathbf{e}_{12} + \mathbf{e}_{34}$ that cannot be generated by the external product of two vectors. This also means they do not have a real, that is scalar, square. In this case

$$(\mathbf{e}_{12} + \mathbf{e}_{34})^2 = \mathbf{e}_{12}\mathbf{e}_{12} + \mathbf{e}_{12}\mathbf{e}_{34} + \mathbf{e}_{34}\mathbf{e}_{12} + \mathbf{e}_{34}\mathbf{e}_{34} = -2 + 2\mathbf{e}_{1234}.$$

The element $\mathbf{e}_{1234}$ is the pseudoscalar in $C\ell_4$, distinct from the scalar, so the square is non-scalar.

All bivectors in four dimensions can be generated using at most two exterior products and four vectors. The above bivector can be written as

$$\mathbf{e}_{12} + \mathbf{e}_{34} = \mathbf{e}_1 \wedge \mathbf{e}_2 + \mathbf{e}_3 \wedge \mathbf{e}_4.$$

Alternately every bivector can be written as the sum of two simple bivectors. It is useful to choose two orthogonal bivectors for this, and this is always possible to do. Moreover for a general bivector the choice of simple bivectors is unique, that is there is only one way to decompose into orthogonal bivectors. This is true also for simple bivectors, except one of the orthogonal parts is zero. The exception is when the two orthogonal bivectors have equal magnitudes (as in the above example): in this case the decomposition is not unique.

Rotations in $\mathbb{R}^4$

As in three dimensions bivectors in four dimension generate rotations through the exponential map, and all rotations can be generated this way. As in three dimensions if **B** is a bivector then the rotor R is $e^{\mathbf{B}/2}$ and rotations are generated in the same way:

$$v' = RvR^{-1}.$$

A 3D projection of an tesseract performing a isoclinic rotation

The rotations generated are more complex though. They can be categorised as follows:

simple rotations are those that fix a plane in 4D, and rotate by an angle "about" this plane.
double rotations have only one fixed point, the origin, and rotate through two angles about two orthogonal planes. In general the angles are different and the planes are uniquely specified
isoclinic rotations are double rotations where the angles of rotation are equal. In this case the planes about which the rotation is taking place are not unique.

These are generated by bivectors in a straightforward way. Simple rotations are generated by simple bivectors, with the fixed plane the dual or orthogonal to the plane of the bivector. The rotation can be said to take place about that plane, in the plane of the bivector. All other bivectors generate double rotations, with the two angles of the rotation equalling the magnitudes of the two simple bivectors the non-simple bivector is composed of. Isoclinic rotations arise when these magnitudes are equal, in which case the decomposition into two simple bivectors in not unique.

Bivectors in general do not commute, but one exception is orthogonal bivectors and exponents of them. So if the bivector $\mathbf{B} = \mathbf{B}_1 + \mathbf{B}_2$, where $\mathbf{B}_1$ and $\mathbf{B}_2$ are orthogonal simple bivectors, is used to generate a rotation it decomposes into two simple rotations that commute as follows:

$$R = e^{\frac{\mathbf{B}_1+\mathbf{B}_2}{2}} = e^{\frac{\mathbf{B}_1}{2}} e^{\frac{\mathbf{B}_2}{2}} = e^{\frac{\mathbf{B}_2}{2}} e^{\frac{\mathbf{B}_1}{2}}$$

It is always possible to do this as all bivectors can be expressed as sums of orthogonal bivectors.

Spacetime rotations

Spacetime is a mathematical model for our universe used in special relativity. It consists of three space dimensions and one time dimension combined into a single four dimensional space. It is naturally described using geometric algebra and bivectors, with the Euclidean metric is replaced by a Minkowski metric. That is the algebra is identical to that of Euclidean space, except the signature is changed, so

$$\mathbf{e}_i^2 = \begin{cases} 1, & i = 1, 2, 3 \\ -1, & i = 4 \end{cases}$$

(Note the order and indices above are not universal – here $\mathbf{e}_4$ is the time-like dimension). The geometric algebra is $C\ell_{3,1}(\mathbb{R})$, and the subspace of bivectors is $\Lambda^2\mathbb{R}^{3,1}$. The bivectors are of two types. The bivectors $\mathbf{e}_{23}$, $\mathbf{e}_{31}$ and $\mathbf{e}_{12}$ have negative squares and correspond to the bivectors of the three dimensional subspace corresponding to Euclidean space, $\mathbb{R}^3$. These bivectors generate normal rotations in $\mathbb{R}^3$.

The bivectors $\mathbf{e}_{14}$, $\mathbf{e}_{24}$ and $\mathbf{e}_{34}$ have positive squares and as planes span a space dimension and the time dimension. These also generate rotations through the exponential map, but instead of trigonometric functions hyperbolic functions are needed, which generates a rotor as follows:

$$e^{\frac{\mathbf{\Omega}\theta}{2}} = \cosh\left(\frac{\theta}{2}\right) + \mathbf{\Omega}\sinh\left(\frac{\theta}{2}\right).$$

These are Lorentz transformations, expressed in a particularly compact way, using the same algebra as in $\mathbb{R}^3$ and $\mathbb{R}^4$. In general all spacetime rotations are generated from bivectors through the exponential map, that is a general rotor generated by bivector $\mathbf{A}$ is is of the form

$$R = e^{\frac{\mathbf{A}}{2}}.$$

The set of all rotations in spacetime form the Lorentz group, and from them most of the consequences of special relativity can be deduced. More generally this show how transformations in Euclidean space and spacetime can all be described using the same algebra.

Maxwell's equations

(Note: in this section traditional 3-vectors are indicated by lines over the symbols and spacetime vector and bivectors by bold symbols, with the vectors **J** and **A** exceptionally in uppercase)

Maxwell's equations are used in physics to describe the relationship between electric and magnetic fields. Normally given as four differential equations they have a particularly compact form when the fields are expressed as a spacetime bivector from $\Lambda^2\mathbb{R}^{3,1}$. If the electric and magnetic fields in $\mathbb{R}^3$ are E and B then the *electromagnetic bivector* is

$$\mathbf{F} = \frac{1}{c}\overline{E}\mathbf{e}_4 + \overline{B}\mathbf{e}_{123},$$

where $\mathbf{e}_4$ is again the basis vector for the time-like dimension and c is the speed of light. The quantity $B\mathbf{e}_{123}$ is the bivector dual to B in three dimensions, as discussed above, while $E\mathbf{e}_4$ as a product of orthogonal vectors is also bivector valued. As a whole it is the electromagnetic tensor expressed more compactly as a bivector, and is used as follows. First it is related to the 4-current **J**, a vector quantity given by

$$\mathbf{J} = \overline{j} + c\rho\mathbf{e}_4,$$

where j is current density and ρ is charge density. They are related by a differential operator ∂, which is

$$\partial = \nabla - \mathbf{e}_4\frac{1}{c}\frac{\partial}{\partial t}.$$

The operator ∇ is a differential operator in geometric algebra, acting on the space dimensions and given by $\nabla\mathbf{M} = \nabla\cdot\mathbf{M} + \nabla\wedge\mathbf{M}$. When applied to vectors $\nabla\cdot\mathbf{M}$ is the divergence and $\nabla\wedge\mathbf{M}$ is the curl but with a bivector rather than vector result, that is dual in three dimensions to the curl. For general quantity **M** they act as grade lowering and raising differential operators. In particular if **M** is a scalar then this operator is just the gradient, and it can be thought of as a geometric algebraic del operator.

Together these can be used to give a particularly compact form for Maxwell's equations in a vacuum:

$$\partial\mathbf{F} = \mathbf{J}.$$

This when decomposed according to geometric algebra, using geometric products which have both grade raising and grade lowering effects, is equivalent to Maxwell's four equations. This is the form in a vacuum, but the general form is only a little more complex. It is also related to the electromagnetic four-potential, a vector **A** given by

$$\mathbf{A} = \overline{A} + \frac{1}{c} V \mathbf{e}_4,$$

where A is the vector magnetic potential and V is the electric potential. It is related to the electromagnetic bivector as follows

$$\partial \mathbf{A} = -\mathbf{F},$$

using the same differential operator ∂.

Higher dimensions

As has been suggested in earlier sections much of geometric algebra generalises well into higher dimensions. The geometric algebra for the real space $\mathbb{R}^n$ is $C\ell_n(\mathbb{R})$, and the subspace of bivectors is $\Lambda^2\mathbb{R}^n$.

The number of simple bivectors needed to form a general bivector rises with the dimension, so for n odd it is $(n - 1) / 2$, for n even it is $n / 2$. So for four and five dimensions only two simple bivectors are needed but three are required for six and seven dimensions. For example in six dimensions with standard basis ($\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, $\mathbf{e}_4$, $\mathbf{e}_5$, $\mathbf{e}_6$) the bivector

$$\mathbf{e}_{12} + \mathbf{e}_{34} + \mathbf{e}_{56}$$

is the sum of three simple bivectors but no less. As in four dimensions it is always possible to find orthogonal simple bivectors for this sum.

Rotations in higher dimensions

As in three and four dimensions rotors are generated by the exponential map, so

$$e^{\frac{\mathbf{B}}{2}}$$

is the rotor generated by bivector **B**. Simple rotations, that take place in a plane of rotation around a fixed blade of dimension $(n - 2)$ are generated by with simple bivectors, while other bivectors generate more complex rotations which can be described in terms of the simple bivectors they are sums of, each related to a plane of rotation. All bivectors can be expressed as the sum of orthogonal and commutative simple bivectors, so rotations can always be decomposed into a set of commutative rotations about the planes associated with these bivectors. The group of the rotors in n dimensions is the spin group, Spin(n).

One notable feature, related to the number of simple bivectors and so rotation planes, is that in odd dimensions every rotation has a fixed axis - it is misleading to call it an axis of rotation as in higher dimensions rotations are taking place in multiple planes orthogonal to it. This is related to bivectors, as bivectors in odd dimensions decompose into the same

number of bivectors as the even dimension below, so have the same number of planes, but one extra dimension. As each plane generates rotations in two dimensions in odd dimensions there must be one dimension, that is an axis, that is not being rotated.

Bivectors are also related to the rotation matrix in n dimensions. As in three dimensions the characteristic equation of the matrix can be solved to find the eigenvalues. In odd dimensions this has one real root, with eigenvector the fixed axis, and in even dimensions it has no real roots, so either all or all but one of the roots are complex conjugate pairs. Each pair is associated with a simple component of the bivector associated with the rotation. In particular the log of each pair is ± the magnitude, while eigenvectors generated from the roots are parallel to and so can be used to generate the bivector. In general the eigenvalues and bivectors are unique, and the set of eigenvalues gives the full decomposition into simple bivectors; if roots are repeated then the decomposition of the bivector into simple bivectors is not unique.

Projective geometry

Geometric algebra can be applied to projective geometry in a straightforward way. The geometric algebra used is $C\ell_n(\mathbb{R})$, $n \geq 3$, the algebra of the real vector space $\mathbb{R}^n$. This is used to describe objects in the real projective space $\mathbb{RP}^{n-1}$. The non-zero vectors in $C\ell_n(\mathbb{R})$ or $\mathbb{R}^n$ are associated with points in the projective space so vectors that differ only by a scale factor, so their exterior product is zero, map to the same point. Non-zero simple bivectors in $\Lambda^2\mathbb{R}^n$ represent lines in $\mathbb{RP}^{n-1}$, with bivectors differing only by a (positive or negative) scale factor representing the same line.

A description of the projective geometry can be constructed in the geometric algebra using basic operations. For example given two distinct points in $\mathbb{RP}^{n-1}$ represented by vectors **a** and **b** the line between them is given by **a** ∧ **b** (or **b** ∧ **a**). Two lines intersect in a point if **A** ∧ **B** = 0 for their bivectors **A** and **B**. This point is given by the vector

$$\mathbf{p} = \mathbf{A} \vee \mathbf{B} = (\mathbf{A} \times \mathbf{B}) J^{-1}.$$

The operation "∨" is the meet, which can be defined as above in terms of the join, $J = \mathbf{A} \wedge \mathbf{B}$ for non-zero **A** ∧ **B**. Using these operations projective geometry can be formulated in terms of geometric algebra. For example given a third (non-zero) bivector **C** the point **p** lies on the line given by **C** if and only if

$$\mathbf{p} \wedge \mathbf{C} = 0.$$

So the condition for the lines given by **A**, **B** and **C** to be collinear is

$$(\mathbf{A} \vee \mathbf{B}) \wedge \mathbf{C} = 0,$$

which in $C\ell_3(\mathbb{R})$ and $\mathbb{RP}^2$ simplifies to

$$\langle \mathbf{ABC} \rangle = 0,$$

where the angle brackets denote the scalar part of the geometric product. In the same way all projective space operations can be written in terms of geometric algebra, with bivectors representing general lines in projective space, so the whole geometry can be developed using geometric algebra.

Tensors and matrices

As noted above a bivector can be written as a skew-symmetric matrix, which through the exponential map generates a rotation matrix that describes the same rotation as the rotor, also generated by the exponential map but applied to the vector. But it is also used with other bivectors such as the angular velocity tensor and the electromagnetic tensor, respectively a 3x3 and 4x4 skew-symmetric matrix or tensor.

Real bivectors in $\Lambda^2\mathbb{R}^n$ are isomorphic to $n \times n$ skew-symmetric matrices, or alternately to antisymmetric tensors of order 2 on $\mathbb{R}^n$. While bivectors are isomorphic to vectors (via the dual) in three dimensions they can be represented by skew-symmetric matrices in any dimension. This is useful for relating bivectors to problems described by matrices, so they can be re-cast in terms of bivectors, given a geometric interpretation, then often solved more easily or related geometrically to other bivector problems.

More generally every real geometric algebra is isomorphic to a matrix algebra. These contain bivectors as a subspace, though often in a way which is not especially useful. These matrices are mainly of interest as a way of classifying Clifford algebras.

Chapter 17

Calculus of Moving Surfaces & Covariance and Contravariance of Vectors

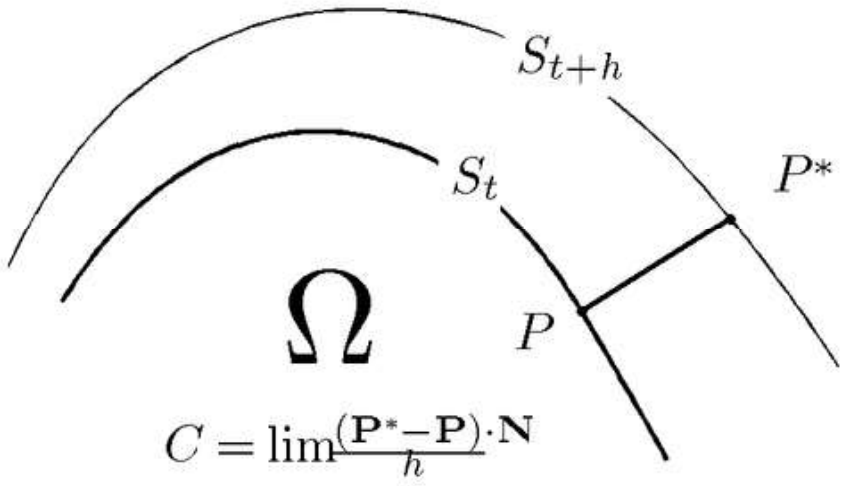
Calculus of moving surfaces

The **calculus of moving surfaces** (**CMS**) is an extension of the classical tensor calculus to deforming manifolds. Central to the CMS is the $\delta / \delta t$-derivative whose original definition was put forth by Jacques Hadamard. It plays the role analogous to that of the covariant derivative ∇_α on differential manifolds. In particular, it has the property that it produces a tensor when applied to a tensor.

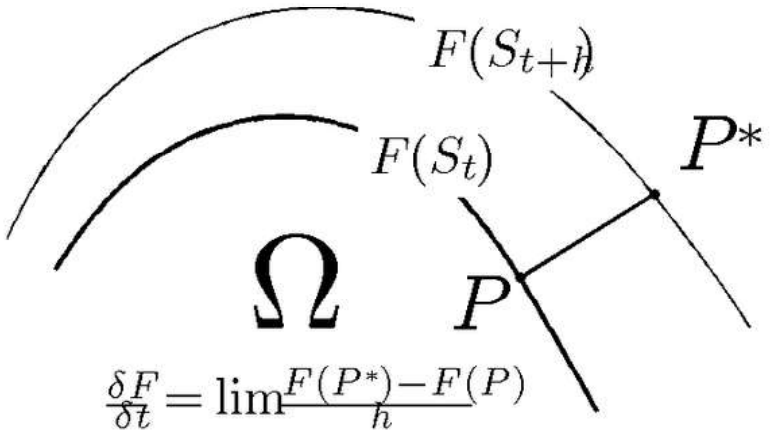
Suppose that S_t is the evolution of the surface S indexed by a time-like parameter t. The definitions of the surface velocity C and the operator $\delta / \delta t$ are the geometric foundations of the CMS. The velocity C is the rate of deformation of the surface S in the instantaneous normal direction. The value of C at a point P is defined as the limit

$$C = \lim_{h \to 0} \frac{\text{Distance}(P, P^*)}{h}$$

where P^* is the point on S_{t+h} that lies on the straight line perpendicular to S_t at point P. This definition is illustrated in the first geometric figure below. The velocity C is a signed quantity: it is positive when $\overline{PP^*}$ points in the direction of the chosen normal, and negative otherwise. The relationship between S_t and C is analogous to the relationship between location and velocity in elementary calculus: knowing either quantity allows one to construct the other by differentiation or integration.



Geometric construction of the surface velocity C



Geometric construction of the δ / δ*t*-derivative of an invariant field F

The δ / δ*t*-derivative for a scalar field F defined on S_t is the rate of change in *F* in the instantaneously normal direction:

$$\frac{\delta F}{\delta t} = \lim_{h \to 0} \frac{F(P^*) - F(P)}{h}$$

This definition is also illustrated in second geometric figure.

The above definitions are *geometric*. In analytical settings, direct application of these definitions may not be possible. The CMS gives *analytical* definitions of C and δ / δ*t* in terms of elementary operations from calculus and differential geometry.

Analytical definitions

For analytical definitions of *C* and δ / δ*t*, consider the evolution of *S* given by

$$Z^i = Z^i(t, S)$$

where Z^i are general curvilinear space coordinates and S^α are the surface coordinates. By convention, tensor indices of function arguments are dropped. Thus the above equations contains *S* rather than S^α.The velocity object v^i is defined as the partial derivative

$$v^i = \frac{\partial Z^i(t, S)}{\partial t}$$

The velocity *C* can be computed most directly by the formula

$$C = v^i N_i$$

where N_i are the covariant components of the normal vector $\vec{N}$.

The definition of the δ / δ*t*-derivative for an invariant *F* reads

$$\frac{\delta F}{\delta t} = \frac{\partial F\left(t, S\right)}{\partial t} - v^i Z_i^\alpha \nabla_\alpha F$$

where Z_i^α is the shift tensor and $\nabla_\alpha$is the covariant derivative on S.

For *tensors*, an appropriate generalization is needed. The proper definition for a representative tensor $T_{j\beta}^{i\alpha}$reads

$$\frac{\delta T_{j\beta}^{i\alpha}}{\delta t} = \frac{\partial T_{j\beta}^{i\alpha}}{\partial t} - v^\eta \nabla_\eta T_{j\beta}^{i\alpha} + v^m \Gamma_{mk}^i T_{j\beta}^{k\alpha} - v^m \Gamma_{mj}^k T_{k\beta}^{i\alpha} + \nabla_\eta v^\alpha T_{j\beta}^{i\eta} - \nabla_\beta v^\eta T_{j\eta}^{i\alpha}$$

where $\Gamma_{mj}^k$are Christoffel symbols.

Properties of the δ/δt-derivative

The δ / δ*t*-derivative commutes with contraction, satisfies the product rule for any collection of indices

$$\frac{\delta}{\delta t}\left(S_\alpha^i T_j^\beta\right) = \frac{\delta S_\alpha^i}{\delta t} T_j^\beta + S_\alpha^i \frac{\delta T_j^\beta}{\delta t}$$

and obeys a chain rule for surface restrictions of spatial tensors:

$$\frac{\delta F_k^j}{\delta t} = \frac{\partial F_k^j}{\partial t} + C N^i \nabla_i F_k^j$$

Chain rule shows that the δ / δ*t*-derivative of spatial "metrics" vanishes

$$\frac{\delta \delta_j^i}{\delta t}, \frac{\delta Z_{ij}}{\delta t}, \frac{\delta Z^{ij}}{\delta t}, \frac{\delta \varepsilon_{ijk}}{\delta t}, \frac{\delta \varepsilon^{ijk}}{\delta t} = 0$$

where Z_{ij} and Z^{ij} are covariant and contravariant metric tensors, $\delta_j^i$is the Kronecker delta symbol, and $\varepsilon_{ijk}$and $\varepsilon^{ijk}$are the Levi-Civita symbols. The main article on Levi-Civita

symbols describes them for Cartesian coordinate systems. The preceding rule is valid in general coordinates, where the definition of the Levi-Civita symbols must include the square root of the determinant of the covariant metric tensor Z_{ij}.

Differentiation table for the δ/δt-derivative

The δ / δ*t*-derivative of the key surface objects leads to highly concise and attractive formulas. When applied to the covariant surface metric tensor $S_{\alpha\beta}$ and the contravariant metric tensor $S^{\alpha\beta}$, the following identities result

$$\frac{\delta S_{\alpha\beta}}{\delta t} = -2CB_{\alpha\beta}$$

$$\frac{\delta S^{\alpha\beta}}{\delta t} = 2CB^{\alpha\beta}$$

where $B_{\alpha\beta}$ and $B^{\alpha\beta}$ are the doubly covariant and doubly contravariant curvature tensors. These curvature tensors, as well as for the mixed curvature tensor B^{α}_{β}, satisfy

$$\frac{\delta B_{\alpha\beta}}{\delta t} = \nabla_\alpha \nabla_\beta C - CB_{\alpha\gamma}B^{\gamma}_{\beta}$$

$$\frac{\delta B^{\alpha}_{\beta}}{\delta t} = \nabla^\alpha \nabla_\beta C + CB^{\alpha}_{\gamma}B^{\gamma}_{\beta}$$

$$\frac{\delta B^{\alpha\beta}}{\delta t} = \nabla^\alpha \nabla^\beta C + 3CB^{\alpha}_{\gamma}B^{\gamma\beta}$$

The shift tensor Z^{i}_{α} and the normal N^i satisfy

$$\frac{\delta Z^{i}_{\alpha}}{\delta t} = \nabla_\alpha \left(CN^i\right)$$

$$\frac{\delta N^i}{\delta t} = -Z^{i}_{\alpha}\nabla^\alpha C$$

Finally, the surface Levi-Civita symbols $\varepsilon_{\alpha\beta}$ and $\varepsilon^{\alpha\beta}$ satisfy

$$\frac{\delta \varepsilon_{\alpha\beta}}{\delta t} = -\varepsilon_{\alpha\beta} C B^{\gamma}_{\gamma}$$

$$\frac{\delta \varepsilon^{\alpha\beta}}{\delta t} = \varepsilon^{\alpha\beta} C B^{\gamma}_{\gamma}$$

Time differentiation of integrals

The CMS provides rules for time differentiation of volume and surface integrals.

Covariance and contravariance of vectors

In multilinear algebra and tensor analysis, **covariance** and **contravariance** describe how the quantitative description of certain geometric or physical entities changes with a change of basis from one coordinate system to another. When one coordinate system is just a rotation of the other, this distinction is invisible. However, when considering more general coordinate systems such as skew coordinates, curvilinear coordinates, and coordinate systems on differentiable manifolds, the distinction becomes critically important.

- For a **vector** (such as a direction vector or velocity vector) to be coordinate system invariant, the components of the vector must contra-vary with a change of basis to compensate. That is, the components must vary in the opposite "direction" (the inverse transformation) as the change of basis. Vectors (as opposed to dual vectors) are said to be **contravariant**. Examples of *contravariant* vectors include the position of an object relative to an observer, or any derivative of position with respect to time, including velocity, acceleration, and jerk. In Einstein notation, contravariant components have **upper indices** as in

  $$\mathbf{v} = v^i \mathbf{e}_i.$$

- For a **dual vector**, (such as a gradient) to be coordinate system invariant, the components of the vector must co-vary with a change of basis to maintain the same meaning. That is, the components must vary by the same transformation as the change of basis. Dual vectors (as opposed to vectors) are said to be **covariant**. Examples of *covariant* vectors generally appear when taking a gradient of a function (effectively dividing by a vector). In Einstein notation, covariant components have **lower indices** as in

  $$\mathbf{v} = v_i \mathbf{e}^i.$$

In physics, vectors often have units of distance or distance times some other unit (such as the velocity), whereas covectors have units the inverse of distance or the inverse of distance times some other unit. The distinction between covariant and contravariant vectors is particularly important for computations with tensors, which often have **mixed variance**. This means that they have both covariant and contravariant components, or both vectors and dual vectors. The **valence** or **type** of a tensor is number of variant and covariant terms. The duality between covariance and contravariance intervenes whenever a vector or tensor quantity is represented by its components, although modern differential geometry uses more sophisticated index-free methods to represent tensors.

The terms covariant and contravariant were introduced by J.J. Sylvester in 1853 in order to study algebraic invariant theory. In this context, for instance, a system of simultaneous equations is contravariant in the variables. The use of both terms in the modern context of multilinear algebra is a specific example of corresponding notions in category theory.

Introduction

In physics, a vector typically arises as the outcome of a measurement or series of measurements, and is represented as a list (or tuple) of numbers such as

$$(v_1,v_2,v_3).$$

This list of numbers depends on the choice of coordinate system. For instance, if the vector represents position with respect to an observer, then the coordinate system may be obtained from a system of rigid rods, or reference axes, along which the components v_1, v_2, and v_3 are measured. For a vector to represent a geometric object, it must be possible to describe how it looks in any other coordinate system. That is to say, the components of the vectors will *transform* in a certain way in passing from one coordinate system to another.

A *contravariant vector* is required to have components that "transform in the same way as the coordinates" (the opposite way as the reference axes) under changes of coordinates such as rotation and dilation. The vector itself does not change under these operations; instead, the components of the vector make a change that cancels the change in the spatial axes, in the same way that co-ordinates change. In other words, if the reference axes were rotated in one direction, the component representation of the vector would rotate in exactly the opposite way. Similarly, if the reference axes were stretched in one direction, the components of the vector, like the co-ordinates, would reduce in an exactly compensating way. Mathematically, if the coordinate system undergoes a transformation described by an invertible matrix M, so that a coordinate vector $\mathbf{x}$ is transformed to $\mathbf{x}' = M\mathbf{x}$, then a contravariant vector $\mathbf{v}$ must be similarly transformed via $\mathbf{v}' = M\mathbf{v}$. This important requirement is what distinguishes a contravariant vector from any other triple of physically meaningful quantities. For example, if v consists of the x, y, and z-components of velocity, then v is a contravariant vector: if the coordinates of space are stretched, rotated, or twisted, then the components of the velocity transform in the same way. On the other hand, for instance, a triple consisting of the length, width, and height

of a rectangular box could make up the three components of an abstract vector, but this vector would not be contravariant, since rotating the box does not change the box's length, width, and height. Examples of contravariant vectors include displacement, velocity, electric field, momentum, force, and acceleration.

By contrast, a *covariant vector* has components that change oppositely to the coordinates or, equivalently, transform like the reference axes. For instance, the components of the gradient vector of a function

$$\nabla f = \frac{\partial f}{\partial x_1}\widehat{x}_1 + \frac{\partial f}{\partial x_2}\widehat{x}_2 + \frac{\partial f}{\partial x_3}\widehat{x}_3$$

transform like the reference axes themselves. When only rotations of the spatial are considered, the components of contravariant and covariant vectors behave in the same way. It is only when other transformations are allowed that the difference becomes apparent.

Definition

The general formulation of covariance and contravariance refers to how the components of a coordinate vector transform under a change of basis (passive transformation). Thus let V be a vector space of dimension n over the field of scalars S, and let each of $\mathbf{f} = (X_1,...,X_n)$ and $\mathbf{f}' = (Y_1,...,Y_n)$ be a basis of V. Also, let the change of basis from **f** to **f′** be given by

$$\mathbf{f} \mapsto \mathbf{f}' = \left(\sum_i a_1^i X_i, \ldots, \sum_i a_n^i X_i\right) = \mathbf{f}A \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (1)$$

for some invertible $n \times n$ matrix A with entries a_j^i. Here, each vector Y_j of the **f′** basis is a linear combination of the vectors X_i of the **f** basis, so that

$$Y_j = \sum_i a_j^i X_i$$

.

Contravariant transformation

A vector v in V is expressed uniquely as a linear combination of the elements of the **f** basis as

$$v = \sum_i v^i[\mathbf{f}] X_i \quad \text{..... (2)}$$

where $v^i[\mathbf{f}]$ are scalars in S known as the **components** of v in the **f** basis. Denote the column vector of components of v by $\mathbf{v}[\mathbf{f}]$:

$$\mathbf{v}[\mathbf{f}] = \begin{bmatrix} v^1[\mathbf{f}] \\ v^2[\mathbf{f}] \\ \vdots \\ v^n[\mathbf{f}] \end{bmatrix}$$

so that (**2**) can be rewritten as a matrix product

$$v = \mathbf{f}\,\mathbf{v}[\mathbf{f}].$$

The vector v may also be expressed in terms of the **f'** basis, so that

$$v = \mathbf{f}'\,\mathbf{v}[\mathbf{f}'].$$

However, since the vector v itself is invariant under the choice of basis,

$$\mathbf{f}\,\mathbf{v}[\mathbf{f}] = v = \mathbf{f}'\,\mathbf{v}[\mathbf{f}'].$$

The invariance of v combined with the relationship (**1**) between **f** and **f'** implies that

$$\mathbf{f}\,\mathbf{v}[\mathbf{f}] = \mathbf{f}A\,\mathbf{v}[\mathbf{f}A],$$

Covariant transformation

A linear functional α on V is expressed uniquely in terms of its **components** (scalars in S) in the **f** basis as

$$\alpha_i[\mathbf{f}] = \alpha(X_i), \quad i = 1, 2, \ldots, n.$$

These components are the action of α on the basis vectors X_i of the **f** basis.

Under the change of basis from **f** to **f'** (**1**), the components transform such that

$$\begin{aligned} \alpha_i[\mathbf{f}A] &= \alpha(Y_i) \\ &= \alpha\left(\textstyle\sum_j a_i^j X_j\right) \\ &= \textstyle\sum_j a_i^j \alpha(X_j) \\ &= \textstyle\sum_j a_i^j \alpha_j[\mathbf{f}] \end{aligned} \quad \text{..... (3)}$$

Denote the row vector of components of **α** by **α**[**f**]:

$$\alpha[\mathbf{f}] = \begin{bmatrix} \alpha_1[\mathbf{f}] & \alpha_2[\mathbf{f}] & \ldots & \alpha_n[\mathbf{f}] \end{bmatrix}$$

so that (**3**) can be rewritten as the matrix product

$$\alpha[\mathbf{f}A] = \alpha[\mathbf{f}]A.$$

Because the components of the linear functional α transform with the matrix A, these components are said to **transform covariantly** under a change of basis.

Had a column vector representation been used instead, the transformation law would be the transpose

$$\alpha^{\mathrm{T}}[\mathbf{f}A] = A^{\mathrm{T}}\alpha^{\mathrm{T}}[\mathbf{f}].$$

giving the transformation rule

$$\mathbf{v}[\mathbf{f}A] = A^{-1}\mathbf{v}[\mathbf{f}].$$

In terms of components,

$$v^i[\mathbf{f}A] = \sum_j \tilde{a}^i_j v^j[\mathbf{f}]$$

where the coefficients $\tilde{a}^i_j$ are the entries of the inverse matrix of A.

Because the components of the vector v transform with the *inverse* of the matrix A, these components are said to **transform contravariantly** under a change of basis.

Coordinates

The choice of basis **f** on the vector space V defines uniquely a set of coordinate functions on V, by means of

$$x^i[\mathbf{f}](v) = v^i[\mathbf{f}].$$

The coordinates on V are therefore contravariant in the sense that

$$x^i[\mathbf{f}A] = \sum_{k=1}^{n} \tilde{a}^i_k x^k[\mathbf{f}].$$

Conversely, a system of n quantities v^i that transform like the coordinates x^i on V defines a contravariant vector. A system of n quantities that transform oppositely to the coordinates is then a covariant vector.

This formulation of contravariance and covariance is often more natural in applications, in which there is a coordinate space (a manifold) on which vectors live as tangent vectors or cotangent vectors. Given a local coordinate system x^i on the manifold, the reference axes for the coordinate system are the vector fields

$$X_1 = \frac{\partial}{\partial x^1}, \ldots, X_n = \frac{\partial}{\partial x^n}.$$

This gives rise to the frame **f** = (X_1,...,X_n) at every point of the coordinate patch.

If y^i is a different coordinate system and

$$Y_1 = \frac{\partial}{\partial y^1}, \ldots, Y_n = \frac{\partial}{\partial y^n},$$

then the frame **f'** is related to the frame **f** by the inverse of the Jacobian matrix of the coordinate transition:

$$\mathbf{f}' = \mathbf{f} J^{-1}, \quad J = \left(\frac{\partial y^i}{\partial x^j} \right)_{i,j=1}^{n}.$$

Or, in indices,

$$\frac{\partial}{\partial y^i} = \sum_{j=1}^{n} \frac{\partial x^j}{\partial y^i} \frac{\partial}{\partial x^j}.$$

A tangent vector is by definition a vector that is a linear combination of the coordinate partials $\partial / \partial x^i$. Thus a tangent vector is defined by

$$v = \sum_{i=1}^{n} v^i[\mathbf{f}] X_i = \mathbf{f}\ \mathbf{v}[\mathbf{f}].$$

Such a vector is contravariant with respect to change of frame. Under changes in the coordinate system, one has

$$\mathbf{v}[\mathbf{f}'] = \mathbf{v}[\mathbf{f} J^{-1}] = J\, \mathbf{v}[\mathbf{f}].$$

Therefore the components of a tangent vector transform via

$$v^i[\mathbf{f}'] = \sum_{j=1}^{n} \frac{\partial y^i}{\partial x^j} v^j[\mathbf{f}].$$

Accordingly, a system of n quantities v^i depending on the coordinates that transform in this way on passing from one coordinate system to another is called a contravariant vector.

Covariant and contravariant components of a vector

In a Euclidean space V, there is little distinction between covariant and contravariant vectors, because the dot product allows for covectors to be identified with vectors. That is, a vector v determines uniquely a covector α via

$$\alpha(w) = v \cdot w$$

for all vectors w. Conversely, each covector α determines a unique vector v by this equation. Because of this identification of vectors with covectors, one may speak of the **covariant components** or **contravariant components** of a vector, that is, they are just representations of the same vector using reciprocal bases.

Given a basis $\mathbf{f} = (X_1,...,X_n)$ of V, there is a unique reciprocal basis $\mathbf{f}^{\#} = (Y^1,...,Y^n)$ of V determined by requiring

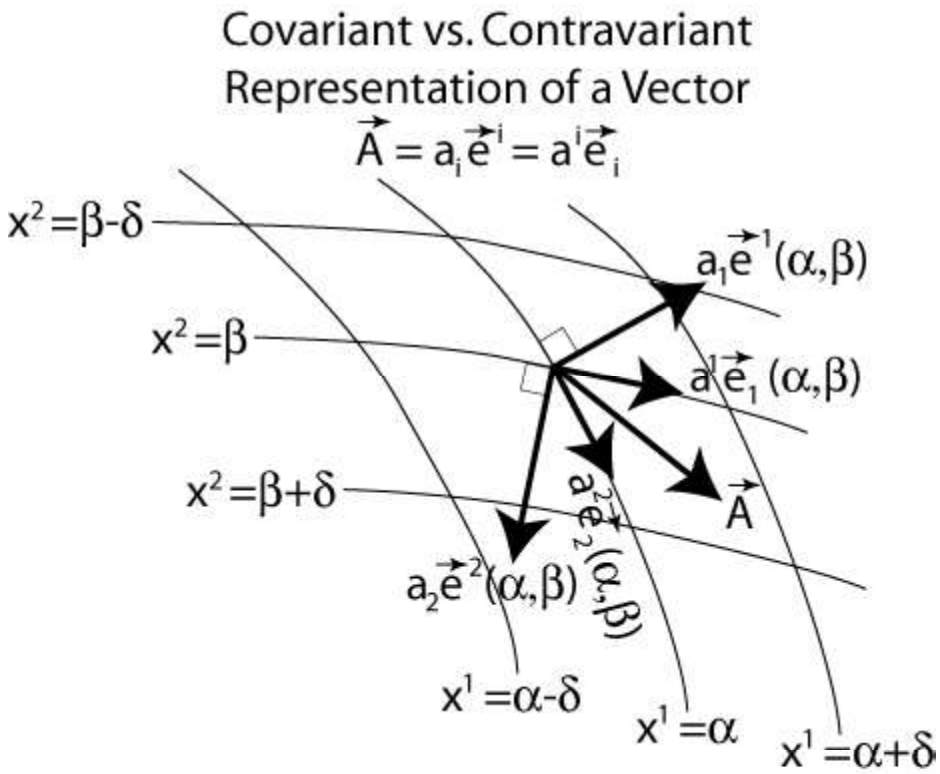
$$Y^i \cdot X_j = \delta^i_j,$$

the Kronecker delta. In terms of these bases, any vector v can be written in two ways:

$$\begin{aligned} v &= \sum_i v^i[\mathbf{f}] X_i = \mathbf{f}\, \mathbf{v}[\mathbf{f}] \\ &= \sum_i v_i[\mathbf{f}] Y^i = \mathbf{f}^\sharp \mathbf{v}^\sharp[\mathbf{f}]. \end{aligned}$$

The components $v^i[\mathbf{f}]$ are the **contravariant components** of the vector v in the basis **f**, and the components $v_i[\mathbf{f}]$ are the **covariant components** of v in the basis **f**. The terminology is justified because under a change of basis,

$$\mathbf{v}[\mathbf{f}A] = A^{-1}\mathbf{v}[\mathbf{f}], \quad \mathbf{v}^\sharp[\mathbf{f}A] = A^T \mathbf{v}^\sharp[\mathbf{f}].$$



The contravariant components of a vector are obtained by projecting onto the coordinate axes. The covariant components are obtained by projecting onto the normal lines to the coordinate hyperplanes.

Euclidean R^3

In the Euclidean space $\mathbf{R}^3$, the dot product allows for covectors to be identified with vectors.

This fact allows one to determine explicitly the dual basis to a given set of basis vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ of $\mathbf{R}^3$ that are not necessarily assumed to be orthogonal nor of unit norm. The contravariant (dual) basis vectors are:

$$\mathbf{e}^1 = \frac{\mathbf{e}_2 \times \mathbf{e}_3}{\mathbf{e}_1 \cdot (\mathbf{e}_2 \times \mathbf{e}_3)}; \qquad \mathbf{e}^2 = \frac{\mathbf{e}_3 \times \mathbf{e}_1}{\mathbf{e}_2 \cdot (\mathbf{e}_3 \times \mathbf{e}_1)}; \qquad \mathbf{e}^3 = \frac{\mathbf{e}_1 \times \mathbf{e}_2}{\mathbf{e}_3 \cdot (\mathbf{e}_1 \times \mathbf{e}_2)}.$$

Even if the $\mathbf{e}_i$ and $\mathbf{e}^i$ are not orthonormal, they are still by this definition mutually dual:

$$\mathbf{e}^i \cdot \mathbf{e}_j = \delta^i_j.$$

Then the contravariant coordinates of any vector $\mathbf{v}$ can be obtained by the dot product of $\mathbf{v}$ with the contravariant basis vectors:

$$q^1 = \mathbf{v} \cdot \mathbf{e}^1; \qquad q^2 = \mathbf{v} \cdot \mathbf{e}^2; \qquad q^3 = \mathbf{v} \cdot \mathbf{e}^3.$$

Likewise, the covariant components of **v** can be obtained from the dot product of **v** with covariant basis vectors, viz.

$$q_1 = \mathbf{v} \cdot \mathbf{e}_1; \qquad q_2 = \mathbf{v} \cdot \mathbf{e}_2; \qquad q_3 = \mathbf{v} \cdot \mathbf{e}_3.$$

Then **v** can be expressed in two (reciprocal) ways, viz.

$$\mathbf{v} = q_i \mathbf{e}^i = q_1 \mathbf{e}^1 + q_2 \mathbf{e}^2 + q_3 \mathbf{e}^3$$

or

$$\mathbf{v} = q^i \mathbf{e}_i = q^1 \mathbf{e}_1 + q^2 \mathbf{e}_2 + q^3 \mathbf{e}_3.$$

Combining the above relations, we have

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{e}_i)\mathbf{e}^i = (\mathbf{v} \cdot \mathbf{e}^i)\mathbf{e}_i$$

and we can convert from covariant to contravariant basis with

$$q_i = \mathbf{v} \cdot \mathbf{e}_i = (q^j \mathbf{e}_j) \cdot \mathbf{e}_i = (\mathbf{e}_j \cdot \mathbf{e}_i) q^j$$

and

$$q^i = \mathbf{v} \cdot \mathbf{e}^i = (q_j \mathbf{e}^j) \cdot \mathbf{e}^i = (\mathbf{e}^j \cdot \mathbf{e}^i) q_j.$$

The indices of covariant coordinates, vectors, and tensors are subscripts. If the contravariant basis vectors are orthonormal then they are equivalent to the covariant basis vectors, so there is no need to distinguish between the covariant and contravariant coordinates, and all indices are subscripts.

Informal usage

In the field of physics, the adjective **covariant** is often used informally as a synonym for invariant. For example, the Schrödinger equation does not keep its written form under the coordinate transformations of special relativity. Thus, a physicist might say that the Schrödinger equation is *not covariant*. In contrast, the Klein-Gordon equation and the Dirac equation do keep their written form under these coordinate transformations. Thus, a physicist might say that these equations are *covariant*.

Despite the dominant usage of "covariant", it is more accurate to say that the Klein-Gordon and Dirac equations are invariant, and that the Schrödinger equation is not invariant. Additionally, to remove ambiguity, the transformation by which the invariance is evaluated should be indicated. Continuing with the above example, neither the Klein-Gordon nor the Dirac equations are universally invariant under any coordinate transformation (e.g. those of general relativity), so unambiguous description of these

equations is that they are *invariant with respect to the coordinate transformations of special relativity*.

Because the components of vectors are contravariant and those of covectors are covariant, the vectors themselves are often referred to as being contravariant and the covectors as covariant. This usage is not universal, however, since vectors push forward – are covariant under diffeomorphism – and covectors pull back – are contravariant under diffeomorphism.

Use in tensor analysis

The distinction between covariance and contravariance is particularly important for computations with tensors, which often have **mixed variance**. This means that they have both covariant and contravariant components, or both vector and dual vector components. The valence of a tensor is the number of variant and covariant terms, and in Einstein notation, covariant components have lower indices, while contravariant components have upper indices. The duality between covariance and contravariance intervenes whenever a vector or tensor quantity is represented by its components, although modern differential geometry uses more sophisticated index-free methods to represent tensors.

In tensor analysis, a **covariant** vector varies more or less reciprocally to a corresponding contravariant vector. Expressions for lengths, areas and volumes of objects in the vector space can then be given in terms of tensors with covariant and contravariant indices. Under simple expansions and contractions of the coordinates, the reciprocity is exact; under affine transformations the components of a vector intermingle on going between covariant and contravariant expression.

On a manifold, a tensor field will typically have multiple indices, of two sorts. By a widely followed convention, covariant indices are written as lower indices, whereas contravariant indices are upper indices. When the manifold is equipped with a metric, covariant and contravariant indices become very closely related to one-another. Contravariant indices can be turned into covariant indices by contracting with the metric tensor. Contravariant indices can be gotten by contracting with the (matrix) inverse of the metric tensor. Note that in general, no such relation exists in spaces not endowed with a metric tensor. Furthermore, from a more abstract standpoint, a tensor is simply "there" and its components of either kind are only calculational artifacts whose values depend on the chosen coordinates.

The explanation in geometric terms is that a general tensor will have contravariant indices as well as covariant indices, because it has parts that live in the tangent bundle as well as the cotangent bundle.

A contravariant vector is one which transforms like $\frac{dx^\mu}{d\tau}$, where x^μ are the coordinates of a particle at its proper time τ. A covariant vector is one which transforms like $\frac{\partial \phi}{\partial x^\mu}$, where ϕ is a scalar field.

Algebra and geometry

In category theory, there are covariant functors and contravariant functors. The dual space of a vector space is a standard example of a contravariant functor. Some constructions of multilinear algebra are of 'mixed' variance, which prevents them from being functors.

In geometry, the same map in/map out distinction is helpful in assessing the variance of constructions. A tangent vector to a smooth manifold *M* is, to begin with, a curve mapping smoothly into *M* and passing through a given point *P*. It is therefore covariant, with respect to smooth mappings of *M*. A contravariant vector, or 1-form, is in the same way constructed from a smooth mapping from *M* to the real line, near *P*. It is in the cotangent bundle, built up from the dual spaces of the tangent spaces. Its *components with respect to* a local basis of one-forms dx_i will be covariant; but one-forms and differential forms in general are contravariant, in the sense that they pull back under smooth mappings. This is crucial to how they are applied; for example a differential form can be *restricted* to any submanifold, while this does not make the same sense for a field of tangent vectors.

Covariant and contravariant components transform in different ways under coordinate transformations. By considering a coordinate transformation on a manifold as a map from the manifold to itself, the transformation of covariant indices of a tensor are given by a pullback, and the transformation properties of the contravariant indices is given by a pushforward.

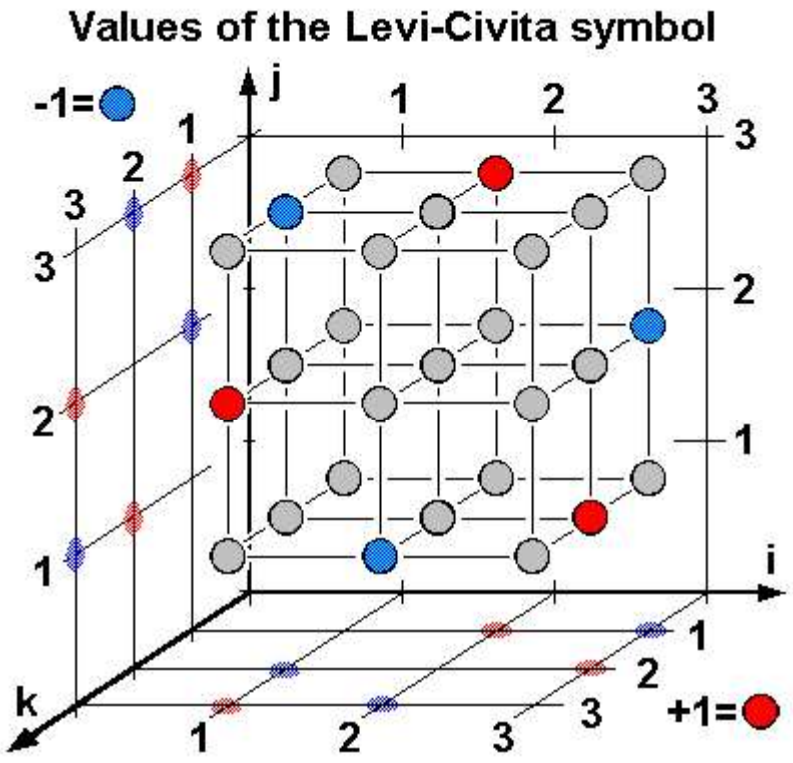
Chapter 18

Levi-Civita Symbol and Metric Tensor

Levi-Civita symbol

The **Levi-Civita symbol**, also called the **permutation symbol**, **antisymmetric symbol**, or **alternating symbol**, is a mathematical symbol used in particular in tensor calculus. It is named after the Italian mathematician and physicist Tullio Levi-Civita.

Definition



Values of the Levi-Civita-Symbol for a right-handed coordinate system

In three dimensions, the Levi-Civita symbol is defined as follows:

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i,j,k) \text{ is } (1,2,3),(3,1,2) \text{ or } (2,3,1), \\ -1 & \text{if } (i,j,k) \text{ is } (1,3,2),(3,2,1) \text{ or } (2,1,3), \\ 0 & \text{otherwise: } i=j \text{ or } j=k \text{ or } k=i, \end{cases}$$

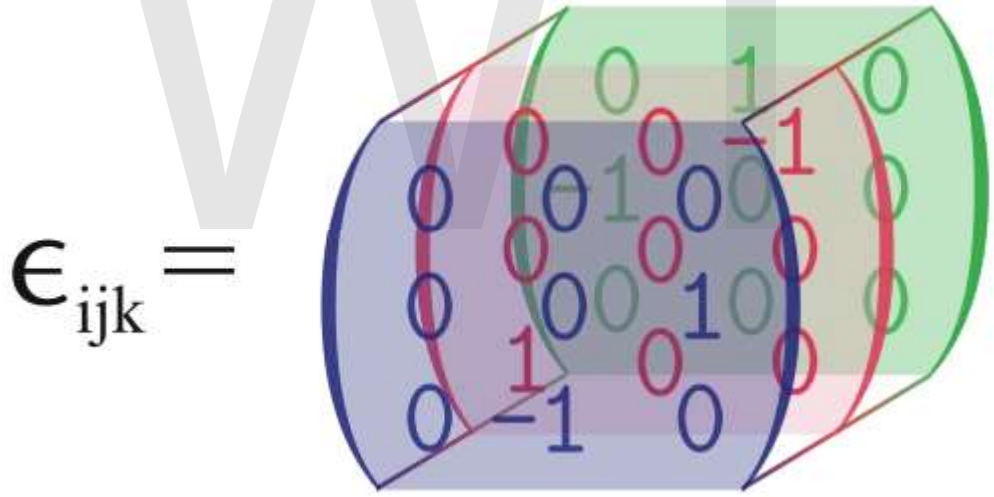
i.e. ε_{ijk} is 1 if (i, j, k) is an even permutation of (1,2,3), −1 if it is an odd permutation, and 0 if any index is repeated.

The formula for the three dimensional Levi-Civita symbol is:

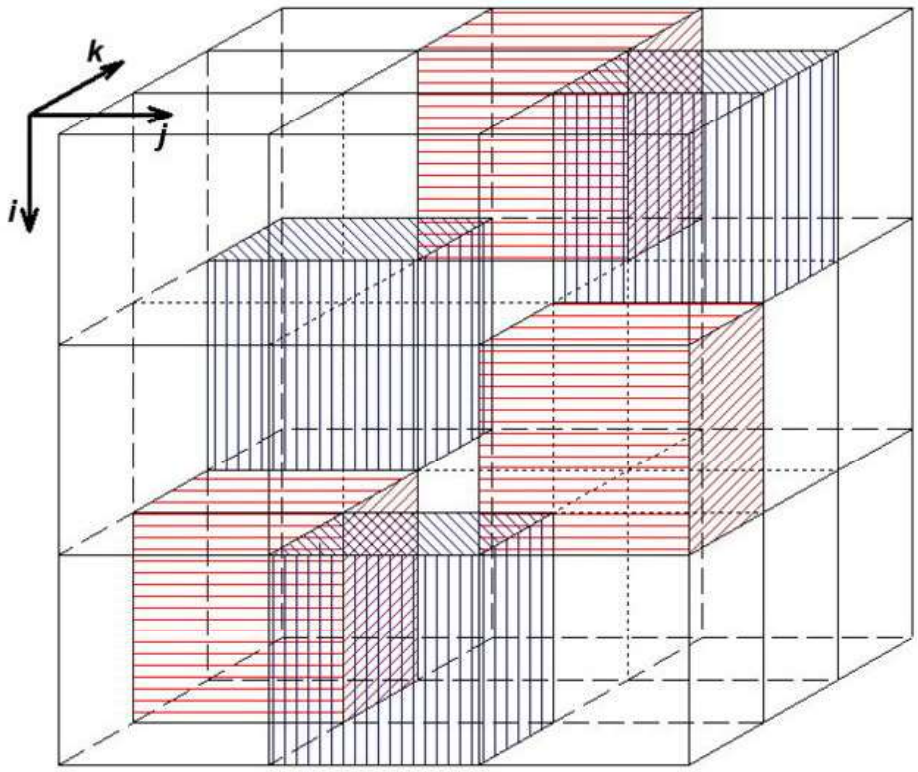
$$\varepsilon_{ijk} = \frac{(j-i)(k-i)(k-j)}{2} = \frac{(i-j)(j-k)(k-i)}{2}$$

The formula in four dimensions is:

$$\varepsilon_{ijkl} = \frac{(j-i)(k-i)(l-i)(k-j)(l-j)(l-k)}{12} = \frac{(i-j)(i-k)(i-l)(j-k)(j-l)(k-l)}{12}$$



Visualization of the Levi-Civita symbol as a 3×3×3 matrix



Corresponding visualization of the Levi-Civita-Symbol for a left-handed coordinate system. Empty cubes mean 0, red ones +1, and blue ones -1.

For example, in linear algebra, the determinant of a 3×3 matrix A can be written

$$\det A = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} \varepsilon_{ijk} a_{1i} a_{2j} a_{3k}$$

(and similarly for a square matrix of general size, see below)

and the cross product of two vectors can be written as a determinant:

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} \varepsilon_{ijk} \mathbf{e_i} a_j b_k$$

or more simply:

$$\mathbf{a} \times \mathbf{b} = \mathbf{c},\ c_i = \sum_{j=1}^{3} \sum_{k=1}^{3} \varepsilon_{ijk} a_j b_k.$$

According to the Einstein notation, the summation symbols may be omitted.

The tensor whose components in an orthonormal basis are given by the Levi-Civita symbol (a tensor of covariant rank n) is sometimes called the **permutation tensor**. It is actually a pseudotensor because under an orthogonal transformation of jacobian determinant −1 (i.e., a rotation composed with a reflection), it acquires a minus sign. Because the Levi-Civita symbol is a pseudotensor, the result of taking a cross product is a pseudovector, not a vector.

Note that under a general coordinate change, the components of the permutation tensor get multiplied by the jacobian of the transformation matrix. This implies that in coordinate frames different from the one in which the tensor was defined, its components can differ from those of the Levi-Civita symbol by an overall factor. If the frame is orthonormal, the factor will be ±1 depending on whether the orientation of the frame is the same or not.

Relation to Kronecker delta

The Levi-Civita symbol is related to the Kronecker delta. In three dimensions, the relationship is given by the following equations:

$$\varepsilon_{ijk}\varepsilon_{lmn} = \det \begin{bmatrix} \delta_{il} & \delta_{im} & \delta_{in} \\ \delta_{jl} & \delta_{jm} & \delta_{jn} \\ \delta_{kl} & \delta_{km} & \delta_{kn} \end{bmatrix}$$

$$= \delta_{il}\left(\delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km}\right) - \delta_{im}\left(\delta_{jl}\delta_{kn} - \delta_{jn}\delta_{kl}\right) + \delta_{in}\left(\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}\right)$$

$$\sum_{i=1}^{3} \varepsilon_{ijk}\varepsilon_{imn} = \delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km}$$

("contracted epsilon identity")

(In Einstein notation, the duplication of the i index implies the sum on i. The previous is then noted: $\varepsilon_{ijk}\varepsilon_{imn} = \delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km}$)

$$\sum_{i=1}^{3}\sum_{j=1}^{3} \varepsilon_{ijk}\varepsilon_{ijn} = 2\delta_{kn}$$

Generalization to *n* dimensions

The Levi-Civita symbol can be generalized to higher dimensions:

$$\varepsilon_{ijkl\ldots} = \begin{cases} +1 & \text{if } (i,j,k,l,\ldots) \text{ is an even permutation of } (1,2,3,4,\ldots) \\ -1 & \text{if } (i,j,k,l,\ldots) \text{ is an odd permutation of } (1,2,3,4,\ldots) \\ 0 & \text{otherwise} \end{cases}$$

Thus, it is the sign of the permutation in the case of a permutation, and zero otherwise.

The generalized formula is:

$$\varepsilon_{a_1 a_2 a_3 \ldots a_n} = \prod_{i=1}^{n-1} \left(\frac{1}{i!} \prod_{j=i+1}^{n} (a_j - a_i) \right)$$

where n is the dimension (rank).

For any *n* the property

$$\sum_{i,j,k,\ldots=1}^{n} \varepsilon_{ijk\ldots}\varepsilon_{ijk\ldots} = n!$$

follows from the facts that (a) every permutation is either even or odd, (b) $(+1)^2 = (-1)^2 = 1$, and (c) the permutations of any *n*-element set number exactly *n*!.

In index-free tensor notation, the Levi-Civita symbol is replaced by the concept of the Hodge dual.

In general *n* dimensions one can write the product of two Levi-Civita symbols as:

$$\varepsilon_{i_1 i_2 \ldots i_n}\varepsilon_{j_1 j_2 \ldots j_n} = \det \begin{bmatrix} \delta_{i_1 j_1} & \delta_{i_1 j_2} & \cdots & \delta_{i_1 j_n} \\ \delta_{i_2 j_1} & \delta_{i_2 j_2} & \cdots & \delta_{i_2 j_n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{i_n j_1} & \delta_{i_n j_2} & \cdots & \delta_{i_n j_n} \end{bmatrix}$$

Properties

(in these examples, superscripts should be considered equivalent with subscripts)

1. In two dimensions, when all *i*,*j*,*m*,*n* are in {1,2},

$$\varepsilon_{ij}\varepsilon^{mn} = \delta_i^m\delta_j^n - \delta_i^n\delta_j^m \qquad (1)$$
$$\varepsilon_{ij}\varepsilon^{in} = \delta_j^n \qquad (2)$$
$$\varepsilon_{ij}\varepsilon^{ij} = 2 \qquad (3)$$

2. In three dimensions, when all *i*,*j*,*k*,*m*,*n* are in {1,2,3},

$$\varepsilon_{jmn}\varepsilon^{imn} = 2\delta_j^i \qquad (4)$$
$$\varepsilon_{ijk}\varepsilon^{ijk} = 6 \qquad (5)$$
$$\varepsilon_{ijk}\varepsilon^{imn} = \delta_j^m\delta_k^n - \delta_j^n\delta_k^m \qquad (6)$$

3. In *n* dimensions, when all $i_1,...,i_n j_1,...,j_n$ are in {1,...,*n*},:

$$\varepsilon_{i_1\ldots i_n}\varepsilon^{j_1\cdots j_n} = n!\delta^{j_1}_{[i_1}\ldots\delta^{j_n}_{i_n]} \qquad (7)$$
$$\varepsilon_{i_1\ldots i_k\ i_{k+1}\ldots i_n}\varepsilon^{i_1\ldots i_k\ j_{k+1}\cdots j_n} = k!(n-k)!\,\delta^{j_{k+1}}_{[i_{k+1}}\ldots\delta^{j_n}_{i_n]} \qquad (8)$$
$$\varepsilon_{i_1\ldots i_n}\varepsilon^{i_1\ldots i_n} = n! \qquad (9)$$

Proofs

For equation 1, both sides are antisymmetric with respect of *ij* and *mn*. We therefore only need to consider the case $i \neq j$ and $m \neq n$. By substitution, we see that the equation holds for $\varepsilon_{12}\varepsilon^{12}$, i.e., for *i* = *m* = 1 and *j* = *n* = 2. (Both sides are then one). Since the equation is antisymmetric in *ij* and *mn*, any set of values for these can be reduced to the above case (which holds). The equation thus holds for all values of *ij* and *mn*. Using equation 1, we have for equation 2 $\varepsilon_{ij}\varepsilon^{in} = \delta_i^i\delta_j^n - \delta_i^n\delta_j^i$

$$= 2\delta_j^n - \delta_j^n$$
$$= \delta_j^n .$$

Here we used the Einstein summation convention with *i* going from 1 to 2. Equation 3 follows similarly from equation 2. To establish equation 4, let us first observe that both sides vanish when $i \neq j$. Indeed, if $i \neq j$, then one can not choose *m* and *n* such that both permutation symbols on the left are nonzero. Then, with *i* = *j* fixed, there are only two ways to choose *m* and *n* from the remaining two indices. For any such indices, we have $\varepsilon_{jmn}\varepsilon^{imn} = (\varepsilon^{imn})^2 = 1$ (no summation), and the result follows. Property (5) follows since 3! = 6 and for any distinct indices *i*,*j*,*k* in {1,2,3}, we have $\varepsilon_{ijk}\varepsilon^{ijk} = 1$ (no summation).

Examples

1. The determinant of an $n \times n$matrix $A = (a_{ij})$ can be written as

$$\det A = \varepsilon_{i_1 \cdots i_n} a_{1 i_1} \cdots a_{n i_n},$$

where each i_l should be summed over $1, \ldots, n$.

Equivalently, it may be written as

$$\det A = \frac{1}{n!} \varepsilon_{i_1 \cdots i_n} \varepsilon_{j_1 \cdots j_n} a_{i_1 j_1} \cdots a_{i_n j_n},$$

where now each i_l and each j_l should be summed over $1, \ldots, n$.

2. If $A = (A^1, A^2, A^3)$ and $B = (B^1, B^2, B^3)$ are vectors in R^3 (represented in some right hand oriented orthonormal basis), then the ith component of their cross product equals

$$(A \times B)^i = \varepsilon^{ijk} A^j B^k.$$

For instance, the first component of $A \times B$is $A^2B^3 - A^3B^2$. From the above expression for the cross product, it is clear that $A \times B = -B \times A$. Further, if $C = (C^1, C^2, C^3)$ is a vector like A and B, then the triple scalar product equals

$$A \cdot (B \times C) = \varepsilon^{ijk} A^i B^j C^k.$$

From this expression, it can be seen that the triple scalar product is antisymmetric when exchanging any adjacent arguments. For example, $A \cdot (B \times C) = -B \cdot (A \times C)$.

3. Suppose $F = (F^1, F^2, F^3)$ is a vector field defined on some open set of R^3 with Cartesian coordinates $x = (x^1, x^2, x^3)$. Then the ith component of the curl of F equals

$$(\nabla \times F)^i(x) = \varepsilon^{ijk} \frac{\partial}{\partial x^j} F^k(x).$$

Notation

A shorthand notation for anti-symmetrization is denoted by a pair of square brackets. For example, in arbitrary dimensions, for a rank 2 covariant tensor M,

$$M_{[ab]} = \frac{1}{2!}(M_{ab} - M_{ba}),$$

and for a rank 3 covariant tensor T,

$$T_{[abc]} = \frac{1}{3!}(T_{abc} - T_{acb} + T_{bca} - T_{bac} + T_{cab} - T_{cba}) .$$

In three dimensions, these are equivalent to

$$M_{[ab]} = \varepsilon_{abc} \frac{1}{2!} \varepsilon^{dec} M_{de} ,$$
$$T_{[abc]} = \varepsilon_{abc} \frac{1}{3!} \varepsilon^{def} T_{def} .$$

While in four dimensions, these are equivalent to

$$M_{[ab]} = \frac{1}{2!} \varepsilon_{abcd} \frac{1}{2!} \varepsilon^{efcd} M_{ef} ,$$
$$T_{[abc]} = \varepsilon_{abcd} \frac{1}{3!} \varepsilon^{efgd} T_{efg} .$$

More generally, in *n* dimensions

$$S_{[a_1 \ldots a_i]} = \frac{1}{(n-i)!} \varepsilon_{a_1 \ldots a_i \ b_1 \ldots b_{n-i}} \frac{1}{i!} \varepsilon^{c_1 \ldots c_i \ b_1 \ldots b_{n-i}} S_{c_1 \ldots c_i} .$$

Tensor density

In any arbitrary curvilinear coordinate system and even in the absence of a metric on the manifold, the Levi-Civita symbol as defined above may be considered to be a tensor density field in two different ways. It may be regarded as a contravariant tensor density of weight +1 or as a covariant tensor density of weight -1. In four dimensions,

$$\varepsilon^{\alpha\beta\gamma\delta} = \varepsilon_{\alpha\beta\gamma\delta} .$$

Notice that the value, and in particular the sign, does not change.

Ordinary tensor

In the presence of a metric tensor field, one may define an ordinary contravariant tensor field which agrees with the Levi-Civita symbol at each event whenever the coordinate system is such that the metric is orthonormal at that event. Similarly, one may also define an ordinary covariant tensor field which agrees with the Levi-Civita symbol at each event whenever the coordinate system is such that the metric is orthonormal at that event. These ordinary tensor fields should not be confused with each other, nor should they be confused with the tensor density fields mentioned above. One of these ordinary tensor fields may be converted to the other by raising or lowering the indices with the metric as is usual, but a minus sign is needed if the metric signature contains an odd number of

negatives. For example, in Minkowski space (the four dimensional spacetime of special relativity)

$$E^{\alpha\beta\gamma\delta} = -g^{\alpha\zeta}g^{\beta\eta}g^{\gamma\theta}g^{\delta\iota}E_{\zeta\eta\theta\iota}\,.$$

Notice the minus sign.

Metric tensor

In the mathematical field of differential geometry, a **metric tensor** is a type of function defined on a manifold (such as a surface in space) which takes as input a pair of tangent vectors *v* and *w* and produces a real number (scalar) *g*(*v*,*w*) in a way that generalizes many of the familiar properties of the dot product of vectors in Euclidean space. In the same way as a dot product, metric tensors are used to define the length of, and angle between, tangent vectors.

A metric tensor is defined to be a nondegenerate symmetric bilinear form on each tangent space that varies smoothly from point to point. It is an example of a tensor field. Relative to a local coordinate system, a metric tensor takes on the form of a symmetric matrix whose entries transform covariantly under changes to the coordinate system, which is to say that the metric tensor is a covariant symmetric tensor.

Introduction

Carl Friedrich Gauss in his 1827 *Disquisitiones generales circa superficies curvas* (*General investigations of curved surfaces*) considered a surface parametrically, with the Cartesian coordinates *x*, *y*, and *z* of points on the surface depending on two auxiliary variables *u* and *v*. Thus a parametric surface is (in today's terms) a vector valued function

$$\vec{r}(u,v) = (x(u,v), y(u,v), z(u,v))$$

depending on an ordered pair of real variables (*u*,*v*), and defined in an open set *D* in the *uv*-plane. One of the chief aims of Gauss' investigations was to deduce those features of the surface which could be described by a function which would remain unchanged if the surface underwent a transformation in space (such as bending the surface without stretching it), or a change in the particular parametric form of the same geometrical surface.

One natural such invariant quantity is the length of a curve drawn along the surface. Another is the angle between a pair of curves drawn along the surface and meeting at a common point, or tangent vectors at the same point of the surface. A third such quantity

is the area of a piece of the surface. The study of these invariants of a surface led Gauss to introduce the predecessor of the modern notion of the metric tensor.

Arclength

If the variables *u* and *v* are taken to depend on a third variable, *t*, taking values in an interval [*a*,*b*], then $\vec{r}(u(t),v(t))$ will trace out a parametric curve in *M*. The arclength of that curve is given by the integral

$$\begin{aligned} s &= \int_a^b \left\| \frac{d}{dt}\vec{r}(u(t),v(t)) \right\| dt \\ &= \int_a^b \sqrt{u'(t)^2\,\vec{r}_u \cdot \vec{r}_u + 2u'(t)v'(t)\,\vec{r}_u \cdot \vec{r}_v + v'(t)^2\,\vec{r}_v \cdot \vec{r}_v}\ dt. \end{aligned}$$

Here the chain rule has been applied, the subscripts denoting partial derivatives. The integrand is the restriction to the curve of the square root of the (quadratic) differential

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2 \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (1)$$

where

$$E = \vec{r}_u \cdot \vec{r}_u, \quad F = \vec{r}_u \cdot \vec{r}_v, \quad G = \vec{r}_v \cdot \vec{r}_v. \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (2)$$

The quantity *ds* in (**1**) is called the **line element**, while ds^2 is called the **first fundamental form** of *M*. Intuitively, it represents the principal part of the square of the displacement undergone by $\vec{r}(u,v)$ when *u* is increased by *du* units, and *v* is increased by *dv* units.

Suppose now that a different parameterization is selected, by allowing *u* and *v* to depend on another pair of variables *u*′ and *v*′. Then the analog of (**2**) for the new variables is

$$E' = \vec{r}_{u'} \cdot \vec{r}_{u'}, \quad F' = \vec{r}_{u'} \cdot \vec{r}_{v'}, \quad G' = \vec{r}_{v'} \cdot \vec{r}_{v'}. \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (2')$$

The chain rule relates *E*′, *F*′, and *G*′ to *E*,*F*, and *G* via the matrix equation

$$\begin{bmatrix} E' & F' \\ F' & G' \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial u'} & \frac{\partial u}{\partial v'} \\ \frac{\partial v}{\partial u'} & \frac{\partial v}{\partial v'} \end{bmatrix}^{\top} \begin{bmatrix} E & F \\ F & G \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial u'} & \frac{\partial u}{\partial v'} \\ \frac{\partial v}{\partial u'} & \frac{\partial v}{\partial v'} \end{bmatrix} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (3)$$

where the superscript *T* denotes the matrix transpose. The matrix with the coefficients *E*, *F*, and *G* arranged in this way therefore transforms by the **Jacobian matrix** of the coordinate change

$$J = \begin{bmatrix} \frac{\partial u}{\partial u'} & \frac{\partial u}{\partial v'} \\ \frac{\partial v}{\partial u'} & \frac{\partial v}{\partial v'} \end{bmatrix}.$$

A matrix which transforms in this way is one kind of what is called a tensor. The matrix

$$\begin{bmatrix} E & F \\ F & G \end{bmatrix}$$

with the transformation law (**3**) is known as the **metric tensor** of the surface.

Ricci-Curbastro & Levi-Civita (1900) first observed the significance of a system of coefficients *E*, *F*, and *G*, that transformed in this way on passing from one system of coordinates to another. The upshot is that the first fundamental form (**1**) is *invariant* under changes in the coordinate system, and that this follows exclusively from the transformation properties of *E*, *F*, and *G*. Indeed, by the chain rule,

$$\begin{bmatrix} du \\ dv \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial u'} & \frac{\partial u}{\partial v'} \\ \frac{\partial v}{\partial u'} & \frac{\partial v}{\partial v'} \end{bmatrix} \begin{bmatrix} du' \\ dv' \end{bmatrix}$$

so that

$$\begin{aligned} ds^2 &= \begin{bmatrix} du & dv \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix} \begin{bmatrix} du \\ dv \end{bmatrix} \\ &= \begin{bmatrix} du' & dv' \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial u'} & \frac{\partial u}{\partial v'} \\ \frac{\partial v}{\partial u'} & \frac{\partial v}{\partial v'} \end{bmatrix}^{\top} \begin{bmatrix} E & F \\ F & G \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial u'} & \frac{\partial u}{\partial v'} \\ \frac{\partial v}{\partial u'} & \frac{\partial v}{\partial v'} \end{bmatrix} \begin{bmatrix} du' \\ dv' \end{bmatrix} \\ &= \begin{bmatrix} du' & dv' \end{bmatrix} \begin{bmatrix} E' & F' \\ F' & G' \end{bmatrix} \begin{bmatrix} du' \\ dv' \end{bmatrix} \\ &= (ds')^2. \end{aligned}$$

Angle

Another interpretation of the metric tensor, also considered by Gauss, is that it provides a way in which to compute the angle between two tangent vectors to the surface. In contemporary terms, the metric tensor allows one to compute the dot product of tangent vectors in a manner independent of the parametric description of the surface. Any tangent vector at a point of the parametric surface *M* can be written in the form

$$X = a\vec{r}_u + b\vec{r}_v$$

for suitable real numbers *a* and *b*. If two tangent vectors are given

$$X_1 = a_1\vec{r}_u + b_1\vec{r}_v, X_2 = a_2\vec{r}_u + b_2\vec{r}_v$$

then using the bilinearity of the dot product,

$$\begin{aligned} X_1 \cdot X_2 &= a_1a_2\vec{r}_u \cdot \vec{r}_u + a_1b_2\vec{r}_u \cdot \vec{r}_v + b_1a_2\vec{r}_v \cdot \vec{r}_u + b_1b_2\vec{r}_v \cdot \vec{r}_v \\ &= a_1a_2E + a_1b_2F + b_1a_2F + b_1b_2G \\ &= \begin{bmatrix} a_1 & b_1 \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \end{aligned}.$$

This is plainly a function of the four variables a_1, b_1, a_2, and b_2. It is more profitably viewed, however, as a function that takes a pair of arguments $\mathbf{a} = [a_1\ b_1]$ and $\mathbf{b} = [a_2\ b_2]$ which are vectors in the *uv*-plane. That is, put

$$g(\mathbf{a}, \mathbf{b}) = a_1a_2E + a_1b_2F + b_1a_2F + b_1b_2G.$$

This is a symmetric function in **a** and **b**, meaning that

$$g(\mathbf{a}, \mathbf{b}) = g(\mathbf{b}, \mathbf{a}).$$

It is also bilinear meaning that it is linear in each variable **a** and **b** separately. That is,

$$\begin{aligned} &g(\lambda\mathbf{a} + \mu\mathbf{a}', \mathbf{b}) = \lambda g(\mathbf{a}, \mathbf{b}) + \mu g(\mathbf{a}', \mathbf{b}), \quad \text{and} \\ &g(\mathbf{a}, \lambda\mathbf{b} + \mu\mathbf{b}') = \lambda g(\mathbf{a}, \mathbf{b}) + \mu g(\mathbf{a}, \mathbf{b}') \end{aligned}$$

for any vectors **a**, **a**′, **b**, and **b**′ in the *uv* plane, and any real numbers μ and λ.

Area

The surface area is another numerical quantity which should depend only on the surface itself, and not on how it is parameterized. If the surface *M* is parameterized by the function $\vec{r}(u, v)$ over the domain *D* in the *uv*-plane, then the surface area of *M* is given by the integral

$$\iint_D |\vec{r}_u \times \vec{r}_v|\, du\, dv$$

where × denotes the cross product, and the absolute value denotes the length of a vector in Euclidean space. By Lagrange's identity for the cross product, the integral can be written

$$\iint_D \sqrt{(\vec{r}_u \cdot \vec{r}_u)(\vec{r}_v \cdot \vec{r}_v) - (\vec{r}_u \cdot \vec{r}_v)^2}\, du\, dv$$
$$= \iint_D \sqrt{EG - F^2}\, du\, dv$$
$$= \iint_D \sqrt{\det \begin{bmatrix} E & F \\ F & G \end{bmatrix}}\, du\, dv$$

where det is the determinant.

Definition

Let M be a smooth manifold of dimension n; for instance a surface (in the case $n = 2$) or hypersurface in the Cartesian space $\mathbf{R}^{n+1}$. At each point $p \in M$ there is a vector space T_pM, called the tangent space, consisting of all tangent vectors to the manifold at the point p. A **metric at *p*** is a function $g_p(X_p,Y_p)$ which takes as inputs a pair of tangent vectors X_p and Y_p at p, and produces as an output a real number (scalar), so that the following conditions are satisfied:

- g_p is **bilinear**. A function of two vector arguments is bilinear if it is linear separately in each argument. Thus if U_p, V_p, Y_p are three tangent vectors at p and a and b are real numbers, then

  $$g_p(aU_p + bV_p, Y_p) = ag_p(U_p, Y_p) + bg_p(V_p, Y_p), \quad \text{and}$$
  $$g_p(Y_p, aU_p + bV_p) = ag_p(Y_p, U_p) + bg_p(Y_p, V_p).$$

- g_p is **symmetric**. A function of two vector arguments is symmetric provided that for all vectors X_p and Y_p,

  $$g_p(X_p, Y_p) = g_p(Y_p, X_p).$$

- g_p is **nondegenerate**. A bilinear function is nondegenerate provided that, for every tangent vector $X_p \neq 0$, the function

  $$Y_p \mapsto g_p(X_p, Y_p)$$
  obtained by holding X_p constant and allowing Y_p to vary is not identically zero. That is, for every $X_p \neq 0$ there exists a Y_p such that $g_p(X_p,Y_p) \neq 0$.

A **metric tensor** g on M assigns to each point p of M a metric g_p in the tangent space at p such that in a way that varies smoothly with p. More precisely, given any open subset U of manifold M and any (smooth) vector fields X and Y on U, the real function

$$g(X, Y)(p) = g_p(X_p, Y_p)$$

is a smooth function of p.

Components of the metric

This section assumes some familiarity with coordinate vectors.

The components of the metric in any basis of vector fields, or frame, $\mathbf{f} = (X_1, \ldots, X_n)$ are given by

$$g_{ij}[\mathbf{f}] = g(X_i, X_j). \quad \textbf{(4)}$$

The n^2 functions $g_{ij}[\mathbf{f}]$ form the entries of an $n \times n$ symmetric matrix, $G[\mathbf{f}]$. If

$$v = \sum_{i=1}^{n} v^i X_i, \quad w = \sum_{i=1}^{n} w^i X_i$$

are two vectors at $p \in U$, then value of the metric applied to v and w is determined by the coefficients (**4**) by bilinearity:

$$g(v, w) = \sum_{i,j=1}^{n} v^i w^j g(X_i, X_j) = \sum_{i,j=1}^{n} v^i w^j g_{ij}[\mathbf{f}]$$

Denoting the matrix $(g_{ij}[\mathbf{f}])$ by $G[\mathbf{f}]$ and arranging the components of the vectors v and w into column vectors $\mathbf{v}[\mathbf{f}]$ and $\mathbf{w}[\mathbf{f}]$,

$$g(v, w) = \mathbf{v}[\mathbf{f}]^\top G[\mathbf{f}]\mathbf{w}[\mathbf{f}] = \mathbf{w}[\mathbf{f}]^\top G[\mathbf{f}]\mathbf{v}[\mathbf{f}]$$

where $\mathbf{v}[\mathbf{f}]^T$ and $\mathbf{w}[\mathbf{f}]^T$ denote the transpose of the vectors $\mathbf{v}[\mathbf{f}]$ and $\mathbf{w}[\mathbf{f}]$, respectively. Under a change of basis of the form

$$\mathbf{f} \mapsto \mathbf{f}' = \left(\sum_k X_k a_{k1}, \ldots, \sum_k X_k a_{kn} \right) = \mathbf{f}A$$

for some invertible $n \times n$ matrix $A = (a_{ij})$, the matrix of components of the metric changes by A as well. That is,

$$G[\mathbf{f}A] = A^\top G[\mathbf{f}]A$$

or, in terms of the entries of this matrix,

$$g_{ij}[\mathbf{f}A] = \sum_{k,\ell=1}^{n} a_{ki} g_{k\ell}[\mathbf{f}] a_{\ell j}.$$

For this reason, the system of quantities $g_{ij}[\mathbf{f}]$ is said to **transform covariantly** with respect to changes in the frame **f**.

Metric in coordinates

A system of *n* real valued functions $(x^1, \ldots, x^n)$, giving a local coordinate system on an open set *U* in *M*, determines a basis of vector fields on *U*

$$\mathbf{f} = \left(X_1 = \frac{\partial}{\partial x^1}, \ldots, X_n = \frac{\partial}{\partial x^n} \right).$$

The metric *g* has components relative to this frame given by

$$g_{ij}[\mathbf{f}] = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right).$$

Relative to a new system of local coordinates, say

$$y^i = y^i(x^1, x^2, \ldots, x^n), \quad i = 1, 2, \ldots, n$$

the metric tensor will determine a different matrix of coefficients,

$$g_{ij}[\mathbf{f}'] = g\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j} \right).$$

This new system of functions is related to the original $g_{ij}(\mathbf{f})$ by means of the chain rule

$$\frac{\partial}{\partial y^i} = \sum_{k=1}^{n} \frac{\partial x^k}{\partial y^i} \frac{\partial}{\partial x^k}$$

so that

$$g_{ij}[\mathbf{f}'] = \sum_{k,\ell=1}^{n} \frac{\partial x^k}{\partial y^i} g_{k\ell}[\mathbf{f}] \frac{\partial x^\ell}{\partial y^j}.$$

Or, in terms of the matrices $G[\mathbf{f}] = (g_{ij}[\mathbf{f}])$ and $G[\mathbf{f}'] = (g_{ij}[\mathbf{f}'])$,

$$G[\mathbf{f}'] = \left((Dy)^{-1} \right)^\top G[\mathbf{f}] (Dy)^{-1}$$

where *Dy* denotes the Jacobian matrix of the coordinate change.

Signature of a metric

Associated to any metric tensor is the quadratic form defined in each tangent space by

$$q_m(X_m) = g_m(X_m, X_m), \quad X_m \in T_m M.$$

If q_m is positive for all non-zero X_m, then the metric is positive definite at m. If the metric is positive definite at every $m \in M$, then g is called a Riemannian metric. More generally, if the quadratic forms q_m have constant signature independent of m, then the **signature of g** is this signature, and g is called a pseudo-Riemannian metric. If M is connected, then the signature of q_m does not depend on m.

By Sylvester's law of inertia, a basis of tangent vectors X_i can be chosen locally so that the quadratic form diagonalizes in the following manner

$$q_m\left(\sum_i \xi^i X_i\right) = (\xi^1)^2 + (\xi^2)^2 + \cdots + (\xi^p)^2 - (\xi^{p+1})^2 - \cdots - (\xi^n)^2$$

for some p between 1 and n. Any two such expressions of q (at the same point m of M) will have the same number p of positive signs. The signature of g is the pair of integers $(p, n-p)$, signifying that there are p positive signs and $n-p$ negative signs in any such expression. Equivalently, the metric has signature $(p,n-p)$ if the matrix g_{ij} of the metric has p positive and $n-p$ negative eigenvalues.

Certain metric signatures which arise frequently in applications are:

- If g has signature $(n, 0)$, then g is a Riemannian metric, and M is called a Riemannian manifold. Otherwise, g is a pseudo-Riemannian metric, and M is called a pseudo-Riemannian manifold (the term semi-Riemannian is also used).
- If M is four-dimensional with signature (1,3) or (3,1), then the metric is called Lorentzian. More generally, a metric tensor in dimension n other than 4 of signature $(1,n-1)$ or $(n-1, 1)$ is sometimes also called Lorentzian.
- If M is $2n$-dimensional and g has signature (n,n), then the metric is called ultrahyperbolic.

Inverse metric

Let $\mathbf{f} = (X_1, \ldots, X_n)$ be a basis of vector fields, and as above let $G[\mathbf{f}]$ be the matrix of coeffients

$$g_{ij}[\mathbf{f}] = g(X_i, X_j).$$

One can consider the inverse matrix $G[\mathbf{f}]^{-1}$, which is identified with the **inverse metric** (or *conjugate* or *dual metric*). The inverse metric satisfies a transformation law when the frame **f** is changed by a matrix A via

$$G[\mathbf{f}\,A]^{-1} = A^{-1}G[\mathbf{f}]^{-1}(A^{-1})^{\top}. \quad \textbf{(5)}$$

The inverse metric transforms *contravariantly*, or with respect to the inverse of the change of basis matrix A. Whereas the metric itself provides a way to measure the length of (or angle between) vector fields, the inverse metric supplies a means of measuring the length of (or angle between) covector fields; that is, fields of linear functionals.

To see this, suppose that α is a covector field. To wit, for each point p, α determines a function α_p defined on tangent vectors at p so that the following linearity condition holds for all tangent vectors X_p and Y_p, and all real numbers a and b:

$$\alpha_p(aX_p + bY_p) = a\alpha_p(X_p) + b\alpha_p(Y_p).$$

As p varies, α is assumed to be a smooth function in the sense that

$$p \mapsto \alpha_p(X_p)$$

is a smooth function of p for any smooth vector field X.

Any covector field α has components in the basis of vector fields **f**. These are determined by

$$\alpha_i = \alpha(X_i), \quad i = 1, 2, \ldots, n.$$

Denote the row vector of these components by

$$\alpha[\mathbf{f}] = [\alpha_1 \quad \alpha_2 \quad \ldots \quad \alpha_n].$$

Under a change of **f** by a matrix A, α[**f**] changes by the rule

$$\alpha[\mathbf{f}\,A] = \alpha[\mathbf{f}]A.$$

That is, the row vector of components α[**f**] transforms as a *covariant* vector.

For a pair α and β of covector fields, define the inverse metric applied to these two covectors by

$$\tilde{g}(\alpha, \beta) = \alpha[\mathbf{f}]G[\mathbf{f}]^{-1}\beta[\mathbf{f}]^{\top}. \quad \textbf{(6)}$$

The resulting definition, although it involves the choice of basis **f**, does not actually depend on **f** in an essential way. Indeed, changing basis to **f**A gives

$$\begin{aligned}\alpha[\mathbf{f}A]G[\mathbf{f}A]^{-1}\beta[\mathbf{f}A]^{\top} &= (\alpha[\mathbf{f}]A)\left(A^{-1}G[\mathbf{f}]^{-1}(A^{-1})^{\top}\right)A^{\top}\beta[\mathbf{f}]^{\top} \\ &= \alpha[\mathbf{f}]G[\mathbf{f}]^{-1}\beta[\mathbf{f}]^{\top}.\end{aligned}$$

So that the right-hand side of equation (**6**) is unaffected by changing the basis **f** to any other basis **f***A* whatsoever. Consequently, the equation may be assigned a meaning independently of the choice of basis. The entries of the matrix $G[\mathbf{f}]$ are denoted by g^{ij}, where the indices *i* and *j* have been raised to indicate the transformation law (**5**).

Raising and lowering indices

In a basis of vector fields $\mathbf{f} = (X_1, \ldots, X_n)$, any smooth tangent vector field *X* can be written in the form

$$X = v^1[\mathbf{f}]X_1 + v^2[\mathbf{f}]X_2 + \cdots + v^n[\mathbf{f}]X_n = \mathbf{f}\begin{bmatrix} v^1[\mathbf{f}] \\ v^2[\mathbf{f}] \\ \vdots \\ v^n[\mathbf{f}] \end{bmatrix} = \mathbf{f}v[\mathbf{f}] \quad \cdot (7)$$

for some uniquely determined smooth functions $v^1, \ldots, v^n$. Upon changing the basis **f** by a nonsingular matrix *A*, the coefficients v^i change in such a way that equation (**7**) remains true. That is,

$$X = \mathbf{f}\mathbf{A}v[\mathbf{f}\mathbf{A}] = \mathbf{f}v[\mathbf{f}].$$

Consequently, $v[\mathbf{f}A] = A^{-1}v[\mathbf{f}]$. In other words, the components of a vector transform *contravariantly* (with respect to the inverse) under a change of basis by the nonsingular matrix *A*. The contravariance of the components of $v[\mathbf{f}]$ is notationally designated by placing the indices of $v^i[\mathbf{f}]$ in the upper position.

A frame also allows covectors to be expressed in terms of their components. For the basis of vector fields $\mathbf{f} = (X_1, \ldots, X_n)$ define the dual basis to be the linear functionals $(\theta^1[\mathbf{f}], \ldots, \theta^n[\mathbf{f}])$ such that

$$\theta^i[\mathbf{f}](X_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

That is, $\theta^i[\mathbf{f}](X_j) = \delta_j^i$, the Kronecker delta. Let

$$\theta[\mathbf{f}] = \begin{bmatrix} \theta^1[\mathbf{f}] \\ \theta^2[\mathbf{f}] \\ \vdots \\ \theta^n[\mathbf{f}] \end{bmatrix}.$$

Under a change of basis $\mathbf{f} \to \mathbf{f}A$ for a nonsingular matrix A, $\theta[\mathbf{f}]$ transforms via

$$\theta[\mathbf{f}A] = A^{-1}\theta[\mathbf{f}].$$

Any linear functional α on tangent vectors can be expanded in terms of the dual basis θ

$$\begin{aligned} \alpha &= a_1[\mathbf{f}]\theta^1[\mathbf{f}] + a_1[\mathbf{f}]\theta^1[\mathbf{f}] + \cdots + a_n[\mathbf{f}]\theta^n[\mathbf{f}] \\ &= \begin{bmatrix} a_1[\mathbf{f}] & a_2[\mathbf{f}] & \dots & a_n[\mathbf{f}] \end{bmatrix} \theta[\mathbf{f}] = a[\mathbf{f}]\theta[\mathbf{f}] \end{aligned} \quad \textbf{(8)}$$

where $a[\mathbf{f}]$ denotes the row vector $[a_1[\mathbf{f}] \dots a_n[\mathbf{f}]]$. The components a_i transform when the basis $\mathbf{f}$ is replaced by $\mathbf{f}A$ in such a way that equation (**8**) continues to hold. That is,

$$\alpha = a[\mathbf{f}A]\theta[\mathbf{f}A] = a[\mathbf{f}]\theta[\mathbf{f}]$$

whence, because $\theta[\mathbf{f}A] = A^{-1}\theta[\mathbf{f}]$, it follows that $a[\mathbf{f}A] = a[\mathbf{f}]A$. That is, the components a transform *covariantly* (by the matrix A rather than its inverse). The covariance of the components of $a[\mathbf{f}]$ is notationally designated by placing the indices of $a_i[\mathbf{f}]$ in the lower position.

Now, the metric tensor gives a means to identify vectors and covectors as follows. Holding X_p fixed, the function

$$g_p(X_p, -) : Y_p \mapsto g_p(X_p, Y_p)$$

of tangent vector Y_p defines a linear functional on the tangent space at p. This operation takes a vector X_p at a point p and produces a covector $g_p(X_p, -)$. In a basis of vector fields $\mathbf{f}$, if a vector field X has components $v[\mathbf{f}]$, then the components of the covector field $g(X, -)$ in the dual basis are given by the entries of the row vector

$$a[\mathbf{f}] = v[\mathbf{f}]^{\top} G[\mathbf{f}].$$

Under a change of basis $\mathbf{f} \to \mathbf{f}A$, the right-hand side of this equation transforms via

$$v[\mathbf{f}A]^{\top} G[\mathbf{f}A] = v[\mathbf{f}]^{\top} (A^{-1})^{\top} A^{\top} G[\mathbf{f}]A = v[\mathbf{f}]^{\top} G[\mathbf{f}]A$$

so that $a[\mathbf{f}A] = a[\mathbf{f}]A$: a transforms covariantly. The operation of associating to the (contravariant) components of a vector field $v[\mathbf{f}] = [v^1[\mathbf{f}]\ v^2[\mathbf{f}] \ldots v^n[\mathbf{f}]\ ]^{\mathrm{T}}$ the (covariant) components of the covector field $a[\mathbf{f}] = [a_1[\mathbf{f}]\ a_2[\mathbf{f}] \ldots a_n[\mathbf{f}]\ ]$ where

$$a_i[\mathbf{f}] = \sum_{k=1}^{n} v^k[\mathbf{f}]g_{ki}[\mathbf{f}]$$

is called **lowering the index**.

To *raise the index*, one applies the same construction but with the inverse metric instead of the metric. If $a[\mathbf{f}] = [a_1[\mathbf{f}]\ a_2[\mathbf{f}] \ldots a_n[\mathbf{f}]\ ]$ are the components of a covector in the dual basis $\theta[\mathbf{f}]$, then the column vector

$$v[\mathbf{f}] = G^{-1}[\mathbf{f}]a[\mathbf{f}]^{\top} \quad \text{(9)}$$

has components which transform contravariantly:

$$v[\mathbf{f}A] = A^{-1}v[\mathbf{f}].$$

Consequently, the quantity $X = \mathbf{f}v[\mathbf{f}]$ does not depend on the choice of basis **f** in an essential way, and thus defines a vector field on M. The operation (**9**) associating to the (covariant) components of a covector $a[\mathbf{f}]$ the (contravariant) components of a vector $v[\mathbf{f}]$ given is called **raising the index**. In components, (**9**) is

$$v^i[\mathbf{f}] = \sum_{k=1}^{n} g^{ik}[\mathbf{f}]a_k[\mathbf{f}].$$

Induced metric

Let U be an open set in $\mathbf{R}^n$, and let φ be a continuously differentiable function from U into the Euclidean space $\mathbf{R}^m$ where $m > n$. The mapping φ is called an immersion if φ is an injective function and the Jacobian matrix of φ has rank n at every point of U. The image of φ is called an immersed submanifold.

Suppose that φ is an immersion onto the submanifold $M \subset \mathbf{R}^m$. The usual Euclidean dot product in $\mathbf{R}^m$ is a metric which, when restricted to vectors tangent to M, gives a means for taking the dot product of these tangent vectors. This is called the **induced metric**.

Suppose that v is a tangent vector at a point of U, say

$$v = v^1\mathbf{e}_1 + \cdots + v^n\mathbf{e}_n$$

where $\mathbf{e}_i$ are the standard coordinate vectors in $\mathbf{R}^n$. When φ is applied to U, the vector v goes over to the vector tangent to M given by

$$\phi_*(v) = \sum_{i=1}^{n} \sum_{a=1}^{m} v^i \frac{\partial \phi^a}{\partial x^i} \mathbf{e}_a.$$

(This is called the pushforward of *v* along φ.) Given two such vectors, *v* and *w*, the induced metric is defined by

$$g(v, w) = \phi_*(v) \cdot \phi_*(w).$$

It follows from a straightforward calculation that the matrix of the induced metric in the basis of coordinate vector fields **e** is given by

$$G(\mathbf{e}) = (D\phi)^\top (D\phi)$$

where Dφ is the Jacobian matrix:

$$D\phi = \begin{bmatrix} \frac{\partial \phi^1}{\partial x^1} & \frac{\partial \phi^1}{\partial x^2} & \cdots & \frac{\partial \phi^1}{\partial x^n} \\ \frac{\partial \phi^2}{\partial x^1} & \frac{\partial \phi^2}{\partial x^2} & \cdots & \frac{\partial \phi^2}{\partial x^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi^m}{\partial x^1} & \frac{\partial \phi^m}{\partial x^2} & \cdots & \frac{\partial \phi^m}{\partial x^n} \end{bmatrix}.$$

Intrinsic definitions of a metric

The notion of a metric can be defined intrinsically using the language of fiber bundles and vector bundles. In these terms, a **metric tensor** is a function

$$g : TM \times_M TM \to \mathbf{R} \quad \text{(5)}$$

from the fiber product of the tangent bundle of *M* to **R** such that the restriction of *g* to each fiber is a nondegenerate bilinear mapping

$$g_p : T_pM \times T_pM \to \mathbf{R}.$$

The mapping (**5**) is required to be continuous, and often continuously differentiable, smooth, or real analytic, depending on the case of interest, and whether *M* can support such a structure.

Metric as a section of a bundle

By the universal property of the tensor product, any bilinear mapping (**5**) gives rise naturally to a section $g_\otimes$ of the dual of the tensor product bundle of T*M* with itself

$$g_\otimes \in \Gamma\left((TM \otimes TM)^*\right).$$

The section $g_\otimes$ is defined on simple elements of TM⊗TM by

$$g_\otimes(v \otimes w) = g(v, w)$$

and is defined on arbitrary elements of TM⊗TM by extending linearly to linear combinations of simple elements. The original bilinear form g is symmetric if and only if

$$g_\otimes \circ \tau = g_\otimes$$

where

$$\tau : TM \otimes TM \xrightarrow{\cong} TM \otimes TM$$

is the braiding map.

Since M is finite-dimensional, there is a natural isomorphism

$$(TM \otimes TM)^* \cong T^*M \otimes T^*M,$$

so that $g_\otimes$ is regarded also as a section of the bundle T*M⊗T*M of the cotangent bundle T*M with itself. Since g is symmetric as a bilinear mapping, it follows that $g_\otimes$ is a symmetric tensor.

Metric in a vector bundle

More generally, one may speak of a metric in a vector bundle. If E is a vector bundle over a manifold M, then a metric is a mapping

$$g : E \times_M E \to \mathbf{R}$$

from the fiber product of E to **R** which is bilinear in each fiber:

$$g_p : E_p \times E_p \to \mathbf{R}.$$

Using duality as above, a metric is often identified with a section of the tensor product bundle $E^*\otimes E^*$.

Tangent-cotangent isomorphism

The metric tensor gives a natural isomorphism from the tangent bundle to the cotangent bundle, sometimes called the musical isomorphism. This isomorphism is obtained by setting, for each tangent vector $X_p \in \mathrm{T}_pM$,

$$S_g X_p \stackrel{def}{=} g(X_p, -),$$

the linear functional on T_pM which sends a tangent vector Y_p at p to $g_p(X_p,Y_p)$. That is, in terms of the pairing $[-,-]$ between T_pM and its dual space T_p^*M,

$$[S_g X_p, Y_p] = g_p(X_p, Y_p)$$

for all tangent vectors X_p and Y_p. The mapping S_g is a linear transformation from T_pM to T_p^*M. It follows from the definition of non-degeneracy that the kernel of S_g is reduced to zero, and so by the rank-nullity theorem, S_g is a linear isomorphism. Furthermore, S_g is a symmetric linear transformation in the sense that

$$[S_g X_p, Y_p] = [S_g Y_p, X_p]$$

for all tangent vectors X_p and Y_p.

Conversely, any linear isomorphism $S : T_pM \to T_pM$ defines a non-degenerate bilinear form on T_pM by means of

$$g_S(X_p, Y_p) = [SX_p, Y_p].$$

This bilinear form is symmetric if and only if S is symmetric. There is thus a natural one-to-one correspondence between symmetric bilinear forms on T_pM and symmetric linear isomorphisms of T_pM to the dual T_p^*M.

As p varies over M, S_g defines a section of the bundle $\mathrm{Hom}(TM,T^*M)$ of vector bundle isomorphisms of the tangent bundle to the cotangent bundle. This section has the same smoothness as g: it is continuous, differentiable, smooth, or real-analytic according as g. The mapping S_g, which associates to every vector field on M a covector field on M gives an abstract formulation of "lowering the index" on a vector field. The inverse of S_g is a mapping $T^*M \to TM$ which, analogously, gives an abstract formulation of "raising the index" on a covector field.

The inverse S_g^{-1} defines a linear mapping

$$S_g^{-1} : T^*M \to TM$$

which is nonsingular and symmetric in the sense that

$$[S_g^{-1}\alpha, \beta] = [S_g^{-1}\beta, \alpha]$$

for all covectors α, β. Such a nonsingular symmetric mapping gives rise (by the tensor-hom adjunction) to a map

$$T^*M \otimes T^*M \to \mathbf{R}$$

or by the double dual isomorphism to a section of the tensor product

$$TM \otimes TM.$$

Arclength and the line element

Suppose that g is a Riemannian metric on M. In a local coordinate system x^i, $i = 1,2,\ldots,n$, the metric tensor appears as a matrix, denoted here by **G**, whose entries are the components g_{ij} of the metric tensor relative to the coordinate vector fields.

Let $\gamma(t)$ be a piecewise differentiable parametric curve in M, for $a \leq t \leq b$. The **arclength** of the curve is defined by

$$L = \int_a^b \sqrt{\sum_{i,j=1}^{n} g_{ij}(\gamma(t)) \left(\frac{d}{dt} x^i \circ \gamma(t)\right) \left(\frac{d}{dt} x^j \circ \gamma(t)\right)}\, dt.$$

In connection with this geometrical application, the quadratic differential form

$$ds^2 = \sum_{i,j=1}^{n} g_{ij}(p) dx^i dx^j$$

is called the first fundamental form associated to the metric, while ds is the line element. When ds^2 is pulled back to the image of a curve in M, it represents the square of the differential with respect to arclength.

For a pseudo-Riemannian metric, the length formula above is not always defined, because the term under the square root may become negative. We generally only define the length of a curve when the quantity under the square root is always of one sign or the other. In this case, define

$$L = \int_a^b \sqrt{\left|\sum_{i,j=1}^{n} g_{ij}(\gamma(t)) \left(\frac{d}{dt} x^i \circ \gamma(t)\right) \left(\frac{d}{dt} x^j \circ \gamma(t)\right)\right|}\, dt\ .$$

Note that, while these formulas use coordinate expressions, they are in fact independent of the coordinates chosen; they depend only on the metric, and the curve along which the formula is integrated.

The energy, variational principles and geodesics

Given a segment of a curve, another frequently defined quantity is the (kinetic) **energy** of the curve:

$$E = \frac{1}{2}\int_a^b \sum_{i,j=1}^{n} g_{ij}(\gamma(t)) \left(\frac{d}{dt}x^i \circ \gamma(t)\right)\left(\frac{d}{dt}x^j \circ \gamma(t)\right) dt.$$

This usage comes from physics, specifically, classical mechanics, where the integral E can be seen to directly correspond to the kinetic energy of a point particle moving on the surface of a manifold. Thus, for example, in Jacobi's formulation of Maupertuis principle, the metric tensor can be seen to correspond to the mass tensor of a moving particle.

In many cases, whenever a calculation calls for the length to be used, a similar calculation using the energy may be done as well. This often leads to simpler formulas by avoiding the need for the square-root. Thus, for example, the geodesic equations may be obtained by applying variational principles to either the length or the energy. In the later case, the geodesic equations are seen to arise from the principle of least action: they describe the motion of a "free particle" (a particle feeling no forces) that is confined to move on the manifold, but otherwise moves freely, with constant momentum, within the manifold.

Canonical measure and volume form

In analogy with the case of surfaces, a metric tensor on an n-dimensional paracompact manifold M gives rise to a natural way to measure the n-dimensional volume of subsets of the manifold. The resulting natural positive Borel measure allows one to develop a theory of integrating functions on the manifold by means of the associated Lebesgue integral.

A measure can be defined, by the Riesz representation theorem, by giving a positive linear functional Λ on the space $C_0(M)$ of compactly supported continuous functions on M. More precisely, if M is a manifold with a (pseudo-)Riemannian metric tensor g, then there is a unique positive Borel measure μ_g such that for any coordinate chart (U,φ),

$$\Lambda f = \int_U f \, d\mu_g = \int_{\phi(U)} f \circ \phi^{-1}(x) \sqrt{|\det g|}\, dx$$

for all f supported in U. Here det g is the determinant of the matrix formed by the components of the metric tensor in the coordinate chart. That Λ is well-defined on functions supported in coordinate neighborhoods is justified by Jacobian change of variables. It extends to a unique positive linear functional on $C_0(M)$ by means of a partition of unity.

If M is in addition oriented, then it is possible to define a natural volume form from the metric tensor. In a positively oriented coordinate system $(x^1,\ldots,x^n)$ the volume form is represented as

$$\omega = \sqrt{|\det g|}\, dx^1 \wedge \cdots \wedge dx^n$$

where the dx^i are the coordinate differentials and the wedge ∧ denotes the exterior product in the algebra of differential forms. The volume form also gives a way to integrate functions on the manifold, and this geometric integral agrees with the integral obtained by the canonical Borel measure.

Examples

The Euclidean metric

The most familiar example is that of elementary Euclidean geometry: the two-dimensional Euclidean metric tensor. In the usual *x*-*y* coordinates, we can write

$$g = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The length of a curve reduces to the formula:

$$L = \int_a^b \sqrt{(dx)^2 + (dy)^2}.$$

The Euclidean metric in some other common coordinate systems can be written as follows.

Polar coordinates: (r, θ)

$x = r\cos\theta$
$y = r\sin\theta$

$$J = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}.$$

So

$$g = J^\top J = \begin{bmatrix} \cos^2\theta + \sin^2\theta & -r\sin\theta\cos\theta + r\sin\theta\cos\theta \\ -r\cos\theta\sin\theta + r\cos\theta\sin\theta & r^2\sin^2\theta + r^2\cos^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}$$

by trigonometric identities.

In general, in a Cartesian coordinate system x^i on a Euclidean space, the partial derivatives $\partial/\partial x^i$ are orthonormal with respect to the Euclidean metric. Thus the metric tensor is the Kronecker delta δ_{ij} in this coordinate system. The metric tensor with respect to arbitrary (possibly curvilinear) coordinates q^i is given by:

$$g_{ij} = \sum_{kl} \delta_{kl} \frac{\partial x^k}{\partial q^i} \frac{\partial x^l}{\partial q^j} = \sum_k \frac{\partial x^k}{\partial q^i} \frac{\partial x^k}{\partial q^j}.$$

The round metric on a sphere

The unit sphere in $\mathbf{R}^3$ comes equipped with a natural metric induced from the ambient Euclidean metric. In standard spherical coordinates (θ,φ), with θ the co-latitude, the angle measured from the z axis, and φ the angle from the x axis in the xy plane, the metric takes the form

$$g = \begin{bmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{bmatrix}.$$

This is usually written in the form

$$ds^2 = d\theta^2 + \sin^2\theta \, d\phi^2.$$

Lorentzian metrics from relativity

In flat Minkowski space (special relativity), with coordinates $r^\mu \to (x^0, x^1, x^2, x^3) = (ct, x, y, z)$, the metric is

$$g = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

For a curve with—for example—constant time coordinate, the length formula with this metric reduces to the usual length formula. For a timelike curve, the length formula gives the proper time along the curve.

In this case, the spacetime interval is written as

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 = dr^\mu dr_\mu = g_{\mu\nu} dr^\mu dr^\nu.$$

The Schwarzschild metric describes the spacetime around a spherically symmetric body, such as a planet, or a black hole. With coordinates $(x^0,x^1,x^2,x^3) = (ct,r,\theta,\varphi)$, we can write the metric as

$$G = (g_{\mu\nu}) = \begin{bmatrix} (1 - \frac{2GM}{rc^2}) & 0 & 0 & 0 \\ 0 & -(1 - \frac{2GM}{rc^2})^{-1} & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -r^2 \sin^2 \theta \end{bmatrix}$$

where G (inside the matrix) is the gravitational constant and M the mass of the body.